

TOPOLOGY AND LOGIC AS A SOURCE OF ALGEBRA

*Retiring Presidential Address*¹

BY SAUNDERS MAC LANE

Table of Contents

Introduction	1
1: Separable extensions	1
2: Homological algebra	5
3: The cohomology of groups	8
4: The bar construction	17
5: Coherence and canonical maps	24
6: Naturality and categories	33
7: What is algebra?	36

Introduction. Each President of the American Mathematical Society is required to present a retiring presidential address. By custom, this address is not in any way directed to the examination of the administrative or professional problems of the Society. This is not because these problems are unimportant, but is presumably because the President's contribution to them should have long since been made, or not made, as the case may be, well before his retirement from office. Instead the custom is that the President's address is concerned with contributions to our science. This is a good custom, since it is the mathematical theorems and the solutions of mathematical problems which possess the quality of permanence and the symmetry of structure which is our primary objective in mathematics.

My own research work has been largely concerned with aspects of algebra—a variety of aspects by no means covering all of algebra, but chiefly involved in some explorations of the relation between algebra and the neighboring fields of logic and geometry. These studies have given me the lively impression that many of the ideas of algebra do indeed arise from these other fields, and that this origin highlights the sense in which the science of mathematics exemplifies the interdependence of its parts. Hence, this address will be devoted to an examination of certain of the ways in which the problems from geometry and logic arising in my own research work have illuminated algebra or contributed new concepts to that field.

1. Separable extensions. My first example is a problem in the theory of fields whose solution required the use of ideas from geometry—more exactly, ideas about an axiomatic treatment of linear independence.

¹ Delivered at the 81st Annual Meeting of the American Mathematical Society in Washington, D.C., on January 24, 1975; received by the editors June 16, 1975.

AMS (MOS) subject classifications (1970). Primary 18-02, 12-02, 55-02, 02-02, 00-xx.

The problem begins with number theory. Here the treatment of various arithmetic questions had been effectively formulated in terms of the p -adic numbers. Recall that these numbers arise from studying the divisibility properties of rational numbers *vis-a-vis* a fixed prime number p . Each rational $x \neq 0$ may be written as $x = p^e(m/n)$ for some integral exponent e and integers m and n prime to p . Then the function $Vx = e$ assigning to x the exponent of p defines the “exponential” p -adic valuation V of x , with the formal properties

$$(1.1) \quad V(xy) = Vx + Vy, \quad V(x + y) \geq \text{Min}(Vx, Vy)$$

for all $x, y \neq 0$. The corresponding norm on the field Q of rational numbers is $|x| = \exp(-Vx)$. Completing Q with respect to this norm yields the field of p -adic numbers. The p -adic integers are then the x with $Vx \geq 0$, and the ring of p -adic integers has a unique maximal ideal (all x with $Vx > 0$), so that the integers modulo this ideal form a field, called the residue field. In this case, it is the familiar finite field with p elements. This structure uniquely determines the p -adic numbers: They are characterized (essentially) as a complete field with valuation having this residue field, and maximal ideal generated by p .

There is a similar structure theorem for more general p -adic fields. Let K be a field of characteristic 0 with an integer valued valuation V with the properties (1.1), complete in the corresponding norm. Its *residue* field k is again formed as the ring of its integers modulo the maximal ideal, and K is called p -adic if k is of characteristic p (whence $Vp > 0$) and the prime p generates the maximal ideal (K is “unramified” over the prime field). The structure theorems then state that such a p -adic field is determined, up to analytic isomorphism, by its residue field k .

Initially, I was engaged in finding two sorts of enlargements of this basic structure theorem—on the one hand, a structure theorem for complete fields with a valuation of higher rank [62]; on the other hand, a relative structure theorem [65], [66]. Is a complete and unramified extension L of K determined (up to relative analytic isomorphism) by the corresponding extension of the residue-class field? For both purposes I needed to examine more carefully the proof of the original “absolute” structure theorem for unramified complete fields. Here the uniqueness of the field K_0 of p -adic numbers for their residue field k_0 (the field of integers modulo p) is established first; then the whole complete field K is obtained by intermediate stages, each stage complete, and each corresponding to a stage in a step-by-step extension of the residue fields from k_0 to the whole residue field k . Typically one considers first transcendental extensions, by independent indeterminates x, y, z, \dots , and then a final algebraic extension to get all of k as a tower

$$k_0 \subset k_0(x) \subset k_0(x, y) \subset \dots \subset k_0(x, y, z, \dots) = k' \subset k.$$

Within the given field K one can try to construct a corresponding tower of p -adic fields

$$K_0 \subset K_1 \subset K_2 \subset \dots \subset K' \subset K,$$

with these respective residue fields. Specifically, if x is transcendental over k_0 , any element X in the coset x is transcendental over k_0 , and we may take K_1 to be the completion of $K_0(x)$. In the remaining case, if a is algebraic over k' , satisfying a separable polynomial equation, then the ‘‘Hensel-Rychlik’’ lemma will produce a corresponding element A algebraic over K' . However, this lemma really uses separability: A separable extension is one generated by roots of an irreducible polynomial whose roots are all different. This is the same notion of separability as that which arises in the Galois theory, where the standard correspondence between subgroups of the Galois prime and intermediate fields of a normal algebraic extension holds when that extension is separable.

However, it is precisely for fields of characteristic the prime p that inseparability can arise. For such fields, the binomial theorem for p th powers has the well-known simplified form:

$$(b + c)^p = b^p + c^p.$$

In consequence, an equation $X^p - a = 0$ over such a field k_0 , which may well be irreducible over that field, can have only one root α in any extended field, because $\alpha^p = a$ and so $(x - \alpha)^p = x^p - a$. This can happen unless k_0 is a *perfect* field of characteristic p ; that is, a field containing p th roots of all of its elements. By the little Fermat theorem, finite fields are perfect, but there are many infinite fields, such as transcendental extensions $k_0(x)$, which are not.

These observations about separability at first seemed to indicate that the structure theorems for p -adic fields worked well only when the residue field k was *separably generated* over k_0 , in the sense that k can be obtained from k_0 by the successive adjunction of elements each of which is either transcendental or separable and algebraic over the previously generated field. When k can be so generated, those elements which are transcendental can, for convenience, be adjoined first. This shows that k is separably generated over k_0 if and only if it has over k_0 a *separating transcendence base*: a string of elements x, y, z, \dots which are algebraically independent over k and such that k is separable and algebraic over the field $k_0(x, y, z, \dots)$ which they generate.

This use of separably generated extensions had led H. Hasse and F. K. Schmidt to an elaborate study of inseparably generated extensions in terms of Steinitz ‘‘towers’’ of separably generated extensions. Difficulties with this theory [64] persuaded me to think that the crucial problem was that of finding criteria for the existence of separating transcendence bases. There were striking examples of fields which did not have such bases; for example (Mac Lane [63, p. 384]) when $k_0 = P(x, y)$ where P is a ‘‘perfect’’ field (for example, a finite field) and x and y are algebraically independent, while $k = k_0(z, u)$ with z transcendental over k_0 and u the root of the inseparable equation

$$(1.2) \quad u^p = y + xz^p,$$

irreducible over $k_0(z)$. To be sure, one could solve equation (1.2) for y and so observe that this whole field k is $k=P(u, z, x)$ and thus is separably generated, with u, z , and x as the separating transcendence basis, over the original perfect field P . Unfortunately, this difficulty cannot always be removed just by starting from a perfect field P , as for example in the field $k_2=P(t_0, t_1, \dots, y_2, y_3, \dots)$ arising from P by adjoining two strings of elements t_n and y_n subject only to the algebraic relations

$$(1.3) \quad y_{n+2}^p = t_n + t_{n+1}(t_{n+2})^p, \quad n = 2, 3, 4, \dots$$

This field k_2 simply cannot be separably generated over P , a sorry fact which I had observed (Mac Lane [63, p. 386] and [64, p. 39]) in studying some of the difficulties with the Steinitz towers.

Study of example (1.2)—or similarly of (1.3)—might suggest that the trouble lies in the fact that equation (1.2) makes the p th root of y expressible in terms of the p th root of x , although in the original field $P(x, y)$ these roots are “independent.” Now Teichmüller [94], [95] had disengaged a notion of p -independence. An element z of a field k_0 of characteristic p is said to be p -dependent on elements y_1, \dots, y_n of k_0 when

$$(1.4) \quad z^{1/p} \in k_0(y_1^{1/p}, \dots, y_n^{1/p});$$

correspondingly, elements x_1, \dots, x_m are p -independent in k_0 where no one of them is p -dependent on the others.

At this point, geometry entered; more exactly *linear* geometry. Hassler Whitney had studied ordinary linear independence relations with a view to understanding the combinatorial structures called “matroids” (Whitney [102], Mac Lane [60]). Stimulated also by the ideas of Garrett Birkhoff and O. Ore, I had observed [61] that these notions of independence could be formulated axiomatically in terms of lattices (here, the lattice of intermediate fields). For ordinary linear independence Steinitz had shown that the invariance of dimension of a vector space rested on the exchange property. This property was still present for p -dependence: If z is p -dependent upon y_1, \dots, y_n but not on y_1, \dots, y_{n-1} , then y_n is p -dependent on y_1, \dots, y_{n-1}, z . Moreover, the lattice involved in this exchange property need *not* be a modular lattice, a surprise for lattice theorists (see Mac Lane [61, p. 463]).

For fields of characteristic p , an extension $k \supset k_0$ is said to *preserve* p -independence when each subset of k_0 , p -independent in k_0 , remains p -independent in the larger field k . This concept explains the trouble in the example of (1.2), where the elements x and y of k_0 are algebraically independent over P and hence p -independent in k_0 , but do not remain p -independent in k —precisely because equation (1.2) can be used to express the p th root of x in terms of the p th root of y . Using this concept, I was able to prove

THEOREM 1. *An extension $k \supset k_0$ of fields of characteristic p preserves p -independence if and only if $k_0(y_1, \dots, y_n) \supset k_0$ is separably generated for each finite string of elements y_1, \dots, y_n of k .*

COROLLARY. *A finitely generated extension can be separably generated if and only if it preserves p -independence.*

This concept also served to settle the *relative structure problem* for p -adic fields (Mac Lane [66]):

THEOREM 2. *Let k be any extension of the residue class field k_0 of a given p -adic field K_0 , and let K and L be two p -adic extensions of K_0 , both with the same residue class field k . Then the identity automorphism of K_0 can always be extended to an isomorphism $K \rightarrow L$ which is analytic (i.e., preserves the valuation) and induces the identity on k if and only if the extension $k \supset k_0$ of residue fields preserves p -independence.*

This completed the structure problem for p -adic fields, and shifted the problem to determining which extensions are separably generated. For fields of characteristic p , A. Weil [100] and N. Bourbaki [8] called an extension $k \supset k_0$ *separable* precisely when it preserves p -independence. For an algebraic extension, this agrees exactly with the classical notion, and it makes the Corollary to Theorem 1 read “A finitely generated extension is separably generated if and only if it is separable.” From this point on, the notion of separability became a small part in the field-theoretic treatment of algebraic geometry, relating in particular to the study of “linearly disjoint” extensions (Weil [100]). But in the process, the algebraic study of fields had profited essentially from Teichmüller’s observations that the “geometry” of linear dependence is like that of p -dependence, plus the fact that this study has a general lattice-theoretic form.

2. Homological algebra. The idea of finding the “homology” of an algebraic system—of a group, of a Lie algebra, or of an associative algebra—was clearly derived from the study of the homology (i.e., the connectivity) of a topological space. Originally, following Poincaré, the connectivity of a space had been described just in terms of its Betti numbers and torsion coefficients. Then it was reputedly Emmy Noether who had pointed out to Paul Alexandroff in Göttingen that the Betti number and the torsion coefficients in each dimension must really be the set of invariants of some finitely generated abelian group—the homology *group* of the space in that dimension. Noether emphasized that the proper subject of study must be this group.

The use of homology groups was soon systematized. They are usually described algebraically in terms of chains and their boundaries. An n -dimensional chain is a formal linear combination of n -simplices (either a simplex in a given triangulation of the space or any singular (continuous) simplex in the space). In any event, all the n -chains form a free abelian group C_n , and the sequence of these groups, with the corresponding boundary homomorphisms ∂ , is a *chain complex*.

$$(2.1) \quad C : C_0 \xleftarrow{\partial} C_1 \xleftarrow{\partial} C_2 \xleftarrow{\partial} \cdots \xleftarrow{\partial} C_n \xleftarrow{\partial} C_{n+1} \cdots$$

with $\partial\partial = 0$. Its homology in dimension n is the quotient group

$$(2.2) \quad H_n = H_n(C) = \{\text{Kernel } \partial : C_n \rightarrow C_{n-1}\} / \{\text{Image } \partial : C_{n+1} \rightarrow C_n\}.$$

This is the “integral” homology of the complex (or the underlying space); from the beginnings (see Veblen’s *Analysis situs* [97]) one had also used the homology modulo 2, formed by first reducing each free group C_n modulo 2, then forming H_n . J. W. Alexander had also pointed out the utility of chains formed modulo a prime p , i.e., with “coefficients” in \mathbf{Z}_p . For example, an outstanding question was that of finding the homotopy classes of continuous maps $g:K^n \rightarrow S^n$ of an n -dimensional polyhedron K^n into the n -sphere S^n ; H. Hopf had found a solution, expressed elaborately in terms of the action of g on the various homology groups of the complex K^n for various coefficients, $H_n(K^n, \mathbf{Z})$ and $H_n(K^n, \mathbf{Z}_p)$.

Hassler Whitney [103] saw how to simplify this result by formulating it in terms of cohomology (an idea possibly also suggested by the de Rham process of measuring connectivity in terms of “closed” and “exact” differential forms). Using “coefficients” in any abelian group A , an n -dimensional cochain of the complex C is a homomorphism $f:C_n \rightarrow A$ of abelian groups, its coboundary δf is the composite homomorphism

$$(2.3) \quad \delta f = f \partial : C_{n+1} \xrightarrow{\partial} C_n \longrightarrow A,$$

the groups C^n of all n -cochains again form a chain complex

$$(2.4) \quad \longrightarrow C^n = \text{Hom}(C_n, A) \xrightarrow{\delta} C^{n+1} = \text{Hom}(C_{n+1}, A) \longrightarrow \dots$$

with cocycles (the kernel of δ) and coboundaries (image δ) yielding

$$(2.5) \quad H^n(C, A) = \{\text{Ker } \delta : C^n \rightarrow C^{n+1}\} / \{\text{Image } \delta : C^{n-1} \rightarrow C^n\},$$

the cohomology of the complex C (or of the associated space) with coefficients in A . Each morphism $h:C \rightarrow C'$ of chain complexes then yields a morphism of cohomology

$$(2.6) \quad h : C \rightarrow C' \mapsto h^* : H^n(C', A) \rightarrow H^n(C, A)$$

in the opposite direction. In particular, the integral cohomology of the n -sphere is infinite cyclic in dimension n , $H^n(S^n, \mathbf{Z}) = \mathbf{Z}$. Under any continuous $g:K^n \rightarrow S^n$ from an n -complex K^n , the image of a generator under g^* is then a single cohomology class in $H^n(K^n, \mathbf{Z})$, and this one cohomology class, according to Whitney, suffices to describe the map g up to homotopy equivalence. This was the decisive result which made it clear that cohomology belonged together with homology in the topological descriptions of connectivity.

At the same time, Whitney introduced the notion of tensor products of abelian groups—a notion perhaps already implicit in multilinear algebra. These tensor products clarified the meaning of coefficient groups for homology. Initially, the homology of a complex C modulo the prime p (say, for $p=2$) had been described as the homology of the chain complex formed by simply taking each chain group C_n “modulo p .” As long as C_n is just a free abelian group, generated by simplices, this description is fine; one simply takes the free vector space over \mathbf{Z}_p generated by the same simplices. For more general groups C a more invariant description was needed; it is

provided by taking the tensor product $\mathbf{Z}_p \otimes C_n$. Again, the boundary homomorphism for the original complex induces a boundary homomorphism on the “tensoring” chains

$$(2.7) \quad 1 \otimes \partial : \mathbf{Z}_p \otimes C_n \rightarrow \mathbf{Z}_p \otimes C_{n-1}$$

(with 1 denoting the identity map of \mathbf{Z}_p), and the homology with coefficients \mathbf{Z}_p is the homology of the resulting complex. This done, one also can form the homology of C with coefficients in any abelian group G by forming the complex $G \otimes C$ and taking its homology as

$$(2.8) \quad H_n(C, G) = H_n(G \otimes C).$$

By this development, the study of connectivity of spaces has led directly to the formulation of the two basic functors

$$(2.9) \quad G \otimes C, \quad \text{Hom}(C, A)$$

of homological algebra. They have appeared here as constructions on abelian groups G, C , and A , but they also apply at once to modules over any commutative ring. Moreover, it was again a geometrical question which led to the study of the corresponding derived functors

$$(2.10) \quad \text{Tor}(G, C), \quad \text{Ext}(C, A)$$

and hence to the whole notion of derived functors in homological algebra.

For example, Hopf’s homotopy classification theorem for maps $g : K^n \rightarrow S^n$ in terms of the homology of the polyhedron K^n had been reformulated by Whitney in terms of the cohomology of K^n . This suggested, to Steenrod and others, that cohomology must somehow be expressible in terms of homology. Since cochains of a complex C are by definition homomorphisms of chains $f : C_n \rightarrow A$, each cocycle (f with $\delta f = 0$) is a homomorphism of cycles, and this assignment yields a “natural” homomorphism

$$(2.11) \quad H^n(C, A) \rightarrow \text{Hom}(H_n(C), A).$$

For A (the additive group of) a field, this is an isomorphism, but not for more general A ’s. Hence arose the problems of expressing the whole cohomology group of the complex C in terms of this homomorphism and other constructions; it was called the problem of “universal coefficients” because it was intended that the solution be given by saying that the cohomology is determined by giving the homology $H_n(C, G)$ for a specified list of coefficient groups G , called the “universal” coefficients, perhaps $G = \mathbf{Z}$ and all the groups \mathbf{Z}_p .

The final solution turned out to have a different conceptual structure. First, and easily, the homomorphism (2.11) is onto, so the problem is essentially that of finding its kernel. This was done in the first joint Eilenberg-Mac Lane paper [20], which showed that this kernel could be expressed as the group $\text{Ext}(H_{n-1}(C), A)$ of all abelian group extensions of A by H_{n-1} . This actually yields an expression of cohomology as a direct sum

$$(2.12) \quad H^n(C, A) \cong \text{Ext}(H_{n-1}(C), A) \oplus \text{Hom}(H_n(C), A),$$

but the isomorphism to the direct sum is not necessarily “natural” for variable C . Hence the invariant description of the resulting “universal coefficient” theorem states only that there is a (natural) short sequence which is *exact* (kernel=image at each point):

$$(2.13) \quad 0 \rightarrow \text{Ext}(H_{n-1}(C), A) \rightarrow H^n(C, A) \rightarrow \text{Hom}(H_n(C), A) \rightarrow 0.$$

Moreover, Eilenberg-Mac Lane showed how to calculate the new group Ext from any representation of the homology $H_{n-1}(C)$ as F/R , for F a free abelian group and R a subgroup. Such a representation is in fact itself a short exact sequence

$$(2.14) \quad 0 \rightarrow R \rightarrow F \rightarrow H_{n-1}(C) \rightarrow 0;$$

that is, what is now called a free (and therefore projective) *resolution* of the abelian group $H_{n-1}(C)$.

The corresponding universal coefficient question for homology is resolved by a corresponding short exact sequence (Mac Lane [71, Theorem V.11.1]),

$$(2.15) \quad 0 \rightarrow H_n(C) \otimes G \rightarrow H_n(C, G) \rightarrow \text{Tor}(H_{n-1}(C), G) \rightarrow 0,$$

involving another basic new functor, the torsion product Tor . Originally, in the hands of Čech, this group Tor was described in terms of generators and relations (Mac Lane [71, V.6]). Cartan-Eilenberg showed that it could be calculated from resolutions like (2.14) and hence that it was a *derived functor* in the general sense they developed for homological algebra.

Thus the algebraic constructions $\text{Ext}(B, A)$ and $\text{Tor}(B, G)$ on groups B , A , and G (or on modules) owe their origin to specific questions of algebraic topology. The resulting rapid development of homological algebra has considerably influenced ring theory and algebraic geometry.

3. The cohomology of groups. The study of a group involves not just the group itself, but its actions on other objects; for example, on vector spaces or modules. The cohomology of groups deals with such a situation: the action of a group G on an (additive) abelian group A . Specifically, let G act on the left on A , so that each element $x \in G$ is represented as an automorphism

$$a \mapsto xa, \quad a \in A,$$

of A , with the usual rules

$$x(a + b) = xa + xb, \quad (xy)a = x(ya), \quad 1a = a.$$

For example, G might be the Galois group of a field K , consisting of endomorphisms of K , while A is the additive or multiplicative group of K . Again, G might be a group of transformations of a space and A a homology or homotopy group of that space. Finally, A might be an arbitrary abelian group on which every element x of G acts as an identity automorphism; this is called the *trivial* action of G on A .

For any such action of G on A , the cohomology of G with “coefficients” in A was defined, dimension by dimension, in terms of certain “cocycles”

which are functions on G to A satisfying appropriate identities. In dimension 1, these cocycles are the *crossed homomorphisms*: functions h on G to A which satisfy the identity

$$(3.1) \quad h(xy) = x(hy) + hx$$

for all elements x and y of G . In particular, if the action of G on A is trivial, this equation becomes $h(xy) = hy + hx$, so in this case a “crossed” homomorphism is just an ordinary homomorphism $G \rightarrow A$. On the other hand, for a nontrivial action there are always such crossed homomorphisms; indeed for each element a in A one can define a *principal* crossed homomorphism k_a by

$$k_a x = xa - a, \quad x \in G.$$

Also, the pointwise sum of two crossed homomorphisms h_1 and h_2 , defined by

$$(h_1 + h_2)x = h_1x + h_2x,$$

is again a crossed homomorphism, and the crossed homomorphisms form a group under this addition. Now count as equivalent crossed homomorphisms h and $h + k_a$ for any principal crossed homomorphism k_a . Then we want the equivalence classes of crossed homomorphisms; they are cosets, hence elements of a factor group now called the *one-dimensional cohomology group* $H^1(G, A)$:

$$H^1(G, A) = \{h : G \rightarrow A \text{ a crossed homo}\} / \{k \text{ a principal c.h.}\}.$$

The two-dimensional analog of this group is defined similarly from “factor sets”. For the same data (G acting on A), a *factor set* f of G in A is a function $f : G \times G \rightarrow A$ which satisfies the equation

$$(3.2) \quad xf(y, z) + f(x, yz) = f(xy, z) + f(x, y)$$

for all x, y , and z in G . Again, the set of all solutions f of this equation forms a group under pointwise addition. Again, this equation (3.2) has “trivial” solutions; each function h on G to A determines a solution f_h , called a *transformation set*, by the formula

$$(3.3) \quad f_h(x, y) = x(hy) - h(xy) + h(x).$$

The two-dimensional cohomology group $H^2(G, A)$ is now defined as

$$H^2(G, A) = \{f : G \times G \rightarrow A \text{ a factor set}\} / \{f = f_h \text{ a trans. set}\}.$$

In other words, it is the group of all solutions of (3.2), counting solutions as equivalent when they differ by a formally trivial solution.

These definitions of $H^1(G, A)$ and $H^2(G, A)$ were chosen to study certain particular circumstances. The most important circumstance is given by a *group extension* E of A by G ; that is, a multiplicative group E with A as normal subgroup and $G \cong E/A$ the corresponding quotient group. Such an

extension may be presented as a sequence of groups and group homomorphisms with is *exact*:

$$(3.4) \quad 0 \longrightarrow A \longrightarrow E \xrightarrow{\pi} G \longrightarrow 1.$$

(A and E are written additively and G multiplicatively, for convenience below.) For any such extension, G does act as a group of automorphisms of A , the action of each $x \in G$ being given by conjugating the elements of A by any element in the coset of E determined by x . Thus, choosing for each $x \in G$ a representative element $u(x)$ in that coset, all the elements of E have the form $a+u(x)$ and the action of x is given by

$$(3.5) \quad xa = u(x) + a - u(x), \quad x \in G, \quad a \in A,$$

and is independent of the choice of $u(x)$ in its coset.

On the other hand, consider automorphisms $\mathcal{O}: E \rightarrow E$ which leave every element of A and every coset of G fixed. Such an automorphism \mathcal{O} carries each representative $u(x)$ into some other element in the same coset, say $\mathcal{O}(u(x)) = hx + u(x)$, where h is then a function $h: G \rightarrow A$. This function h actually determines the action of the automorphism \mathcal{O} on all elements $a+u(x)$ of E ; moreover, an easy calculation shows that the \mathcal{O} determined by this formula is an automorphism if and only if $h: G \rightarrow A$ is a crossed homomorphism. Also, the principal crossed homomorphisms k_a correspond to conjugation (i.e., give \mathcal{O} the conjugation by $-a$). Hence, the one-dimensional cohomology group is exactly the group of equivalence classes of such automorphisms \mathcal{O} .

Next consider how the “addition table” of the elements $a+u(x)$ of the group E might be constructed from the addition of A and the multiplication of G . The sum of two coset representatives $u(x)$ and $u(y)$ must lie in the coset of the product xy , and so must have the form

$$(3.6) \quad u(x) + u(y) = f(x, y) + u(xy),$$

where f is then a function $f: G \times G \rightarrow A$. This function f determines the whole addition table of E , while an easy calculation shows that the resulting addition is associative if and only if this function f satisfies (3.2), and so is a factor set. Moreover, a change in the choice of coset representatives, from $u(x)$ to (say) $h(x)+u(x)$, simply adds to the factor set f the transformation set f_h . Hence the two-dimensional cohomology group is exactly the group of all (equivalence classes of) those extensions E of A by G which realize the given action of G on A .

Classically, the subject stopped here at dimension 2. There were several indications that it was not really complete. For example, the “2-dimensional” factor sets also arose in class field theory, in the case when G is the Galois group of a finite separable normal extension $N \supset K$ of fields. If we take A to be the multiplicative group N^* of nonzero elements of the field N and so rewrite (3.6) in multiplicative form as $u(x)u(y) = f(x, y)u(xy)$, then the factor set f determines via this formula and distributivity a

multiplication table on the vector space D consisting of all finite formal sums $\sum a_x u(x)$ with $a_x \in N$ (i.e., the vector space D over N with basis the $u(x)$ for $x \in G$). Moreover, this multiplication table makes D a central simple algebra over the base field K . These D are the *crossed product algebras* first investigated (in many special cases) by L. E. Dickson and his students and then baptised (as “verschränkte produkte”) by Emmy Noether. They were used extensively in local class field theory, in the special case when K is a local field (of p -adic numbers); in this case the crossed product algebras formed from N give (up to “similarity”) all central simple algebras over K which are split by N . Hence they could be used to analyse the Brauer group of similarity classes of algebras over K .

O. F. G. Schilling, in joint work with the author, made a substantial attempt [78] to extend the class field theory from *abelian* extension fields to general normal extensions. The attempt did not succeed, but it did leave us with the clear impression that the 2-dimensional factor sets were not enough; something “higher up” was going on. At about the same time, Teichmüller’s study of noncommutative Galois theory [96] led him to consider algebras C which were central simple over the field N (of the extension $N \supset K$); he called such algebras “ G -normal” when every automorphism x of the Galois group G of $N \supset K$ could be extended to an automorphism of the algebra C , and he showed (by devices like the considerations of the obstructions to be indicated below) that each such algebra gave rise to a function $t: G \times G \times G \rightarrow N^*$ which satisfied a three-dimensional identity like that for factor sets:

$$(3.7) \quad xt(y, z, u)t(x, yz, u)t(x, y, z) = t(xy, z, u)t(x, y, zu)$$

for all $x, y, z, u \in G$. He was evidently aware that there were analogous identities of dimension higher than 3.

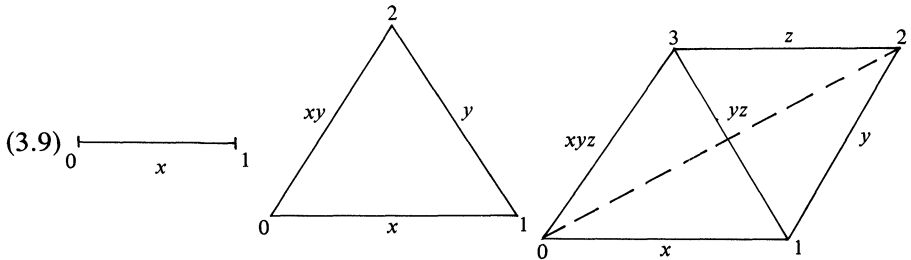
A unified understanding of these identities came from topology. It had long been known that the fundamental group $\Pi_1(X)$ of an arcwise connected space X determined the first integral homology group $H_1(X)$; indeed H_1 is just Π_1 modulo its commutative subgroup. Heinz Hopf in [44] and [45] studied the influence of Π_1 on the *second* integral homology group $H_2(X)$, and found that Π_1 determined not all of H_2 but a certain quotient group $H_2(X)/S_2(X)$, where $S_2(X)$ is the subgroup of $H_2(X)$ spanned by those two-dimensional cycles represented by 2-spheres. The situation simplifies if X is a space with vanishing second homotopy group $\Pi_2(X)$; for then $S_2(X)=0$ and Hopf’s result states that $\Pi_1(X)$ determines $H_2(X)$ —moreover, it determines H_2 by an explicit formula. Going beyond this, Hopf [46] considered spaces with $\Pi_2=\Pi_3=\dots=\Pi_{q-1}=0$ and showed that the fundamental group Π_1 of such a space does determine all the integral homology groups H_1, H_2, \dots, H_{q-1} and a (suitable) quotient group of $H_q(X)$. He gave no explicit formula for this latter determination.

Eilenberg and I decided that there ought to be an algebraic formula expressing this determination. It turned out to be more convenient to

describe a formula to determine the cohomology of X ; that is, the cohomology of the singular complex of X . Eilenberg had just completed the definitive description of this complex $S(X)$, to consist of chain groups S_n and boundary homomorphisms

$$(3.8) \quad S(X): S_0 \xleftarrow{\partial} S_1 \xleftarrow{\partial} S_2 \xleftarrow{\partial} \cdots \xleftarrow{\partial} S_{n-1} \xleftarrow{\partial} S_n \xleftarrow{\partial} \cdots$$

with $\partial\partial=0$ and S_n the free abelian group with generators all singular n -simplices T ; i.e., all continuous maps $T: \Delta_n \rightarrow X$ from the standard n -simplex Δ_n . Then ∂T is just the alternating sum of the $n+1$ faces of T ; each regarded as a map $\Delta_{n-1} \rightarrow X$. Since X is arcwise connected, one may fix a base point p and assume that all the singular simplices send each vertex of Δ_n to that base point p . A singular 1-simplex $T: \Delta_1 \rightarrow X$ is then just a continuous path from p to p , and hence represents an element $x \in \Pi_1(X)$. In the same way each n -simplex $T: \Delta_n \rightarrow X$ for $n > 1$ determines on each edge of Δ_n an element of the fundamental group; thus T may be pictured by the pattern of those edge elements, as in



Numbering the vertices in order, the edges 01 and 12 in the triangle represent elements x, y in Π_1 . Their product, in the fundamental group Π_1 , is the path 01 followed by 12, hence is equal (is homotopic to) the element represented by the edge 02 of Δ_2 . When $\Pi_2 = \cdots = \Pi_{q-1} = 0$ in the space X , two simplices (in dimensions less than q) with the same edge patterns will be homotopic; this observation reduces the singular complex to a so-called minimal subcomplex with one simplex for each possible edge pattern. For a simplex of n -dimensions, this pattern is described by n elements x_1, \dots, x_n of $\Pi_1(X)$ —the elements on the edges $(i-1, i)$. Thereby the original singular chain complex $S(X)$, defined topologically, is replaced by a complex $B(\Pi)$ determined algebraically by $\Pi = \Pi_1$. This complex $B(\Pi)$ consists of groups $B_n = B_n(\Pi)$ and boundary homomorphisms

$$(3.10) \quad B(\Pi_1): B_0 \xleftarrow{\partial} B_1 \xleftarrow{\partial} B_2 \xleftarrow{\partial} \cdots \xleftarrow{\partial} B_{n-1} \xleftarrow{\partial} B_n \xleftarrow{\partial} \cdots$$

where B_n is the free abelian group generated by the elements $[x_1, \dots, x_n]$, taken for all choices of $x_i \in \Pi_1$. The boundary of each generator $[x_1, \dots, x_n]$ may be read off from the geometric figure as the alternating sum of the edge

pattern on the faces. In particular, this yields the formulas:

$$\begin{aligned} \partial[x, y] &= [y] - [xy] + [x], \\ \partial[x, y, z] &= [y, z] - [xy, z] + [x, yz] - [x, y], \\ \partial[x_1, \dots, x_n] &= [x_2, \dots, x_{n+1}] + \sum_{i=1}^{n-1} (-1)^i [x_1, \dots, x_i x_{i+1}, \dots, x_n] \\ &\quad + (-1)^n [x_1, \dots, x_{n-1}]. \end{aligned}$$

The cohomology of the space X with coefficients in an abelian group A is therefore equal to the cohomology of the complex $\text{Hom}(B(\Pi_1), A)$, consisting of n -dimensional cochain groups $C^n = \text{Hom}(B_n, A)$. Since the group B_n is the free abelian group with generators the set $\Pi_1 \times \dots \times \Pi_1$, the group C^n of n -cochains is the group of all functions $f : \Pi_1 \times \dots \times \Pi_1 \rightarrow A$, with coboundary δf given by the general formula $(\delta f)c = f \partial c$; that is,

$$(\delta f)[x_1, \dots, x_n] = f \partial[x_1, \dots, x_n].$$

For $n=1$ and $n=2$ this becomes

$$\begin{aligned} (\delta h)[x, y] &= h(y) - h(xy) + h(x), \\ (\delta f)[x, y, z] &= f[y, z] - f[xy, z] + f[x, yz] - f[y, z]. \end{aligned}$$

In particular, $\delta h=0$ states that h is a (crossed) homomorphism $\Pi_1 \rightarrow A$, and $\delta f=0$ for $n=2$ that f is a factor set (both for the group Π_1 acting trivially on the group A of coefficients). Thereby the original algebraic identities (3.1), (3.2), and (3.7) for functions on G are interpreted geometrically in all dimensions, when G is taken to be the fundamental group Π_1 of a suitable space.

Consider more generally the case when Π_1 acts (nontrivially) on the abelian group A ; this amounts exactly to asserting that A is a (left) module over the integral group ring $\mathbf{Z}(\Pi_1)$. This case also has a geometric interpretation, if we replace the space X by its universal covering space \tilde{X} . The fundamental group Π_1 of X then becomes the group of covering transformations of \tilde{X} , so that Π_1 acts on each chain group $S_n(\tilde{X})$, making S_n a $\mathbf{Z}(\Pi_1)$ -module. The chain complex $B(\Pi_1)$ used above is now replaced by a corresponding complex for \tilde{X} of the form

$$(3.11) \quad E(\Pi_1): E_0 \xleftarrow{\partial} E_1 \xleftarrow{\partial} E_2 \longleftarrow \dots \longleftarrow E_{n-1} \xleftarrow{\partial} E_n \longleftarrow \dots,$$

where each E_n is the free $\mathbf{Z}(\Pi_1)$ -module generated by the elements $[x_1, \dots, x_n]$ and ∂ is the Π_1 -module homomorphism with

$$(3.12) \quad \begin{aligned} \partial[x_1, \dots, x_n] &= x_1[x_2, \dots, x_n] + \sum (-1)^i [x_1, \dots, x_i x_{i+1}, \dots, x_n] \\ &\quad + (-1)^{n+1} f(x_1, \dots, x_n). \end{aligned}$$

Now spaces (like \tilde{X}) with group actions (like covering transformations by Π_1) have an ‘‘equivariant’’ cohomology. Following Eilenberg’s earlier study

of this equivariant cohomology we now construct the complex C^n of equivariant cochains on $E(\Pi_1)$ with coefficients in the $\mathbf{Z}(\Pi_1)$ -module A . Here $C^n = \text{Hom}_{\Pi_1}(E_n, A)$ is the additive group of module maps $E_n \rightarrow A$, and so again consists of all functions $f: \Pi_1 \times \cdots \times \Pi_1 \rightarrow A$, with coboundary involving the Π_1 action as

$$(3.13) \quad \begin{aligned} (\delta f)(x_1, \cdots, x_{n+1}) &= x_1 f(x_2, \cdots, x_{n+1}) \\ &+ \sum_{i=1}^n (-1)^i f(x_1, \cdots, x_i x_{i+1}, \cdots, x_{n+1}) \\ &+ (-1)^{n+1} f(x_1, \cdots, x_n). \end{aligned}$$

The n -dimensional cohomology $H^n(\Pi_1, A)$ is now defined in all cases as the cohomology of this complex

$$(3.14) \quad H^n(\Pi_1, A) = \{\text{all } f: \Pi_1^n \rightarrow A \text{ with } \delta f = 0\} / \{f = \delta g, g: \Pi_1^{n-1} \rightarrow A\}.$$

The construction applies for any group $\Pi_1 = G$, and clearly includes the previously considered cases of crossed homomorphisms, factor sets, and Teichmüller cocycles. Thus the cohomology groups in all dimensions have geometric meanings, as summarized by the theorem whose proof we have just now indicated:

THEOREM 1. *If an arcwise connected space X has vanishing homotopy groups $\Pi_2 = \cdots = \Pi_n = 0$, its n -dimensional cohomology (and homology) depends only on the fundamental group $\Pi_1 = \Pi_1(X)$ and is given, for each abelian group A , by the group-cohomology:*

$$H^n(X, A) \cong H^n(\Pi_1(X), A).$$

THEOREM 2. *Let the group G act as a group of homeomorphisms on a space Y with vanishing homotopy groups $\Pi_1 = \cdots = \Pi_n = 0$ in such a way that no element $x \neq 1$ in G has a fixed point in Y . Then for each $\mathbf{Z}(G)$ -module A the n th equivariant cohomology group H_c^n of Y is given by the group cohomology as*

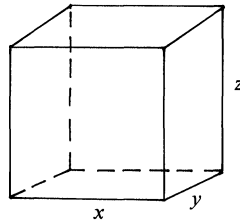
$$H_c^n(Y, A) \cong H^n(G, A).$$

The cohomology of groups is thus intrinsically present in the topological situation; hence it is not surprising that it was independently discovered in several ways: by Eilenberg-Mac Lane in 1943 ([22] and [24]), by Eckmann, with a slightly different complex ([16], 1945), by Hopf ([47], 1944) and by Freudenthal ([33a], 1946). Actually the last two authors described the homology of a group Π_1 , which can be defined as the homology $H_n(B(\Pi_1), A)$ of the complex $B(\Pi_1)$ described above.

As befits its multiple origin, there have been multiple developments of group cohomology: to Postnikov systems, to class field theory, to homological algebra and to higher dimensional interpretations. We report briefly on each.

Topologically, one immediately wishes to calculate *all* the homology and cohomology groups of a space X from *all* its homotopy groups Π_1, Π_2, \cdots . But this can't be done, because the homotopy groups alone do not suffice to

determine the space or even its homotopy type (which would be enough). To see why, it is easiest to again replace the space X by its universal covering space \tilde{X} , noting for all higher dimensions $n > 1$ the natural isomorphism $\Pi_n(X) \cong \Pi_n(\tilde{X})$: This indicates also why each $\Pi_n(X)$ is a Π_1 -module. Now we will construct two different such covering spaces Y and W with the same covering group Π_1 . Take Π_1 to be the free abelian group F on three generators $x, y,$ and $z,$ realized as a group of translations of Euclidean three-space \mathbf{R}^3 with $x, y,$ and z the unit translations along the three axes of \mathbf{R}^3 . Form a space Y by attaching a 2-sphere to each integral lattice point of \mathbf{R}^3 , and form W by removing from \mathbf{R}^3 the interior of each cube of the rectangular lattice (in other words, W is the set of all the points of \mathbf{R}^3 with at least one integer coordinate). The free group F acts as a group of translations on both Y and $W,$ both $\Pi_1(Y)$ and $\Pi_1(W)$ are zero, and both spaces have the same second homotopy group Π_2 —indeed, $\Pi_2(Y) \cong \Pi_2(W)$ is the free F -module on one generator (represented in Y as the sphere attached at the origin, and in W as the boundary of a lattice cube at the origin). Nevertheless they have different homology. The difference may be analyzed by forming in both spaces the boundary faces of a singular 3-simplex (or, equivalently, a singular cube) with leading edges $x, y,$ and z along the orbits of the three generators of the covering group F . This “sphere” may be pictured as a cube thus:



In $W,$ the sphere represents a nonzero element of Π_2 (the interior has been removed) while in Y the corresponding element is zero. This difference can be expressed in terms of a certain invariant k of the space X (covered by Y or W); it is a 3-dimensional cohomology class $k \in H^3(\Pi_1(X), \Pi_2(X))$. The cube pictured above records the fact that $k = 0$ for the space Y and that $k \neq 0$ for W . For any arcwise connected space X with vanishing homotopy beyond $\Pi_2,$ the homology is determined completely by the group $\Pi_1,$ the Π_1 -module Π_2 and this invariant k (see [26]). This list of three invariants $\Pi_1, \Pi_2,$ and k is the first example of a Postnikov system—with more higher homotopy groups, there are more such invariants. Put differently, the cohomology of groups provides the setting for describing invariants, like $k,$ sufficient to determine the homotopy type of spaces (more exactly, of spaces representable as CW-complexes).

As already indicated, the cohomology of groups was motivated in part by the needs of class-field theory; I vividly recall a conference at Indiana

University about 1948 at which Emil Artin pointed out to me some of the prospects. It remained for Artin's student Tate, exploiting also Hochschild's earlier treatment of local class field theory, to show decisively [93] how the higher cohomology groups could be used effectively in class field theory. This is the starting point of Galois cohomology (see for example Serre [88]).

We have described the cohomology of a group G in terms of a standard complex; specifically, $H^n(G, A)$ is the cohomology of $\text{Hom}_G(B_n(G), A)$, where $B_n(G)$ is the specific complex of free G -modules given in (3.10). Eckmann's alternative description of the cohomology had already indicated that one might use *different* complexes to get the *same* cohomology. It is possible to describe exactly the character of these differing complexes. In the complex (3.10), $B_0(G) = \mathbf{Z}(G)$ and the homology of the complex $B(G)$ is zero except in dimension 0, where it is the abelian group \mathbf{Z} (regarded as a G -module with trivial action $x \cdot n = n$ for all $x \in G$). Now "augment" the complex $B(G)$ of (3.10) by adding at the bottom the G -module map $\varepsilon : B_0(G) \rightarrow \mathbf{Z}$ which sends each x in G to 1. This augmented complex may then be compared to any other complex

$$(3.15) \quad \mathbf{Z} \leftarrow P_0 \leftarrow P_1 \leftarrow \cdots$$

starting with the G -module \mathbf{Z} , consisting of projective G -modules P_n and having zero homology. The comparison shows that $\text{Hom}_G(B, A)$ and $\text{Hom}_G(P, A)$ have the same cohomology, thus showing that the cohomology of G with coefficients A may be calculated from any such P —called a *projective resolution* of \mathbf{Z} . This use of projective resolutions, combined with the similar use in the Eilenberg-Mac Lane calculation of the groups Ext , was the first step toward homological algebra (Cartan-Eilenberg [10], Mac Lane [71]).

Our starting point was the interpretation of the cohomology groups $H^1(G, A)$ as a group of crossed homomorphisms and $H^2(G, A)$ as a group of group extensions. It was natural to ask for corresponding interpretations in higher dimensions. For H^3 , Eilenberg-Mac Lane found such an interpretation, building on ideas of Baer [5] and Teichmüller [96]. Consider an extension of a group K , no longer abelian, by G ; that is, a short exact sequence of groups (K additive still)

$$(3.16) \quad 0 \rightarrow K \rightarrow E \rightarrow G \rightarrow 1.$$

Since K is a normal subgroup of E , conjugation of K by an element of E gives a homomorphism $E \rightarrow \text{Aut } K$ to the group of automorphisms of K , and thus induces on G a homomorphism $\psi : G \rightarrow \text{Aut } K / \text{In } K$, where $\text{In } K$ denotes the group of inner automorphisms of K . In case K is abelian, this is just the "action" of G on $K = A$ which we have already exploited in (3.1)–(3.3). When K is *not* abelian, the first step in studying group extensions then becomes: Given K , G , and $\psi : G \rightarrow \text{Aut } K / \text{In } K$, does there exist, as in (3.16), an extension E realizing ψ ? Baer in 1934 showed that this was not always possible. Eilenberg-Mac Lane, starting also with the observation that topologists could measure the impossibility of certain constructions by

finding an “obstruction” cocycle, proceeded to find such an obstruction to the realization of ψ . The basic description of this obstruction is straightforward: In each automorphism class $\psi(x)$, pick an automorphism $\varphi(x)$. Since ψ is a homomorphism, the product $\varphi(x)\varphi(y)$ is $\varphi(xy)$ up to an inner automorphism, which we can write as the inner automorphism $\mu(f(x, y))$ induced by some element $f(x, y) \in K$. Thus, in $\text{Aut } K$,

$$\varphi(x)\varphi(y) = \mu(f(x, y))\varphi(xy).$$

The associative law now shows that μf is essentially a “factor set”. This implies that $f(x, y)$ itself is a factor set only up to an extra term k lying in the kernel of μ , so that, modifying equation (3.2) by this extra term,

$$(3.17) \quad \varphi(x)f(y, z) + f(x, yz) = k(x, y, z) + f(xy, z) + f(x, y).$$

Now $\mu : K \rightarrow \text{Aut } K$ sends each $b \in K$ to conjugation by b , so the kernel of μ is the center C of K . This C is abelian, and G acts on C via φ . Thus k is a function $k : G \times G \times G \rightarrow C$. Straightforward calculation shows that k is a three-dimensional cocycle, and that its cohomology class in $H^3(G, C)$ is independent of the choices made in its construction. This cohomology class is exactly the obstruction to the given problem of realizing ψ by a group extension, in the sense that there is such a realization if and only if this class is 0. Moreover, any three-dimensional cohomology class in $H^3(G, C)$ can be realized in this way as the obstruction to some extension problem. These results, presented in Eilenberg-Mac Lane [29] (or in Mac Lane [71, Chapter IV]) provide a satisfactory interpretation for H^3 . They are analogous to the Teichmüller constructions for G -normal algebras, as discussed above. Moreover, this interpretation of H^3 for the cohomology of groups has been a model for similar interpretations of three-dimensional cohomologies of other sorts of algebraic systems—for Lie algebras (Hochschild [42], [43] and Mori [83]), for rings (Mac Lane [69] and Shukla [89]), and for cohomologies of other systems (Barr [6], Gerstenhaber [38]). All these cases do indicate that the topological idea of an obstruction cocycle has extensive algebraic meaning and use.

For the higher cohomology groups $H^n(G, A)$ with $n \geq 4$ attempts to get similar interpretations have not been generally successful. A more systematic new attack by J. Duskin, using his new concept of a torsor, [13], [14] does, however, offer hopeful prospects.

4. The bar construction. The cohomology groups $H^n(G, A)$ of a group G , as we have just seen, provide a way of calculating the cohomology groups of an arcwise connected space X whose only nonzero homotopy group is the fundamental group $\Pi_1(X) = G$. Corresponding questions arise for other spaces X with a single nonzero homotopy group in some higher dimension n . Indeed Hurewicz, shortly after inventing the higher homotopy groups, had proved that two arcwise connected spaces X and X' with the same n th homotopy group $\Pi_n(X) = \Pi_n(X')$ and with all other homotopy groups zero were necessarily of the same homotopy type, and hence must have the same

homology and cohomology groups in all dimensions. This qualitative result again presented the same sort of algebraic problem: To find algebraic formulas which give the homology of such a space in each dimension as a function of the abelian group Π —and of the dimension n in which Π appears.

In principle, this functorial dependence problem could be solved by analysis of the singular complex $S(X)$ of the space X in question. Singular simplices of dimension less than n could all be deformed homotopically to the degenerate simplices at the chosen base point. In dimension n , a singular simplex $T: \Delta_n \rightarrow X$ would then map all the boundary of Δ_n to the base point, and so would be essentially a map of the n -sphere S_n to X ; each such T thus represents an element y in the n th homotopy group Π . In higher dimensions $q > n$ each singular simplex $T: \Delta_q \rightarrow X$ would determine on each n -dimensional face of Δ_q an element $y \in \Pi$, and the whole simplex T could thus be replaced (in a minimal subcomplex) by the pattern of elements of Π represented by T on the various n -dimensional faces of Δ_q . Each such pattern is a Π -valued function on the n -dimensional faces; when we regard Δ_q as a simplicial set, the pattern is exactly an n -dimensional cocycle (an element of $Z^n(\Delta_q, \Pi)$). In this way it was possible to describe in terms of Π and n a chain complex with the appropriate boundary ∂ ,

$$(4.1) \quad K(\Pi, n): \cdots \longleftarrow K_q(\Pi, n) \xleftarrow{\partial} K_{q+1}(\Pi, n) \longleftarrow \cdots,$$

with $\partial\partial=0$ as usual, in which the group K_q of q -dimensional chains is exactly the free abelian group generated by the set $Z^n(\Delta_q, \Pi)$. This complex $K(\Pi, n)$ is thus a wholly algebraic object depending only on Π and on n and having the desired homology. This complex is also very cumbersome. The problem lay in simplifying this complex to the point where its homology and cohomology groups could be explicitly determined.

The simplification proved to be possible thanks to long and elaborate calculations of many special cases. (As customary, the extent of these calculations was largely covered up in the resulting papers by Eilenberg-Mac Lane [28], [29], and [30].) These calculations ultimately revealed two leading principles:

- (a) $K(\Pi, n+1)$ should be expressed in terms of $K(\Pi, n)$;
- (b) $K(\Pi, n)$ has a multiplicative structure.

The first allowed us to start with $K(\Pi, 0)$, which is just the integral group ring $\mathbf{Z}(\Pi)$, and with $K(\Pi, 1)$, which is of course exactly the complex $B(\Pi)$ already discussed for this case—that of the cohomology of an (abelian) group Π , realized as the fundamental group of a space.

A multiplicative structure on a complex K is a rule under which a p -chain c and a q -chain c' have as product a $p+q$ chain cc' , so as to yield a bilinear (and associative) map

$$(4.2) \quad K_p(\Pi, n) \otimes K_q(\Pi, n) \rightarrow K_{p+q}(\Pi, n)$$

which satisfies the usual (Leibniz) formula for the boundary of a product

$$(4.3) \quad \partial(cc') = (\partial c)c' + (-1)^p c(\partial c').$$

Generally, a chain complex with such a product is called a *differential graded ring*. In the present case, the product arose first in the simplified expression of $K(\Pi, 2)$ in terms of $K(\Pi, 1)$; specifically, the boundary formula for $K(\Pi, 2)$ required in $B(\Pi)=K(\Pi, 1)$ a product given by “shuffles.” For elements $x_1, \dots, x_p, y_1, \dots, y_q$ of Π this product in $B(\Pi)$ was explicitly

$$(4.4) \quad [x_1, \dots, x_p][y_1, \dots, y_q] = \sum \pm [z_1, \dots, z_{p+q}],$$

where each list z_1, \dots, z_{p+q} on the right is obtained by “shuffling” the letters x_1, \dots, x_p (preserving their order) through y_1, \dots, y_q , where the sum is taken over all such shuffles, and where the sign of each term is the sign of the shuffle, regarded as a permutation of $p+q$. From this purely algebraic and formal description of the product it is possible to prove the associativity of the product as well as the boundary formula (4.3). However, it turned out that this product formula had a geometric—or at least a simplicial—origin, based on the fact that the cell $[x_1, \dots, x_p]$ really stands for a pattern of elements x_i of Π on the edges of a p -simplex, and hence a 1-dimensional cocycle in $Z^1(\Delta_p, \Pi)$. Now the cartesian product $\Delta_p \times \Delta_q$ of two simplices is a prism which can be triangulated into simplices in a standard way—think of a square divided into two triangles by a diagonal, or of a triangular prism $\Delta_2 \times \Delta_1$ divided into three tetrahedra. The general picture is this: Let the simplex Δ_p have vertices (a_0, \dots, a_p) while Δ_q has vertices (b_0, \dots, b_q) . Then $\Delta_p \times \Delta_q$, a prism with vertices $v_k=(a_i, b_j)$, is triangulated by those $p+q$ simplices (v_0, \dots, v_{p+q}) which are lists from among the given vertices and taken in some increasing order, so that a vertex $v_k=(a_{i-1}, b_{j-1})$ in such a list is to be followed either by (a_i, b_{j-1}) or by (a_{i-1}, b_j) —in the first case moving along the edge labelled x_i from a_{i-1} to a_i , in the second case moving along the edge y_j . Hence each such list of $p+q$ vertices is given by successive choices of x 's (in order) or y 's (in order); that is, exactly by a shuffle of the x 's through the y 's. It is this simplicial structure which is the real source of the “shuffle” product structure in the complexes $K(\Pi, n)$. The algebraic formula has a geometric origin. (See Mac Lane [71, VIII, 8].)

This observation also emphasizes the presence of a simplicial structure in the complex $K(\Pi, n)$. For each generator (or pattern) c of $K_q(\Pi, n)$ is a cocycle $c \in Z^n(\Delta_q, \Pi)$, while the boundary of c is defined as an alternating sum of faces

$$(4.5) \quad \partial c = F_0 c - F_1 c + \dots + (-1)^q F_q c$$

where $F_i c$ designates the cocycle c restricted to the i th face of Δ_q . There are corresponding “degeneracy” operations; the simplicial map $D^i: \Delta_{q+1} \rightarrow \Delta_q$ collapsing the successive vertices b_j and b_{j+1} of Δ_{q+1} induces a map $D_j: Z^n(\Delta_q, \Pi) \rightarrow Z^n(\Delta_{q+1}, \Pi)$. Together, these operations F_i and D_j satisfy certain identities which make the complex $K(\Pi, n)$ into what we then called an F - D complex—now called a simplicial abelian group. The extensive manipulation of these face and degeneracy operators needed for the calculations of $K(\Pi, n)$ were one of the sources which led at that time to the Eilenberg-Zilber definition [31] of “complete semisimplicial complexes.”

These complexes (now called *simplicial sets*, Mac Lane [71]) are exactly the minimal structures in which face and degeneracy operations can appear. They were used initially [27], [31] to prove the Eilenberg-Zilber theorem. This is the theorem which, applied to the singular complex of the product of two spaces X and Y , yields a chain equivalence of singular complexes

$$(4.6) \quad S(X \times Y) \rightleftarrows S(X) \otimes S(Y).$$

With the Künneth theorem on the homology of a tensor product of complexes (Kan [48], May [81], Lamotke [56]) as well as a systematic Moreover, in this Eilenberg-Zilber theorem, the map from right to left in (4.6) is essentially based on shuffles, because singular simplices $T: \Delta_p \rightarrow X$ and $U: \Delta_q \rightarrow Y$ yield $T \times U: \Delta_p \times \Delta_q \rightarrow X \times Y$ which by subdivision (shuffles) produces a sum of $(p+q)$ -dimensional singular simplices of $X \times Y$.

From this notion of a simplicial set, D. M. Kan developed the possibility of a wholly algebraic formulation of homotopy theory in terms of his Kan complexes (Kan [48], May [81], Lamotke [56]) as well as a systematic view of topology, as in Gabriel-Zisman [36]). This is another case in which geometrical ideas led to algebraic developments—in this case turning back to geometry again in the various uses of simplicial sets in recent differential topology.

After this exploration of the geometric reasons for the presence of a product structure in the complex $K(\Pi, n)$ we return to the original problem of constructing from $K(\Pi, n)$ an equivalent of $K(\Pi, n+1)$. This will be done in terms of the bar construction B , which applies to $K(\Pi, n)$ or, for that matter, to any differential graded ring U with an augmentation $\varepsilon: U \rightarrow \mathbf{Z}$. Here an *augmentation* means a morphism of differential graded rings, and the ring \mathbf{Z} of integers is regarded as a differential graded ring with zero differential and all elements taken to have degree 0. The *bar construction* on U is now the additive group

$$(4.7) \quad B(U) = \sum_{p=0}^{\infty} U^p, \quad U^p = U \otimes \cdots \otimes U \quad (p\text{-factors})$$

with $U^0 = \mathbf{Z}$, with elements written with “bars” in place of \otimes as

$$(4.8) \quad u_1 \otimes \cdots \otimes u_p = [u_1 | \cdots | u_p], \quad u_i \in U,$$

and graded by degrees defined from the degrees in U by

$$(4.9) \quad \deg[u_1 | \cdots | u_p] = p + \deg u_1 + \cdots + \deg u_p.$$

This graded group is made into a chain complex (i.e., a “differential graded group”) by introducing a boundary operator ∂ which is the sum of two terms, $\partial = \partial' + \partial''$. The first summand ∂' is almost exactly the boundary operator of $B(\Pi_1)$,

$$(4.10) \quad \begin{aligned} \partial[u_1 | \cdots | u_p] &= \varepsilon(u_1)[u_2 | \cdots | u_p] + \sum_{i=1}^{p-1} (-1)^{\varepsilon_i} [u_1 | \cdots | u_i u_{i+1} | \cdots | u_p] \\ &\quad + (-1)^{\varepsilon_p} [u_1 | \cdots | u_{p-1}] \varepsilon(u_p), \end{aligned}$$

with signs determined by $e_i = \text{deg}[u_1 | \cdots | u_i]$. The second summand depends on the boundary already present in U ,

$$(4.11) \quad \partial^n[u_1 | \cdots | u_p] = \sum_{i=1}^p (-1)^{e_{i-1}} [u_1 | \cdots | u_{i-1} | \partial u_i | \cdots | u_p].$$

It is immediate to verify that $\partial\partial=0$. The essential result is

THEOREM 1 (EILENBERG-MAC LANE [28]). *For any abelian group Π there is a chain equivalence*

$$(4.12) \quad g : B(K(\Pi, n)) \rightarrow K(\Pi, n+1).$$

The original proof involved an explicit formula for g , a normalization of the bar construction B and a description of the simplicial set $K(\Pi, n+1)$ directly in terms of $K(\Pi, n)$ by a straightforward but elaborate simplicial construction W (one that has subsequently been used by John Moore and others): $K(\Pi, n+1) = W(K(\Pi, n))$. Since chain equivalent complexes have the same homology and cohomology, this result does replace a computation of the homology for $K(\Pi, n+1)$ by that for a substantially simpler complex $B(K(\Pi, n))$.

For really effective calculations one wishes to iterate this reduction. This involves applying a bar construction to a bar construction, and hence requires the use of a suitable product in $B(U)$. To do this it is essential to observe that the product in each $K(\Pi, n)$ is commutative (or should one say “skew commutative”?) in the sense that always

$$cc' = (-1)^{(\text{deg } c)(\text{deg } c')} c'c, \quad c, c' \in K(\Pi, n).$$

For example, the group ring $K(\Pi, 0) = \mathbf{Z}(\Pi)$ is commutative because the group Π is assumed to be abelian, while the shuffle product (4.4) used in describing $K(\Pi, 1) = B(\Pi)$ is commutative in virtue of elementary properties of permutations. More generally, if the differential graded ring U is commutative, the bar construction $B(U)$ becomes a differential graded ring under the same type of shuffle product,

$$[u_1 | \cdots | u_p][v_1 | \cdots | v_q] = \sum \pm [w_1 | \cdots | w_{p+q}],$$

where the sum on the right is taken over all shuffles w_k of the u 's through the v 's and the exponent of the sign $(-1)^-$ is the sum of all products $(\text{deg } u_i)(\text{deg } v_j)$ for those v_j which have been shuffled in front of u_i . With this shuffle product, one proves that the chain equivalence g of the main theorem preserves products. Hence the reduction to the bar construction can be iterated to prove the

COROLLARY. *For any abelian group Π there is a chain equivalence $B^n(\mathbf{Z}(\Pi)) \rightarrow K(\Pi, n)$.*

By these means the original and elaborate complex $K(\Pi, n)$ is reduced to a much simpler complex $B^n(\mathbf{Z}(\Pi))$ on which effective calculation of homology and cohomology of $K(\Pi, n)$ is possible. This of course gives the

homology of the corresponding spaces of type $K(\Pi, N)$ —the arcwise connected spaces X with $\Pi_n(X) = \Pi$ and all other homotopy groups trivial. They are now commonly called “Eilenberg-Mac Lane spaces” and used as “building blocks” in Postnikov systems and elsewhere in topology.

We record the results of a few calculations by these means, writing $H_n(\Pi, n)$ for $H_n(K(\Pi, n))$:

$$\begin{aligned} H_0(\Pi, n) &= \mathbf{Z}, & H_q(\Pi, n) &= 0, & 0 < q < n; \\ H_n(\Pi, n) &= \Pi, \end{aligned}$$

which results from the Hurewicz isomorphism: The first nonvanishing homotopy group (made abelian) is isomorphic to the corresponding homology group;

$$\begin{aligned} H_{n+1}(\Pi, n) &= 0, \\ H_4(\Pi, 2) &= \Gamma(\Pi), & H_{n+2}(\Pi, n) &= \Pi/2\Pi, & n > 2. \end{aligned}$$

Here $\Gamma(\Pi)$, a group discovered by J. H. C. Whitehead [101], is the group generated by symbols $\gamma(x)$, one for each $x \in \Pi$, and subject to the relations $\gamma(-x) = \gamma(x)$ and

$$\gamma(x + y + z) - \gamma(x + y) - \gamma(x + z) - \gamma(y + z) + \gamma(x) + \gamma(y) + \gamma(z) = 0$$

for all x, y , and z . These relations state that γ is a *quadratic* function (or a quadratic form). They have the consequence that Γ , as a function of Π , is not additive but quadratic; for the direct sum of two groups Π_1 and Π_2 , $\Gamma(\Pi_1 + \Pi_2)$ is not $\Gamma(\Pi_1) + \Gamma(\Pi_2)$; instead there is a “deviation from additivity” given by the formula

$$\Gamma(\Pi_1 + \Pi_2) \cong \Gamma(\Pi_1) + \Gamma(\Pi_2) + \Pi_1 \otimes \Pi_2.$$

In other words, the geometric study of the spaces $K(\Pi, n)$ leads explicitly to the construction of functors like $H_4(\Pi, 2)$ which are quadratic (and similarly to functors of higher degrees).

The list continues as

$$H_6(\Pi, 3) = {}_2\Pi + \Pi \wedge \Pi, \quad H_{n+3}(\Pi, n) = {}_2\Pi, \quad n > 3,$$

where ${}_2\Pi$ denotes the subgroup of all elements of order 2 in Π and $\Pi \wedge \Pi$ is the exterior square of Π , generated by symbols $x \wedge y$, for $x, y \in \Pi$, which are bilinear and alternating ($x \wedge x = 0$). This functor $\Pi \wedge \Pi$ is another quadratic functor; there is a similar (and more complex) formula for $H_5(\Pi, 2)$. Finally, for $n > 5$,

$$H_8(\Pi, 4) = \Gamma(\Pi) + \Pi/3\Pi, \quad H_{n+4}(\Pi, n) = \Pi/2\Pi + \Pi/3\Pi.$$

But let us return from these results of calculation to the two principal observations which made these calculations possible: The construction of $K(\Pi, n+1)$ out of $K(\Pi, n)$ and the presence of a (commutative) product in $K(\Pi, n)$. In the Eilenberg-Mac Lane calculations these were first discovered empirically, as means to make the calculations go. They are actually deeply

rooted in the geometry, and so really exhibit our main theme of the influence of geometry upon algebra. For let X be an arcwise connected space of type $K(\Pi, n+1)$, choose a base point x_0 in X and construct the path space $E(X)$ whose points are all the paths starting at x_0 (and topologize $E(X)$ by the compact-open topology). Then the space $E(X)$ is contractible, the function $p : E(X) \rightarrow X$ sending each path to its endpoint in X is a (Serre) fibration and the fiber $p^{-1}(x_0)$ over the base point is exactly the space $\Omega(X)$ of all closed loops (at x_0) in X . A standard argument in homotopy theory shows that each homotopy group $\Pi_k(\Omega(X))$ is isomorphic to $\Pi_{k+1}(X)$, hence in our case $\Omega(X)$ has only one homotopy group, the group Π in dimension n . In other words, the loop space on a space $K(\Pi, n+1)$ gives a space $K(\Pi, n)$. Put differently, the bar construction is a type of "inverse" construction to the loop space. The parallel between geometry and algebra should read

$$\begin{array}{lcl}
 \text{loop space} & \Omega(X) & K(\Pi, n) \\
 & \downarrow & \downarrow \\
 \text{path space} & E(X) & E(K(\Pi, n)) \\
 & \downarrow & \downarrow \\
 K(\Pi, n+1) = & X & B(K(\Pi, n)) \xrightarrow{\sim} K(\Pi, n+1).
 \end{array}$$

The complex on the bottom at the right is the bar construction, but hitherto the algebraic treatment missed the middle complex E on the right. On the left, the loop-space $\Omega(X)$ has a multiplication (the evident product of two loops) and is an H -space under this multiplication; moreover this H -space $\Omega(X)$ acts on the path space $E(X)$. Hence on the right E should be contractible and the differential graded ring $K(\Pi, n)$ should act on E . In the case $n=1$ we used exactly such an E in forming the resolution (3.11). In general, there is always a suitable chain complex E to fit these properties in this diagram. Indeed, for each augmented differential graded ring we may form $E(U)$ as the sum $\sum U \otimes U^p$, with suitable grading and differential, making U act on $E(U)$ via the first factor U in $U \otimes U^p$. (The exact formulas appear in Mac Lane [71], where E is called the *bar* construction and $B = \bar{B}$ the *reduced bar* construction.) Thus the parallel between geometry and algebra is complete; moreover, the product (of loops) present in $\Omega(X)$ does exactly match the shuffle product found algebraically in $K(\Pi, n)$ —as demonstrated, after the fact, by Mac Lane in [68].

This parallel between geometry and algebra can also be used to facilitate the calculations. It was first clearly observed by H. Cartan, who continued in his 1954/55 seminar [9] to show that the bar construction (and the acyclic bar construction E) could be supplemented by many other types of small but similarly acyclic constructions. He then used these constructions to obtain a complete calculation of $H_q(\Pi, n, \mathbf{Z}_p)$, with p a prime, with the results expressed in terms of suitable homology operations. He also obtained a description of $H_q(\Pi, n; \mathbf{Z})$; much more recently, Ross Hamsher [41] and Gerald J. Decker [12] have shown how this description can be further developed to yield a whole family of invariant descriptions of homology groups $H_q(\Pi, n)$ like the few listed in the formulas above.

The bar constructions $B(G)$ for a group and $B(U)$ for a differential graded ring are algebraic entities, inspired by topological considerations. In particular, $B(G)$ is, by the parallel above, exactly the analog of the classifying space of a group G . This analogy led many authors to replace the algebraic bar construction $B(G)$ by a suitably topological construction of a space (Rothenberg-Steenrod [87], Steenrod [91]). The most definitive formulation is that of Milgram [82]. Since the formulas for $B(G)$ do not use the inverse, but just the multiplication and unit in G , Milgram replaces the group G by a topological monoid; more exactly, by a monoid in the category of compactly generated spaces; from this G he constructs a space $B(G)$. For each cell $[x_1, \dots, x_p]$ in the algebraic bar construction he takes a copy of the standard p -simplex, forms the sum of all these products,

$$B(G) = \sum_{p, x_i} [x_i, \dots, x_p] \times \Delta_p / \sim,$$

and then identifies suitable faces and degeneracies. Indeed, we already have noted that the faces and degeneracies act on the cells $[x_i, \dots, x_p]$; they act in the opposite or dual sense on the simplices Δ_p 's. As a result the identifications required can be expressed conceptually by regarding both

$$p \mapsto \underbrace{G \times \dots \times G}_p \quad \text{and} \quad p \mapsto \Delta_p$$

as functors, of opposite variance, on the basic simplicial category Δ . The geometric bar construction is thus expressed simply as a *coend* in the categorical sense (*Categories work* [16], Chapter IX; the coend, indicated by the integral sign, serves to make exactly the desired identifications).

$$B(G) = \int^p G^p \times \Delta_p.$$

At this central point, algebra and topology are thoroughly mixed! (For details, see Mac Lane [75].)

5. Coherence and canonical maps. Stimulated by considerations in topology, Norman Steenrod raised the question: When is there a canonical map between two specified *formal* combinations of modules? He was considering (the category of) all modules over a commutative ring, and the combinations of such modules formed by applying the functors Hom and \otimes . In order to describe a suitable meaning for "canonical maps" between such combinations we must describe these circumstances in more detail.

First consider only a tensor product \otimes . A (symmetric) *monoidal category* M is one equipped with a bifunctor $\otimes : M \times M \rightarrow M$ which is associative and commutative up to isomorphism, in the sense that there are isomorphisms

$$a : A \otimes (B \otimes C) \rightarrow (A \otimes B) \otimes C, \quad c : A \otimes B \rightarrow B \otimes A$$

which are natural (in the objects A, B, C of M) and which satisfy two additional conditions, expressed by the commutativity of the (pentagonal

and hexagonal) diagrams

$$\begin{array}{ccc}
 A \otimes [B \otimes (C \otimes D)] & \xrightarrow{a} & (A \otimes B) \otimes (C \otimes D) \\
 \downarrow 1 \otimes a & & \searrow a \\
 A \otimes [(B \otimes C) \otimes D] & \xrightarrow{a} & [A \otimes (B \otimes C)] \otimes D \\
 & & \nearrow a \otimes 1
 \end{array}$$

$$\begin{array}{ccccc}
 A \otimes (B \otimes C) & \xrightarrow{a} & (A \otimes B) \otimes C & \xrightarrow{c} & C \otimes (A \otimes B) \\
 \downarrow 1 \otimes c & & & & \downarrow a \\
 A \otimes (C \otimes B) & \xrightarrow{a} & (A \otimes C) \otimes B & \xrightarrow{c \otimes 1} & (C \otimes A) \otimes B
 \end{array}$$

the first of these asserts that the two ways of reassociating an iterated product of four factors are equal, while the second states that the two possible ways of “permuting” a factor C past both A and B are equal. In addition, a symmetric monoidal category is equipped with an object I and an isomorphism $b : A \otimes I \rightarrow A$, natural in A , which renders commutative the (triangular) diagram

$$\begin{array}{ccc}
 A \otimes (B \otimes I) & \xrightarrow{a} & (A \otimes B) \otimes I \xrightarrow{b} A \otimes B \\
 & & \xrightarrow{1 \otimes b}
 \end{array}$$

Clearly the modules (or the graded modules, or the differential graded modules) over a commutative ring R , with I the module R and with a, b , and c the usual canonical maps, do then form a monoidal category. A “canonical” map in such a category may now be described (more formally below) as any map formed by tensor products and compositions of these three maps a, b , and c . The three diagrammatic conditions on these maps are a way of insuring that we have used the “right” maps (for example, for modules again, *not* the “wrong” isomorphism $\bar{a} : A \otimes (B \otimes C) \rightarrow (A \otimes B) \otimes C$ given as

$$x \otimes (y \otimes z) \mapsto -(x \otimes y) \otimes z, \quad x \in A, y \in B, z \in C;$$

this map evidently does not make the pentagonal diagram above commutative. Put differently, these conditions on a, b , and c allow us to handle with precision the familiar identifications: The usual identification of $A \otimes (B \otimes C)$ with the logically distinct triple product $(A \otimes B) \otimes C$ can be done via the given isomorphism a . The pentagonal diagram insures that this provides exactly *one* isomorphism serving to identify two quadruple tensor products; the first coherence theorems (independently found by Mac Lane [73] and Epstein [33]) assert the same for n -fold products, for all n . Moreover, this applies to many other instances of monoidal categories, such as the category of sets with \otimes taken to be cartesian product. Here again it explains the process of identifying two formally different iterated n -fold cartesian products.

Now we supplement the tensor product with a functor Hom , defining a *closed category* to be a monoidal category M with an added bifunctor $\text{Hom} : M^{\text{op}} \times M \rightarrow M$ (the exponent “op” indicates that this functor Hom is

contravariant in its first argument) such that each (covariant) functor $\text{Hom}(A, -) : M \rightarrow M$ is right adjoint to the functor “tensor product by A ”, $- \otimes A : M \rightarrow M$. This statement means, as usual, that there is an isomorphism

$$(5.1) \quad \pi : \text{hom}(B \otimes A, C) \cong \text{hom}(B, \text{Hom}(A, C))$$

natural in B and C , where the “outside” hom is a set; specifically, $\text{hom}(B, D)$ the set of all morphisms from B to D in the category M . This formulation emphasizes the sense in which a closed category is one with an *internal* hom-functor Hom , with values in M .

For our purposes of describing canonical maps it is necessary to recall that an adjunction isomorphism, such as π above, can be completely described by its unit and its counit. To get the *unit* we set $C = B \otimes A$ in the formula above, take the identity $B \otimes A \rightarrow B \otimes A$ in the left hand hom -set and its image under the adjunction isomorphism, giving some $d : B \rightarrow \text{Hom}(A, B \otimes A)$ (for modules, d is the map which sends $y \in B$ to the homomorphism $x \mapsto y \otimes x$, all $x \in A$). Similarly, to get the *counit*, set $B = \text{Hom}(A, C)$, take the identity of $\text{Hom}(A, C)$ and its image on the left in the adjunction, which is a map $e : \text{Hom}(A, C) \otimes A \rightarrow C$ called “evaluation.” For the category of modules, it is just the familiar map $f \otimes x \mapsto f(x)$ which “evaluates” a morphism $f : A \rightarrow C$ in $\text{Hom}(A, C)$ on an element $x \in A$.

A closed category may thus be described as a monoidal category M together with an added bifunctor $\text{Hom} : M^{\text{op}} \times M \rightarrow M$ and two families of maps

$$d : B \rightarrow \text{Hom}(A, B \otimes A), \quad e : \text{Hom}(A, C) \otimes A \rightarrow C,$$

natural in B and in C , respectively, which render commutative the diagrams

$$\begin{array}{ccc} \text{Hom}(A, C) & \xrightarrow{d} & \text{Hom}(A, \text{Hom}(A, C) \otimes A) & B \otimes A & \xrightarrow{d \otimes 1} & \text{Hom}(A, B \otimes A) \otimes A \\ & \searrow 1 & \downarrow \text{Hom}(1, e) & & \searrow 1 & \downarrow e \\ & & \text{Hom}(A, C) & & & B \otimes A \end{array}$$

and, for every arrow $f : A \rightarrow D$, the diagrams

$$\begin{array}{ccccc} B & \xrightarrow{d} & \text{Hom}(A, B \otimes A) & \text{Hom}(D, C) \otimes A & \xrightarrow{1 \otimes f} & \text{Hom}(D, C) \otimes D \\ \downarrow d & & \downarrow \text{Hom}(1, 1 \otimes f) & \downarrow \text{Hom}(f, 1) \otimes 1 & & \downarrow e \\ \text{Hom}(D, B \otimes D) & \xrightarrow{\text{Hom}(f, 1)} & \text{Hom}(A, B \otimes D) & \text{Hom}(A, C) \otimes A & \xrightarrow{e} & C. \end{array}$$

The first two diagrams are the “triangular identities” for unit d and counit e ; they always suffice for the descriptions of the adjunction in terms of unit and counit—see Mac Lane, *Categories work* [76]. The other two diagrams state that the unit d is “diagonally natural” in A and that e is too. The modules over a commutative ring do form, in the familiar way, a closed category, as do sets, with \otimes taken to be cartesian product.

Now we can explicate Steenrod’s question about canonical maps between formal combinations of modules. The formal combinations in question will

be called *words*. Assume given an infinite stock of letters A, B, C, \dots and a symbol I , not here regarded as a “letter.” Then I and each of the letters is called a word, covariant in that letter. If W and W' are two words in *disjoint* (finite) sets L and L' of letters then both $W \otimes W'$ and $\text{Hom}(W, W')$ are words in the set $L + L'$ of letters. Each letter A in W has the opposite variance (\pm) when regarded as a letter in $\text{Hom}(W, W')$, while the other letters retain their variance. The words so constructed are all words. Clearly each word W in n distinct letters determines for each particular closed category M a functor W_M of n variables in M , with values in M , and with the variance of the functor in each of its variables determined by the variance of the corresponding letter in the word W . (All this formality about letters is necessary to avoid misstatements, as will presently appear.)

The canonical maps f are intended to be those generated by the maps a, b, c, d , and e defining the structure of the closed category M , but regarded just formally, as maps $f: W \rightarrow W'$ between words. However it is convenient to so arrange matters that the domain and codomain words W and W' have no letters in common. For example, we will rewrite the map $c: A \otimes B \rightarrow B \otimes A$ as $C: A \otimes B \rightarrow B^* \otimes A^*$, to involve four different letters A, B, A^*, B^* together with an involution $A \leftrightarrow A^*, B \leftrightarrow B^*$ on this set of four letters. More generally, we consider any *instance* of c formed by taking any two words V and W , copying them as V^* and W^* in different (but corresponding) letters, and using $c_{V,W}: V \otimes W \rightarrow W^* \otimes V^*$ as a canonical “map” in the letters in the disjoint union $V + W + V^* + W^*$ of four sets of letters. Moreover, this process determines an involution, called the *graph* of $c_{V,W}$, on the total set of letters, written $\text{graph}(c_{V,W}): V \otimes W \rightarrow W^* \otimes V^*$. One defines similarly the *instances* of a, b, c, d , and e and their graphs, noting however that in the graph of $e: \text{Hom}(V, W) \otimes V^* \rightarrow W^*$ two letters corresponding under the involution may appear on the same side of the arrow—and similarly for d . Canonical maps are now defined by recursion, requiring that all instances of a, b, c, d , and e are canonical, that if $f: V \rightarrow W$ and $g: V' \rightarrow W'$ are canonical, so are $f \otimes g: V \otimes V' \rightarrow W \otimes W'$ and $\text{Hom}(f, g): \text{Hom}(V', W) \rightarrow \text{Hom}(V, W')$ and that the composite of two canonical maps is canonical. Thus each canonical map $f: V \rightarrow W$ appears as a word built up by applying \otimes, Hom , and composition to letters standing for instances of a, b, c, d , and e ; it is easily possible to regard the canonical maps as arrows in a suitably constructed category with objects the words V and W , and this is indeed so done in the original treatment by Kelly-Mac Lane [51].

Each canonical $f: W \rightarrow W'$, by this construction, has a *graph*, which is an involution on the set of letters involved; specifically, if W is a word in letters L and W' a word in the (necessarily disjoint) set of letters L' , then $\text{graph}(f)$ is an involution on the set $L + L'$. In actually writing down diagrams, this graph is customarily indicated by the labels of the variables, corresponding letters being labelled alike. Formally, we can regard the graph as an arrow $W \rightarrow W'$ in the category of graphs, with objects all words and arrows all (possible) graphs; it turns out that this category of graphs is itself a closed category.

Now in $e : \text{Hom}(A, B) \otimes A \rightarrow B$, and likewise in the maps d , a , etc., one may note that corresponding letters have the *same* variance when they occur on *opposite* sides of the arrow, and otherwise have opposite variance. This variance rule may be readily captured in the following definition. Given words W and W' in disjoint sets L and L' of letters, a graph $\varphi : W \rightarrow W'$ is an involution of the disjoint sum $L + L'$ such that corresponding letters would have opposite variance in the (additional) word $\text{Hom}(W, W')$. This involution of the letters is needed in order to describe the naturality of our intended transformations, and the notion of graph is due to Eilenberg and Kelly in the pioneering study [19] of diagonal naturality. Moreover, these authors showed how to compose graphs, in a way which would match with the composition of the corresponding natural transformations. A small adjustment is required for $W \rightarrow W' \rightarrow W''$ in the cases where the letters of W and W'' are not disjoint; with this adjustment, the graphs form a category which is actually a closed category (Kelly-Mac Lane [51]).

Steenrod's original question now has the specific form: Given a graph $\varphi : W \rightarrow W'$ (with words W and W' on disjoint sets of letters), does there exist a canonical arrow $f : W \rightarrow W'$ with this graph? If so, how many such arrows are there? For example, are there canonical maps

$$(5.2) \quad A \rightarrow \text{Hom}(\text{Hom}(A, B), B), \quad \text{Hom}(\text{Hom}(A, B), A) \rightarrow B?$$

Again, there should be a canonical map

$$(5.3) \quad \text{Hom}(B, C) \otimes \text{Hom}(A, B) \rightarrow \text{Hom}(A, C);$$

namely, composition. To obtain it, one first applies an instance of d , to get

$$\text{Hom}(B, C) \otimes \text{Hom}(A, B) \rightarrow \text{Hom}(A, \text{Hom}(B, C) \otimes \text{Hom}(A, B) \otimes A).$$

The inside term on the right then goes to C by two evaluations,

$$\text{Hom}(B, C) \otimes \text{Hom}(A, B) \otimes A \rightarrow \text{Hom}(B, C) \otimes B \rightarrow C.$$

All this is just a formulation of the usual definition of the composite $g \circ f$ as $(gf)(a) = g(fa)$, via two evaluations on an element $a \in A$. It does also exhibit the essential point of difficulty: To get a canonical map $W \rightarrow W'$ as in (5.2) one may need to go through a composite which involves intermediate words longer—possibly much longer—than W or W' . Therefore it is not obvious how to make a finite test for the existence of a canonical map $f : W \rightarrow W'$.

Now a formally similar question had long since arisen in logic, more exactly in proof theory as developed by G. Gentzen [37]. One deals not with words but with formulas. The letters stand for propositions, $p \otimes q$ is replaced by $p \& q$, standing for “ p and q ”, while $\text{Hom}(p, q)$ becomes $p \supset q$, standing for p implies q . One asks not for canonical maps built up from constituent instances, but for proofs built up in the accustomed way from constituent inferences. One of these is the famous inference *modus ponens* (also called “cut”): Given p and $p \supset q$, one may infer q . The crucial difficulty is the same: To find all proofs leading to q , one may go through premises involving intermediate formulas p much longer than q . Now in this case Gentzen had

found a way around, recorded in what he called a *cut elimination* theorem. Lambek [54], [55] in studying the corresponding categorical problem had the insight to see that this method from logic might be carried over to closed categories. Kelly and Mac Lane, studying Lambek’s work, could not understand some of the details (for good reason; it turns out that they are wrong, and there seems to be no easy direct repair), so recast the problem to arrive at a solution.

This method first describes two special classes of canonical maps called respectively the “central” and the “constructible” maps. The central maps are to be the canonical maps for the monoidal structure; they are defined formally by recursion: All instances of a , b , and c are central; if f and g are central, so is $f \otimes g$, and the composite of two central maps is central. For any central $f : V \rightarrow W$ any pair of corresponding letters in the graph of f are on opposite sides of the arrow \rightarrow , simply because this is evidently the case for the basic central maps a , b , and c . Any word V can be written as an iterated product $V = P_1 \otimes \cdots \otimes P_m$, with suitable parentheses, of prime factors P_i , where a word is *prime* if it is not a tensor product; that is, if it is I , a letter, or a word of the form $\text{Hom}(U, W)$. Intuitively, a central map $f : V \rightarrow W$ does not change the prime factors P_i of V (except for I ’s), but amounts only to addition or removal of factors I , rearrangement of parentheses, and permutation of factors. Thus the centrals can be described exactly:

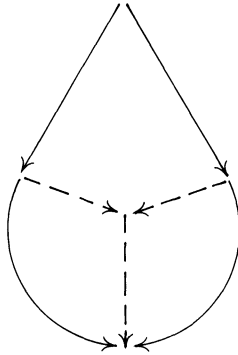
THEOREM 5.1. *If V and W are words not involving Hom , and with no letters in common, each graph $\varphi : V \rightarrow W$ which is a bijection from letters of V to letters of W is the graph of exactly one central map $f : V \rightarrow W$.*

Here “exactly one” uses the appropriate notion of equality, that two central maps $f, g : V \rightarrow W$ are called equal when they are formally equal on the grounds of the axioms for a monoidal category. Hence, this theorem really amounts to the statement that, in a monoidal category, every diagram of canonical maps a, b, c must commute. It was proved by Mac Lane [73] and independently by Epstein. Mac Lane’s then proof of the theorem was quite correct, but his statement was deficient. He did not discuss words at all, so that he considered not the formal word $A \otimes B$ in two letters A and B , but only the functor $M \times M \rightarrow M$ which it determines (on any monoidal category M). As a result, his formulation, carefully examined, confused the word $A \otimes B$ with the word $B \otimes A$ and his theorem would thus assert that the two central maps $c : A \otimes B \rightarrow B \otimes A$ and $1 : A \otimes B \rightarrow A \otimes B$ are equal. Kelly subsequently set this all straight [50].

The analog of Theorem 5.1 for associativity only is simpler: Any diagram of tensor products of instances of the associativity isomorphism a must commute, given only that a is natural in its three arguments and that it satisfies the pentagonal condition. In this case the letters involved in the words stay in the same order under any associativity, so the refinement involved in using graphs is needless. But the theorem does have the following two interesting aspects.

First, its proof is a simple paraphrase of the elementary proof of the

Jordan-Hölder theorem for finite groups: Given two centrals $V \rightarrow W$, represented as composites of instances of a , regard these chains as if they were composition series and show that they are equal by a suitable induction



by fitting in a polygon deforming the one first step to the other. For Jordan-Hölder, that polygon at the top is a diamond, while for the coherence theorem it is either a diamond (a use of naturality) or a pentagon; the details appear in [73] or in *Categories work* [76, p. 163].

Second, the coherence theorem, though legally a theorem of algebra, really came up in topology first. It was probably found first by Stasheff in his study of H -spaces which were not associative, but only associative up to a homotopy: $x(yz) \sim (xy)z$. Following pregnant ideas of Dyer and Lashof he considered also cases where this associativity homotopy, call it a , satisfies a further condition, which was a pentagonal condition *up to homotopy*. These ideas, represented geometrically in all higher dimensions, implicitly contain the above coherence theorem (and more), though I confess that I would not have discovered this from just inspecting Stasheff's paper [90]; it took personal contact, at one of the Society's Summer Research Institutes (1963). There is still another topological source: When Epstein, after working with Steenrod, decided to develop the properties of Steenrod operations in an arbitrary abelian category he was forced to consider coherence theorems, and so rediscovered Theorem 5.1 and other results about coherence.

Returning to closed categories, define a *constructible* $f: V \rightarrow W$ by recursion as follows: Every central map is constructible, as are the maps of the following three types (A, B, C, D arbitrary words):

$$(i) \quad V \xrightarrow{x} A \otimes B \xrightarrow{f \otimes g} C \otimes D \xrightarrow{y} W,$$

where x and y are central, $f: A \rightarrow C$ and $g: B \rightarrow D$ constructible;

$$(ii) \quad V \xrightarrow{\pi f} \text{Hom}(B, C) \xrightarrow{y} W$$

where y is central, $f: V \otimes B \rightarrow C$ constructible, and πf is its transpose under

the basic adjunction, \otimes to Hom ;

$$(iii) \quad \begin{array}{ccc} V \longrightarrow & H(B, C) \otimes A \otimes D & \xrightarrow{1 \otimes f \otimes 1} H(B, C) \otimes B \otimes D \\ & \xrightarrow{e} C \otimes D & \xrightarrow{g} W \end{array}$$

where H is short for Hom , while $f:A \rightarrow B$ and $g:C \otimes D \rightarrow W$ are constructible and x is central.

This definition allows us to test (in a finite number of steps) *whether* there is a constructible $V \rightarrow W$ with a given graph. In fact, the definition instructs us thus: Ask first if there is a central map with this graph. If not, try to factor the graph as a tensor product of two graphs, and test these (simpler) graphs, as in type (ii) above. If this fails, ask if the codomain W has the form $W = \text{Hom}(B, C)$, except for some factors I , and then test instead for constructible maps $V \otimes B \rightarrow C$ with the appropriate graph. Finally, if all these fail, consider in the domain V all prime factors of the form $\text{Hom}(B, C)$. For each, ask if there is another factor A (prime or composite of primes) and a constructible map $f:A \rightarrow B$ of appropriate graph, and determine whether this map, followed by evaluation e as in type (iii) above, will give a constructible map of the required graph. At each step in this process the problem is reduced to an evidently simpler one (or simpler several), so the whole process is finite.

Of the types (i)—(iii), the most characteristic one is (iii), which we call *prepared evaluation*, because it consists of an application of an evaluation e , followed by some other constructible g , but applied only after “preparation” by the prior application of $1 \otimes f \otimes 1$. For example, if we ask for the transpose of a canonical map like (5.3) above, with codomain $\text{Hom}(A, C)$, this method applies and gives precisely the expected map, composition. Similarly, given the two questionable graphs formulated in (5.2), prepared evaluation does produce a constructible map with the first graph, namely the transpose of

$$A \otimes \text{Hom}(A, B) \xrightarrow{e} \text{Hom}(A, B) \otimes A \xrightarrow{e} B.$$

In the second case of (5.2), the three tests quickly indicate that there is *no* constructible map with this graph. Hence these tests are indeed effective.

Now testing for constructibles with a given graph is the same as testing for canonical maps, and so provides an answer to Steenrod’s query, all in virtue of

THEOREM 5.2. *Every canonical map is constructible.*

Since the constructible maps include all instances of a, b, c, d , and e , the crucial part of the proof of this theorem is the demonstration that the composite of two constructible maps is again constructible. The proof (given in detail in Kelly-Mac Lane [51]) proceeds by induction and so amounts essentially to the elaborate but straightforward examination of nine cases of composition—each factor of the composite can be one of the three types (i), (ii), (iii) in the definition of constructible. Thus the proof of this theorem is

essentially a cut-elimination (=composition-elimination) process. However, I do not know any way in which *this* cut-elimination theorem can be derived from any version of the Gentzen cut-elimination theorem of proof theory. At present, there exists only a striking analogy.

Using this cut-elimination theorem, it is plausible that we can also exhibit all possible constructible maps $V \rightarrow W$ of given graph and so determine when such a map is unique. As Kelly-Mac Lane observed, this gives a coherence theorem:

THEOREM 5.3. *If V and W are words which do not involve the (nonletter) I , there is at most one canonical $V \rightarrow W$ with given graph.*

To get such a coherence (uniqueness) theorem, some such provision about I is necessary. Indeed, in any closed category one may define the “dual” $A^* = \text{Hom}(A, I)$ of an object A and hence also the double dual A^{**} . The search for a constructible map $\kappa_A : A \rightarrow A^{**}$ quickly yields one—a special case of the first map of (5.2), with $B = I$. For the category of modules or of vector spaces this map κ_A is just the familiar map sending each A to its double dual. By this example, it is easy to see that the composite

$$A^{***} \xrightarrow{\kappa_{(A^*)}} A^* \xrightarrow{(\kappa_A)^*} A^{***}$$

cannot be the identity. This produces a triangular diagram which does not commute and hence a graph which is realized by two different canonical maps. This is only one of a whole family of examples. For example, in 1950 Arens [1], [2], [3], in his study of canonical products for Banach algebras, produced two different canonical maps

$$A^{**} \otimes B^{**} \longrightarrow (A \otimes B)^{**}.$$

It is interesting to note that Arens at that date had essentially (under the name *phylon*) a definition of closed category.

Thus the coherence Theorem 5.3 must have some qualifications about the use of I 's. Actually, the Kelly-Mac Lane theorem has a more cautious qualification—they call a word *constant* if it involves no letters (i.e., is built up only from I), and they prove Theorem 5.3 for words V and W which involve a word $\text{Hom}(U, C)$ with C constant only when U is also constant. Going further, Voreadou has used similar methods to handle also graphs which additionally record linkages of the I 's, and she can establish a best-possible coherence theorem stateable in terms of graphs alone. It remains to find complete conditions (graphs+?) for the equality of two canonical maps $V \rightarrow W$.

There are a number of other coherence problems. The methods used by Kelly-Mac Lane will also prove that “all diagrams commute” in situations involving several categories relative to a given closed category, with various functors and natural transformations between them; again there are restrictions on the presence of the nonletter I (see Kelly-Mac Lane [51]). A number of other coherence problems are handled in the same Springer

Lecture Notes, Volume 281 on *Coherence in categories*. For example, Laplaza treats the question of the coherence of a tensor product which is distributive over a direct sum \oplus , as in the case of modules. He assumes that the distributive law is an isomorphism $A \otimes (B \oplus C) \cong (A \otimes B) \oplus (A \otimes C)$ and that a whole sequence of appropriate identities are satisfied, and then shows that two canonical maps $V \rightarrow W$ are equal when they have the same “distortion”, suitably defined (and resembling the notion of the same “graph”).

Another difficult case is that of a closed category in the sense in which this was originally defined by Eilenberg-Kelly: A category not equipped with a tensor product, but having an internal hom-functor Hom together with a structural map

$$L : \text{Hom}(B, C) \rightarrow \text{Hom}(\text{Hom}(A, B), \text{Hom}(A, C))$$

—the map which would have been the transpose of composition

$$\text{Hom}(B, C) \otimes \text{Hom}(P, B) \rightarrow \text{Hom}(A, C)$$

had a \otimes -product has been present. This structural map L satisfies a suitable (pentagonal, even) condition, plus other identities. The problem is again that of proving that “all diagrams commute.” Currently there appears to be considerable progress, in work of Day and Laplaza, on this question.

6. Naturality and categories. Arrows $f : X \rightarrow Y$ were first used to visualize continuous functions by some topologist—reportedly by Hurewicz in seminar lectures about 1940. Once used, they present intriguing questions which raise new points of view about algebra. Each continuous map f “induces” a corresponding map on (integral) homology

$$(6.1) \quad H_n(f) : H_n(X) \rightarrow H_n(Y),$$

but if f is just an inclusion (of a subspace $X \subset Y$), the map $H_n(f)$ on homology is by no means necessarily also an inclusion; for example the inclusion $S^1 \subset D^2$ of the circle as boundary of the disc D^2 sends the one-dimensional homology of the circle to zero. It is probably for this reason that early pictures of maps by inclusions $f(X) \subset Y$ turned out to be inadequate. Also, a continuous map of spaces does induce homomorphisms $H_n(f)$ of groups—but that homomorphism is not necessarily onto, in flagrant disregard of the earlier view by algebraists (notably in van der Waerden’s *Modern algebra*) that a homomorphism of groups is a mapping *onto*.

Some useful maps arising in geometry are clearly “natural” ones, like the map of a vector space V to its double dual $V \rightarrow V^{**}$ or the map of an abelian group G to its double character group, $G \rightarrow G^{**}$. Initially, it was easy enough to leave this notion informal, with the side observation that for finite-dimensional vector spaces V the usual isomorphism $V \cong V^*$ was *not* natural, because it depended on choices (here, on a choice of bases). The issue of naturality becomes more perplexing when it turns out that the desired “universal coefficient formula” for cohomology can’t really be

expressed by a single “formula”, because the resulting isomorphism, as expressed in (2.12), is not a natural one. One is forced instead to use the language of exact sequences, as in (2.13), and one then really is relying on the fact that the maps present in the sequence are natural ones. For example, for a chain complex C and an abelian group A , why is the canonical map

$$(6.2) \quad H^n(C, A) \rightarrow \text{Hom}(H_n(C), A)$$

a natural one? The question becomes more pressing when one needs, for example for Čech cohomology, to compute this expression Hom or the corresponding Ext (of (2.12)) over an inverse system of complexes or groups. Then the naturality of these maps becomes a technical question and not a theological one.

These were the sorts of circumstances which led Eilenberg-Mac Lane in 1942 to analyse naturality. The canonical map $\chi: G \rightarrow G^{**}$ to the double character group is natural because, for each group homomorphism $f: G \rightarrow H$, one has $\chi f = f^{**} \chi$, as expressed by the commutative diagram

$$(6.3) \quad \begin{array}{ccc} G & \xrightarrow{\chi} & G^{**} \\ f \downarrow & & \downarrow f^{**} \\ H & \xrightarrow{\chi} & H^{**} \end{array}$$

Similarly, the map (6.2) from cohomology is natural because, for any map $h: C \rightarrow C'$ of chain complexes one has a commutative diagram

$$(6.4) \quad \begin{array}{ccc} H^n(C', A) & \longrightarrow & \text{Hom}(H_n(C'), A) \\ H^n(h) \downarrow & & \downarrow \text{Hom}[(H_n h, 1)] \\ H^n(C, A) & \longrightarrow & \text{Hom}(H_n(C), A) \end{array}$$

This formulation in turn required clarity as to the induced maps involved—the map $H_n(h): H_n(C) \rightarrow H_n(C')$ induced on homology and going in the same direction as the given h , plus the induced map $H^n(h): H^n(C', A) \rightarrow H^n(C, A)$ on cohomology going in the opposite direction. Thereby we were led to the general notion of a *functor* on chain complexes to abelian groups—covariant in the case H_n of homology and contravariant in the case of cohomology. In its turn, a general treatment of functors on mathematical objects of any kind requires the development of the notion of category, while the evidently relevant study of the category of *all* objects of a given kind required attention to the foundational questions involved in that “all.”

These ideas of category, functor, and natural transformation motivated thus by topological questions and explicitly formulated by Eilenberg-Mac Lane in 1945, have provided a fitting framework for later developments. The idea of a “universal construction” was first formulated explicitly by Pierre Samuel in 1948, and directed to the understanding of questions about free topological groups; it really required a categorical setting. This was

finally provided by D. M. Kan's formulation of the notion of adjoint functor in 1958—again based on needs dictated by examples of adjoints acting on simplicial sets, as well as by the basic fact (already noted in §5) that $\text{Hom}(A, -)$ can be described as the left adjoint of $-\otimes A$. Once disengaged, this notion of adjoint functor turned out to be present and helpful in many strikingly different contexts. The basic existence theorem for free objects, known after a fashion to Bourbaki, was effectively formulated in Freyd's adjoint functor theorem (Freyd [34]). In this connection, I had long been persuaded that categorical notions would provide an appropriate setting for universal algebra. This was indeed accomplished; on the one hand, Lawvere's 1963 notion of an "algebraic theory" showed that one should formulate varieties of universal algebras not in terms of *particular* defining operators (unary, binary, ternary, \dots) but in terms of *all* the operations possible for that variety. On the other hand, Eilenberg and Moore in 1965 showed how a pair of adjoint functors produced in the "underlying" category a single functor with a monad (or "triple") structure, and how the upper category could be reconstituted from that monad as a category of "algebras" for the monad. The elegance of this construction was emphasized by the fact that their definition of "algebras" for the monad turned out to be the exact and simple analog of the definition of the modules over a ring or the actions of a group (cf. Mac Lane [76, Chapter VI]). On these grounds Beck was also able to obtain his decisive theorem characterizing the category of all algebras among other categories leading to the same monad. In this way the "universal" parts of algebra have been drastically reformed by the ideas of category, natural transformation, and adjoint functors, all arising from topology. The transformation still continues, sparked currently by the extensive applications of ideas from universal algebra to automata theory (Eilenberg [18], Elgot [32], Manes [80]).

The development of these categorical ideas originating from topology has also had a reverse use in topology and geometry. The notion of the localization of a category (Gabriel-Zisman [36]) has led in particular, in the hands of Quillen [84] and Sullivan [92], to the localization of homotopy theory—in particular, to the study of rational homotopy. The recent development of algebraic K -theory (see, e.g., Bass [7]) has involved many categorical constructions, and this is especially true of Quillen's recent definition [85] of the higher K -groups. The Grothendieck school of algebraic geometry, depending as it does on maximal abstractions and generality, has been extensively categorical; indeed some portion of this development, notably work by Hakim [40] and Rivano [86], are practically purely categorical.

The influence of logic and foundations has also been strong. At several points, for example in 1959 [70], I had pointed out that the proper treatment of big categories, such as categories of functors, did require a different approach to the foundations and to the use of set theory. In 1964, Lawvere [58] had observed that the familiar foundational scheme of using axioms (Zermelo-Fraenkel or Gödel-Bernays) for set membership might be

replaced by a different sort of axioms on the category of sets—using composition of functions rather than membership of elements as the decisive primitive notion. This approach did not become fully effective until it was combined by Lawvere and Tierney with an idea coming from the Grothendieck school of algebraic geometry: the characterization by Giraud of the category of all sheaves of sets on a given (Grothendieck) topology (see Artin, Grothendieck, and Verdier [4]). It turned out that the axioms for the category of sheaves could be suitably restructured in “first order” form so that they also applied to the category of sets. Both the classical two-valued logic and the intuitionistic logic play suitable roles; at present, this study of “topoi” is developing very rapidly. (Koch-Wraith [53], Lawvere [59], Freyd [35], Mac Lane [77]). It points up in clear form the advantages to be gained from not isolating mathematical logic as a sort of separate and self-contained specialty.

7. What is algebra? This essay has recounted a number of instances in which the development of algebraic ideas has been stimulated by problems or queries arising outside algebra; in particular, in logic or in geometry. These developments have been varied and often unexpected: One would hardly have guessed ahead of time that the study of separably generated extensions of a field would be assisted by axiomatic formulations of linear independence or that the foundations of mathematics would be vitally altered by the consequences of a theorem characterizing the category of sheaves on a topological space. The examples of outside influences on algebra which I have recounted are largely drawn from my own work. More narrowly, many of them are dominated by the results of the constructions Hom and \otimes on groups or spaces. There are many other instances of effects of geometry upon algebra, such as the continuing input from algebraic geometry in the theory of commutative rings or the stimulus to multilinear algebra from geometry and mechanics.

On some occasions I have been tempted to try to define what algebra is, can, or should be—most recently in concluding a survey [72] on *Recent advances in algebra*. But no such formal definitions hold valid for long, since algebra and its various subfields steadily change under the influence of ideas and problems coming not just from logic and geometry, but from analysis, other parts of mathematics, and extra mathematical sources. The progress of mathematics does indeed depend on many interlocking, unexpected and multiform developments.

REFERENCES

1. Richard F. Arens, *Operations induced in conjugate spaces*, Proc. Internat. Congr. of Math. (Cambridge, Mass., 1950), vol. I, Amer. Math. Soc., Providence, R.I., 1952, pp. 532–533.
2. ———, *The adjoint of a bilinear operator*, Proc. Amer. Math. Soc. **2** (1951), 839–848. MR **13**, 659.
3. ———, *Operations induced in function classes*, Monatsh. Math. **55** (1951), 1–19. MR **13**, 372.
4. M. Artin, A. Grothendieck and J. L. Verdier, *La théorie des topos et cohomologie étale des*

schémas (SGA 4), vols. I, II, III, (Séminaire de géométrie algébrique du Bois-Marie 1963/64), Lecture Notes in Math., vols. 269, 270, 305, Springer-Verlag, Berlin, Heidelberg and New York, 1972, 1973.

5. R. Baer, *Erweiterung von Gruppen und ihren Isomorphismen*, Math. Z. **38** (1934), 375–416.

6. Michael Barr, *Cohomology and obstructions: Commutative algebras*, Seminar on Triples and Categorical Homology Theory (ETH, Zürich, 1966/67), Lecture Notes in Math., vol. 80, Springer-Verlag, Berlin, Heidelberg and New York, 1969, pp. 357–373. MR **42** #6075.

7. H. Bass, *Algebraic K-theory*, Benjamin, New York and Amsterdam, 1968. MR **40** #2736.

8. N. Bourbaki, *Éléments de mathématique*. XI. Part I: *Les structures fondamentales de l'analyse*. Livre II: *Algèbre*, Chap. 5: *Corps commutatifs*, Actualités Sci. Indust., no. 1102, Hermann, Paris, 1950. MR **12**, 6.

9. H. Cartan, *Séminaire Henri Cartan de l'École Normale Supérieure, 1954/1955, Algèbres d'Eilenberg-Mac Lane et homotopie*, Secrétariat mathématique, Paris, 1955. MR **19**, 438.

10. H. Cartan and S. Eilenberg, *Homological algebra*, Princeton Univ. Press, Princeton, N.J., 1956. MR **17**, 1040.

11. E. Čech, *Les groupes de Betti d'un complexe infini*, Fund. Math. **25** (1935), 33–44.

12. G. J. Decker, *The integral homology algebra of an Eilenberg-Mac Lane space*, Thesis, Univ. of Chicago, Chicago, 1974.

13. J. Duskin, *$K(\Pi, n)$ -torsors and the interpretation of "triple" cohomology*, Proc. Nat. Acad. Sci. U.S.A. **71** (1974), 2554–2557. MR **49** #5139.

14. —, *Simplicial methods and the interpretation of "triple" cohomology*, Mem. Amer. Math. Soc. No. 163 (to appear).

15. E. Dyer and R. K. Lashof, *Homology of iterated loop spaces*, Amer. J. Math. **84** (1962), 35–88. MR **25** #4523.

16. B. Eckmann, *Der Cohomologie-Ring einer beliebigen Gruppe*, Comment. Math. Helv. **18** (1946), 232–282. MR **8**, 166.

17. S. Eilenberg, *Singular homology theory*, Ann. of Math. (2) **45** (1944), 407–447. MR **6**, 96.

18. —, *Automata, languages, and machines*. Vol. A, Academic Press, New York, 1974.

19. S. Eilenberg and G. M. Kelly, *A generalization of the functorial calculus*, J. Algebra **3** (1966), 366–375. MR **32** #7618.

20. S. Eilenberg and S. Mac Lane, *Group extensions and homology*, Ann. of Math. (2) **43** (1942), 757–831. MR **4**, 88.

21. —, *Natural isomorphisms in group theory*, Proc. Nat. Acad. Sci. U.S.A. **28** (1942), 537–543. MR **4**, 134.

22. —, *Relations between homology and homotopy groups*, Proc. Nat. Acad. Sci. U.S.A. **29** (1943), 155–158. MR **4**, 224.

23. —, *General theory of natural equivalences*, Trans. Amer. Math. Soc. **58** (1945), 231–294. MR **7**, 109.

24. —, *Relations between homology and homotopy groups of spaces*, Ann. of Math. (2) **46** (1945), 480–509. MR **7**, 137.

25. —, *Homology of spaces with operators*. II, Trans. Amer. Math. Soc. **65** (1949), 49–99. MR **11**, 379.

26. —, *Relations between homology and homotopy groups of spaces*. II, Ann. of Math. (2) **51** (1950), 514–533. MR **11**, 735.

27. —, *Acyclic models*, Amer. J. Math. **75** (1953), 189–199. MR **14**, 670.

28. —, *On the groups $H(\Pi, n)$* . I, Ann. of Math. (2) **58** (1953), 55–106. MR **15**, 54.

29. —, *On the groups $H(\Pi, n)$* . II. *Methods of computation*, Ann. of Math. (2) **60** (1954), 49–139. MR **16**, 391.

30. —, *On the groups $H(\Pi, n)$* . III. *Operations and obstructions*, Ann. of Math. (2) **60** (1954), 513–557. MR **16**, 392.

31. S. Eilenberg and J. A. Zilber, *Semi-simplicial complexes and singular homology*, Ann. of Math. (2) **51** (1950), 499–513. MR **11**, 734.

- 32.** C. Elgot, *Monadic computation and iterative algebraic theories*, Proc. Logic Colloquium (Bristol, 1973) edited by H. E. Rose and J. C. Shepherdson, North-Holland Publishing Co., Amsterdam. Oxford 1975, pp. 175–230.
- 33.** D. B. A. Epstein, *Functors between tensored categories*, Invent. Math. **1** (1966), 221–228. MR **35** #4276.
- 33a.** H. Freudenthal, *Der Einfluss der Fundamentalgruppe auf die Bettischen Gruppen*, Ann. of Math. (2) **47** (1946), 274–316. MR **8**, 166.
- 34.** P. Freyd, *Abelian categories. An introduction to the theory of functors*, Harper's Ser. in Modern Math., Harper & Row, New York, 1964. MR **29** #3517.
- 35.** ———, *Aspects of topoi*, Bull. Austral. Math. Soc. **7** (1972), 1–76.
- 36.** P. Gabriel and M. Zisman, *Calculus of fractions and homotopy theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 35, Springer-Verlag, Berlin, Heidelberg and New York, 1967. MR **35** #1019.
- 37.** G. Gentzen, *Untersuchungen über das logische Schliessen*. I, II, Math. Z. **39** (1934), 176–210, 405–431.
- 38.** M. Gerstenhaber, *On the deformation of rings and algebras*. II, Ann. of Math. (2) **84** (1966), 1–19. MR **34** #7608.
- 39.** S. J. Goldberg, *Extension of Lie algebras and the third cohomology group*, Canad. J. Math. **5** (1953), 470–476. MR **15**, 282.
- 40.** M. Hakim, *Topos annelés et schémas relatifs*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 64, Springer-Verlag, Berlin and New York, 1972.
- 41.** R. Hamsher, *Eilenberg-Mac Lane algebras and their computation*, Thesis, Univ. of Chicago, Chicago, Ill., 1973.
- 42.** G. Hochschild, *Cohomology classes of finite type and finite dimensional kernels for Lie algebras*, Amer. J. Math. **76** (1954), 763–778. MR **16**, 443.
- 43.** ———, *Lie algebra kernels and cohomology*, Amer. J. Math. **76** (1954), 698–716. MR **16**, 109.
- 44.** H. Hopf, *Fundamentalgruppe und zweite Bettische Gruppe*, Comment. Math. Helv. **14** (1942), 257–309. MR **3**, 316.
- 45.** ———, *Relations between the fundamental group and the second Betti group*, Lectures in Topology, Univ. of Michigan Press, Ann Arbor, Mich., 1941, pp. 315–316. MR **3**, 135.
- 46.** ———, *Nachtrag zu der Arbeit Fundamentalgruppe und zweite Bettische Gruppe*, Comment. Math. Helv. **15** (1943), 27–32. MR **4**, 173.
- 47.** ———, *Über die Bettischen Gruppen, die zu einer beliebigen Gruppe gehören*, Comment. Math. Helv. **17** (1945), 39–79, MR **6**, 279.
- 48.** D. M. Kan, *Adjoint functors*, Trans. Amer. Math. Soc. **87** (1958), 294–329. MR **24** #A1301.
- 49.** ———, *A combinatorial definition of homotopy groups*, Ann. of Math. (2) **67** (1958), 282–312. MR **22** #1897.
- 50.** G. M. Kelly, *On Mac Lane's conditions for coherence of natural associativities, commutativities, etc.*, J. Algebra **1** (1964), 397–402. MR **32** #132.
- 51.** G. M. Kelly and S. Mac Lane, *Coherence in closed categories*, J. Pure Appl. Algebra **1** (1971), no. 1, 97–140; erratum, *ibid.* no. 2, 219. MR **44** #278; **45** #1988.
- 52.** ———, *Closed coherence for a natural transformation*, Coherence in Categories, Lecture Notes in Math., vol. 281, Springer-Verlag, Berlin, Heidelberg and New York, 1972, pp. 1–28. MR **48** #8591.
- 53.** A. Kock and G. C. Wraith, *Elementary toposes*, Lecture Notes Series, no. 30, Aarhus Universitet, Aarhus, Denmark, 1971.
- 54.** J. Lambek, *Deductive systems and categories*. I. *Syntactic calculus and residuated categories*, Math. Systems Theory **2** (1968), 287–318. MR **38** #4277.
- 55.** ———, *Deductive systems and categories*. II. *Standard constructions and closed categories*, Category Theory, Homology Theory and their Applications, I (Battelle Institute Conf., Seattle, Wash., 1968, Vol. One), Lecture Notes in Math., vol. 86, Springer-Verlag, Berlin, Heidelberg and New York, 1969, pp. 76–122. MR **39** #3967.

- 56.** K. Lamotke, *Semisimpliziale algebraische Topologie*, Die Grundlehren der math. Wissenschaften, Band 147, Springer-Verlag, Berlin, Heidelberg and New York, 1968. MR **39** #6318.
- 57.** F. W. Lawvere, *Functorial semantics of algebraic theories*, Proc. Nat. Acad. Sci. U.S.A. **50** (1963), 869–872. MR **28** #2143.
- 58.** ———, *An elementary theory of the category of sets*, Proc. Nat. Acad. Sci. U.S.A. **52** (1964), 1506–1511. MR **30** #3025.
- 59.** ———. (editor), *Toposes, algebraic geometry and logic*, Lecture Notes in Math., vol. 274, Springer-Verlag, Berlin, Heidelberg and New York, 1972. MR **48** #8592.
- 60.** S. Mac Lane, *Some interpretations of abstract linear dependence in terms of projective geometry*, Amer. J. Math. **58** (1936), 236–240.
- 61.** ———, *A lattice formulation for transcendence degrees and p -bases*, Duke Math. J. **4** (1938), 455–468.
- 62.** ———, *The uniqueness of the power series representation of certain fields with valuations*, Ann. of Math. **39** (1938), 370–382.
- 63.** ———, *Modular fields, I. Separating transcendence bases*, Duke Math. J. **5** (1939), 372–393.
- 64.** ———, *Steinitz field towers for modular fields*, Trans. Amer. Math. Soc. **46** (1939), 23–45. MR **1**, 3.
- 65.** ———, *Subfields and automorphism groups of p -adic fields*, Ann. of Math. **40** (1939), 423–442.
- 66.** ———, *Note on the relative structure of p -adic fields*, Ann. of Math. (2) **41** (1940), 751–753. MR **2**, 123.
- 67.** ———, *Modular fields*, Amer. Math. Monthly **47** (1940), 259–274. MR **1**, 328.
- 68.** ———, *Homology products in $K(\Pi, n)$* , Proc. Amer. Math. Soc. **5** (1954), 642–651. MR **16**, 160.
- 69.** ———, *Extensions and obstructions for rings*, Illinois J. Math. **2** (1958), 316–345. MR **20** #5228.
- 70.** ———, *Locally small categories and the foundations of set theory*, Infinitistic Methods (Proc. Sympos. Foundations of Math., Warsaw, 1959), Pergamon, Oxford; PWN, Warsaw, 1961, pp. 25–43. MR **33** #168.
- 71.** ———, *Homology*, Die Grundlehren der math. Wissenschaften, Band 114, Academic Press, New York: Springer-Verlag, Berlin, Göttingen, Heidelberg, 1963; 3rd corrected printing, 1975. MR **28** #122.
- 72.** ———, *Some additional advances in algebra*, Studies in Modern Algebra (A. A. Albert, editor), Studies in Math., vol. 2, The Math. Assoc. of Amer.; distributed by Prentice-Hall, Englewood Cliffs, N.J., 1963, pp. 35–58. MR **26** #3750.
- 73.** ———, *Natural associativity and commutativity*, Rice Univ. Studies **49** (1963), no. 4, 28–46. MR **30** #1160.
- 74.** ———, *The influence of M. H. Stone on the origins of category theory*, Functional Analysis and Related Fields (F. E. Browder, editor), Springer-Verlag, Berlin, Heidelberg and New York, 1970, pp. 228–241.
- 75.** ———, *The Milgram bar construction as a tensor product of functors*, The Steenrod Algebra and its Applications (Proc. Conf. to Celebrate N. E. Steenrod's Sixtieth Birthday, Battelle Memorial Inst., Columbus, Ohio, 1970), Lecture Notes in Math., vol. 168, Springer-Verlag, Berlin, Heidelberg and New York, 1970; pp. 135–152. MR **42** #8495.
- 76.** ———, *Categories for the working mathematician*, Springer-Verlag, Berlin, Heidelberg and New York, 1971.
- 77.** ———, *Sets, topoi, and internal logic in categories*, Proc. Logic Colloquium (Bristol, 1973) edited by H. E. Rose and J. C. Shepherdson, North-Holland Publishing Co., Amsterdam. Oxford 1975, pp. 119–134.
- 78.** S. Mac Lane and O. F. G. Schilling, *Normal algebraic number fields*, Trans. Amer. Math. Soc. **50** (1941), 295–384. MR **3**, 102.
- 79.** S. Mac Lane and F. K. Schmidt, *The generation of inseparable fields*, Proc. Nat. Acad. Sci. U.S.A. **27** (1941), 583–587. MR **3**, 263.

- 80.** E. Manes, *Algebraic theories*, Graduate Texts in Math., Springer-Verlag, Berlin, Heidelberg and New York (to appear).
- 81.** J. P. May, *Simplicial objects in algebraic topology*, Van Nostrand Math. Studies, no. 11, Van Nostrand, Princeton, N.J., 1967. MR **36** #5942.
- 82.** R. J. Milgram, *The bar construction and abelian H-spaces*, Illinois J. Math. **11** (1967), 242–250. MR **34** #8404.
- 83.** M. Mori, *On the three-dimensional cohomology group of Lie algebras*, J. Math. Soc. Japan **5** (1953), 171–183. MR **15**, 282.
- 84.** D. G. Quillen, *Homotopical algebra*, Lecture Notes in Math, vol 43, Springer-Verlag, Berlin, Heidelberg and New York, 1967. MR **36** #6480.
- 85.** ———, *On the cohomology and K-theory of the general linear groups over a finite field*, Ann. of Math. (2) **96** (1972), 552–586. MR **47** #3565.
- 86.** N. Saavedra Rivano, *Catégories tannakiennes*, Lecture Notes in Math., vol. 265, Springer-Verlag, Berlin, Heidelberg and New York, 1972. MR **49** #2769.
- 87.** M. Rothenberg and N. E. Steenrod, *The cohomology of classifying spaces of H-spaces*, Bull. Amer. Math. Soc. **71** (1965), 872–875. MR **34** #8405.
- 88.** J.-P. Serre, *Cohomologie galoisienne*, Cours au Collège de France, 1962–1963, 2nd. ed., Lectures Notes in Math., vol. 5, Springer-Verlag, Berlin, Heidelberg and New York, 1964. MR **31** #4785.
- 89.** U. Shukla, *Cohomologie des algèbres associatives*, Ann. Sci. École Norm. Sup. (3) **78** (1961), 163–209. MR **24** #A2605.
- 90.** J. D. Stasheff, *Homotopy associativity of H-spaces. I*, Trans. Amer. Math. Soc. **108** (1963), 275–292. MR **28** #1623.
- 91.** N. E. Steenrod, *Milgram's classifying space of a topological group*, Topology **7** (1968), 349–368. MR **38** #1675.
- 92.** D. P. Sullivan, *Geometric topology, part 1, Localization, periodicity, and Galois symmetry*, Mimeographed notes, M.I.T., Cambridge, Mass., 1970.
- 93.** J. Tate, *The higher dimensional cohomology groups of class field theory*, Ann. of Math (2) **56** (1952), 294–297. MR **14**, 252.
- 94.** O. Teichmüller, *p-Algebren*, Deutsche Math. **1** (1936), 362–388.
- 95.** ———, *Diskret bewertete perfekte Körper mit unvollkommenem Restklassenkörper*, J. Reine Angew Math. **176** (1936), 141–152.
- 96.** ———, *Über die sogenannte nichtkommutative Galoissche Theorie und die Relation $\xi_{\lambda, \mu, \nu} \xi_{\mu, \nu, \pi} \xi_{\nu, \pi, \lambda} = \xi_{\lambda, \mu, \nu} \xi_{\mu, \nu, \pi}$* , Deutsche Math. **5** (1940), 138–149. MR **2**, 122.
- 97.** O. Veblen, *Analysis situs*, 2nd ed., Amer. Math. Soc. Colloq. Publ., vol. 5, part II, Amer. Math. Soc., Providence, R. I., 1931.
- 98.** R. Voreadou, *A coherence theorem for closed categories* (to appear).
- 99.** ———, *Non-commutative diagrams in closed categories* (to appear).
- 100.** A. Weil, *Foundations of algebraic geometry*, Amer. Math. Soc. Colloq. Publ., vol. 29, Amer. Math. Soc., Providence, R. I., 1946. MR **9**, 303.
- 101.** J. H. C. Whitehead, *A certain exact sequence*, Ann. of Math. (2) **52** (1950), 51–110. MR **12**, 43.
- 102.** H. Whitney, *On the abstract properties of linear dependence*, Amer. J. Math. **57** (1935), 509–533.
- 103.** ———, *The maps of an n-complex into an n-sphere*, Duke Math. J. **3** (1937), 51–55.
- 104.** ———, *Tensor products of abelian groups*, Duke Math. J. **4** (1938), 495–528.