

NATURAL ASSOCIATIVITY AND COMMUTATIVITY

by Saunders Mac Lane¹

1. Introduction. The usual associative law $a(bc) = (ab)c$ is known to imply the “general associative law,” which states that any two iterated products of the same factors in the same order are equal, irrespective of the arrangement of parentheses. Here we are concerned with an associativity given by an isomorphism $a: A(BC) \cong (AB)C$; more exactly, with the case where the product AB is a covariant functor of its arguments A and B , while associativity is an isomorphism a natural in its arguments A, B , and C . The general associative law again shows that any two iterated products F and F' of the n arguments A_1, \dots, A_n are naturally isomorphic, under a natural isomorphism $F \cong F'$ given by “iteration” of a . We then ask: what conditions must be placed upon a if there is to be just *one* such isomorphism $F \cong F'$ for each pair F, F' ? This question arises in categorical algebra, as do the corresponding questions for a natural commutative law $AB \cong BA$ and for an identity element K for the multiplication AB , with $KA \cong A$ natural. Here we present answers to each of these questions. The first question (associativity alone) has already been treated by Stasheff [10] in connection with homotopy associative H -spaces. After these lectures formulating our answers had been delivered, we found that the same questions had been answered by Epstein and formulated in a privately circulated preprint of his study [3] of Steenrod operations in Abelian categories. His results are certainly independent of and probably prior to ours.

For categories we employ the usual terminology (see for example, [7], Chapters I and IX). A *category* \mathcal{C} is a class of “objects” A, B, C, \dots together with a family of disjoint sets $\text{hom}(A, B) = \text{hom}_{\mathcal{C}}(A, B)$, one for each ordered pair of objects. When $f \in \text{hom}(A, B)$ we also write $f: A \rightarrow B$ and we call f a *morphism* of \mathcal{C} with *domain* A and *range* B . A *composite* gf is defined whenever $f: A \rightarrow B$ and $g: B \rightarrow C$ are morphisms with $\text{range}(f) = \text{domain}(g)$: it is a morphism $gf: A \rightarrow C$. There are two axioms: the triple composite is associate

¹ This research was supported in part by the Air Office of Scientific Research, with a grant No. AF-AFOSR-62-170.

Editor's Note: This paper was presented in three lectures in Anderson Hall, Rice University, 23, 24, 26 September 1963. Mr. Mac Lane is the Max Mason Distinguished Service Professor of Mathematics at the University of Chicago.

(whenever defined); to each object B there is a morphism $1_B: B \rightarrow B$ which acts as a left and a right identity under composition.

If \mathcal{B} and \mathcal{C} are categories, a (covariant) *functor* $F: \mathcal{B} \rightarrow \mathcal{C}$ consists of an *object function* and a *mapping function*. The object function assigns to each $B \in \mathcal{B}$ an object $F(B)$ in \mathcal{C} : the mapping function assigns to each morphism $f: B \rightarrow B'$ in \mathcal{B} a morphism $F(f): F(B) \rightarrow F(B')$ in \mathcal{C} in such a way that

$$F(1_B) = 1_{F(B)}, \quad F(gf) = (Fg)(Ff),$$

the latter whenever the composite gf is defined.

If $F, G: \mathcal{B} \rightarrow \mathcal{C}$ are functors, a *natural transformation* $t: F \rightarrow G$ is a function t which assigns to each object B of \mathcal{B} a morphism

$$t(B): F(B) \rightarrow G(B) \quad \text{in } \mathcal{C}$$

such that for every morphism $f: B \rightarrow B'$ in \mathcal{B} the diagram

$$\begin{array}{ccc} F(B) & \xrightarrow{F(f)} & F(B') \\ \downarrow t(B) & & \downarrow t(B') \\ G(B) & \xrightarrow{G(f)} & G(B') \end{array} \quad \text{in } \mathcal{C}$$

is commutative. Moreover, t is a *natural equivalence* (or a natural isomorphism) if each $t(B)$ has in \mathcal{C} a two-sided inverse. It follows that this inverse is a natural transformation $t^{-1}: G \rightarrow F$.

If \mathcal{C} and \mathcal{C}' are categories, their (cartesian) product $\mathcal{C} \times \mathcal{C}'$ has as objects the pairs (C, C') of objects and as morphism the pairs

$$(f, f'): (C, C') \rightarrow (D, D')$$

of morphisms $f: C \rightarrow D$ and $f': C' \rightarrow D'$, with the evident composition and identities. A functor $F: \mathcal{C} \times \mathcal{C}' \rightarrow \mathcal{D}$ is called a *bifunctor* on the categories \mathcal{C} and \mathcal{C}' , and similarly for functors of more arguments and for their natural transformations. In particular, \mathcal{C}^n will denote the n -fold product of the category \mathcal{C} with itself, so that a functor $F: \mathcal{C}^n \rightarrow \mathcal{D}$ is a covariant functor on n arguments in \mathcal{C} with values in \mathcal{D} . We say that F has *multiplicity* n .

2. Categories with a multiplication. By a category with a multiplication we mean a category \mathcal{C} together with a covariant bifunctor on \mathcal{C} to \mathcal{C} . This bifunctor will be denoted by the symbol \otimes , written between its arguments. Thus the statement that \otimes is a bifunctor means:

- (i) Each pair A, B of objects of \mathcal{C} yields an object $A \otimes B$ of \mathcal{C} ;
- (ii) Each pair $f: A \rightarrow A'$ and $g: B \rightarrow B'$ of morphisms of \mathcal{C} yields

$$(2.1) \quad f \otimes g: A \otimes B \rightarrow A' \otimes B',$$

a morphism of \mathcal{C} ;

(iii) Identity morphisms $1_A: A \rightarrow A$ and $1_B: B \rightarrow B$ of \mathcal{C} yield

$$(2.2) \quad 1_A \otimes 1_B = 1_{A \otimes B}: A \otimes B \rightarrow A \otimes B,$$

the identity morphism of $A \otimes B$;

(iv) If the composites $f'f$ and $g'g$ are defined in \mathcal{C} , then

$$(2.3) \quad (f'f) \otimes (g'g) = (f' \otimes g')(f \otimes g).$$

Examples abound. For instance, in a category \mathcal{C} with finite products the product is a multiplication. In detail, a diagram

$$(2.4) \quad A \xleftarrow{p} P \xrightarrow{q} B \quad \text{in } \mathcal{C}$$

(with fixed ends A and B) is called a *product diagram* if to each diagram

$$A \xleftarrow{f} C \xrightarrow{g} B \quad \text{in } \mathcal{C}$$

with the same ends there exists a unique morphism $h: C \rightarrow P$ such that $f = ph$ and $g = qh$. The category has *finite products* if there is such a product diagram for each pair of objects A and B of \mathcal{C} . When this is the case, the middle object P of the product diagram is uniquely determined, up to equivalence, by A and B . Choosing one $P = A \times B$ for each A and B yields a bifunctor \times on \mathcal{C} , hence a multiplication. This includes examples such as the cartesian product of sets or of topological spaces and the direct product of groups (in the category of all groups, with morphisms all group homomorphisms).

The dual notion is that of a coproduct. In a category \mathcal{C} , a diagram

$$A \xrightarrow{i} D \xleftarrow{j} B$$

is called a *coproduct diagram* if to any diagram

$$A \xrightarrow{f} C \xleftarrow{g} B$$

with the same ends there is a unique $h: D \rightarrow C$ such that $f = hi$ and $g = hj$. The category \mathcal{C} has *finite coproducts* if there is such a coproduct diagram for each pair of objects A and B . When this is the case, the middle object D in the coproduct diagram is uniquely determined, up to equivalence, by A and B , and a choice $D = A * B$ provides a bifunctor $*$ which is a multiplication for \mathcal{C} . Examples are the free product of two groups (in the category of groups) or the “wedge” of two spaces (in the category of topolo-

gical spaces with a selected base point, with morphisms continuous maps carrying base point to base point).

More important for our applications are the categories with a multiplication given by one of the usual tensor products. For example, in the category \mathcal{M}_F of all vector spaces V, W, \dots over a fixed field F , the usual tensor product $V \otimes W$ of two vector spaces is a multiplication. The same holds for the category \mathcal{M}_K of all modules over a fixed commutative ring K . Similarly, let $DG(\mathcal{M}_K)$ be the category of all differential graded K -modules (i.e., of all chain complexes of K -modules). The usual tensor product of differential graded modules (defined, say, as in [7], Chap. VI. 7) is a multiplication in this category.

If \mathcal{C} is a category, a subcategory \mathcal{C}' (with the *same* objects as \mathcal{C}) is given by a subclass of the class of all morphisms of \mathcal{C} , such that this subclass contains every identity morphism of \mathcal{C} and with each pair of morphisms their composite (whenever defined in \mathcal{C}). If \mathcal{C} is a category with a multiplication \otimes , a *multiplicative subcategory* is a subcategory \mathcal{C}' which contains with any two morphisms f and g the morphism $f \otimes g$; then \mathcal{C}' is itself a category with a multiplication. Given any set S of morphisms of \mathcal{C} , we may speak of the multiplicative subcategory *generated* by the morphisms of S ; it is defined to be the intersection of all multiplicative subcategories of \mathcal{C} which contain S . The process of generation can be described more explicitly. By an *expansion* of a morphism f we mean a morphism such as $f \otimes 1_A$, $1 \otimes (f \otimes 1)$, $[1 \otimes (f \otimes 1)] \otimes 1$, etc. More formally, the set of *expansions* of f is the smallest set of morphisms of \mathcal{C} which contains f and with any morphism e all morphisms $e \otimes 1_C$ and $1_C \otimes e$ for any object C of the category.

THEOREM 2.1. *Let S be a set of morphisms in a category \mathcal{C} with a multiplication. The multiplicative subcategory of \mathcal{C} generated by S consists of all identity morphisms of \mathcal{C} and of all composites of expansions of morphisms of S .*

For a proof, we must show that the indicated set of morphisms is closed under composition and under \otimes -multiplication. For composition, this is immediate. For \otimes -multiplication, use (2.3), which states that a \otimes -product of two composites can be rewritten as a composite of two tensor products. By iterated application of this result, the problem is reduced to the tensor product of two expansions, say $f \otimes g$. But $f = f1$ and $g = 1'g$, where 1 and $1'$ are suitable identity morphisms, so that (2.3) yields

$$(2.4) \quad f \otimes g = (f1) \otimes (1'g) = (f \otimes 1')(1 \otimes g);$$

this states that the \otimes -product of f and g can be rewritten as a composite of expansions of f and g . This completes the proof.

For any category \mathcal{C} , let $\text{Fct}(\mathcal{C})$ be the category whose objects are all functors $F: \mathcal{C}^n \rightarrow \mathcal{C}$, of any multiplicity n , and whose morphisms are all natural transformations $t: F \rightarrow F'$ between two functors of the same multiplicity. The composite of t with $t': F' \rightarrow F''$ is defined in the expected way, for any arguments A_1, \dots, A_n , as

$$(2.5) \quad (t't)(A_1, \dots, A_n) = t'(A_1, \dots, A_n)t(A_1, \dots, A_n).$$

If \mathcal{C} has a \otimes -multiplication, so does $\text{Fct}(\mathcal{C})$; to functors $F: \mathcal{C}^n \rightarrow \mathcal{C}$ and $G: \mathcal{C}^m \rightarrow \mathcal{C}$ construct the functor $F \otimes G: \mathcal{C}^{n+m} \rightarrow \mathcal{C}$ with object function defined for any arguments C_1, \dots, C_{n+m} by

$$(2.6) \quad (F \otimes G)(C_1, \dots, C_{n+m}) = F(C_1, \dots, C_n) \otimes G(C_{n+1}, \dots, C_{n+m}),$$

and with the corresponding definition for the mapping function. To natural transformations $t: F \rightarrow F'$ and $u: G \rightarrow G'$ construct the natural transformation $t \otimes u: F \otimes G \rightarrow F' \otimes G'$ defined for arguments C_i by

$$(2.7) \quad (t \otimes u)(C_1, \dots, C_{n+m}) = t(C_1, \dots, C_n) \otimes u(C_{n+1}, \dots, C_{n+m}).$$

These definitions give $\text{Fct}(\mathcal{C})$ a \otimes -multiplication.

An *iterate* of the functor $\otimes: \mathcal{C}^2 \rightarrow \mathcal{C}$ will mean any functor formed by repeated applications of \otimes -multiplication. More exactly, the set of iterates of \otimes is the smallest set of functors $F: \mathcal{C}^n \rightarrow \mathcal{C}$ which contains the identity functor $1: \mathcal{C} \rightarrow \mathcal{C}$ and with any two functors F and G the functor $F \otimes G$. By $\text{It}_{\otimes}(\mathcal{C})$ we denote the category whose objects are all iterates of \otimes and whose morphisms are all natural transformations between such iterates. Then $\text{It}_{\otimes}(\mathcal{C})$ is a category with a \otimes -multiplication.

3. Higher Associativity Laws. Let \mathcal{C} be a category with a multiplication \otimes and with a natural transformation

$$(3.1) \quad a = a(A, B, C): A \otimes (B \otimes C) \rightarrow (A \otimes B) \otimes C$$

such that each $a(A, B, C)$ has a two-sided inverse $a^{-1}(A, B, C)$ in \mathcal{C} . Call a the *associativity isomorphism*. If the functors F , G , and H are three iterates of \otimes , of multiplicities n , m , and k , respectively, the natural isomorphism

$$(3.2) \quad a = a(F, G, H): F \otimes (G \otimes H) \rightarrow (F \otimes G) \otimes H: \mathcal{C}^{n+m+k} \rightarrow \mathcal{C}$$

given for arguments C_i as

$$a(F(C_1, \dots, C_n), G(C_{n+1}, \dots, C_{n+m}), H(C_{m+n+1}, \dots, C_{m+n+k}))$$

will be called an *instance* of a ; define instances of a^{-1} similarly.

An *iterate* of a will be any morphism in the multiplicative subcategory of $\text{It}_{\otimes}(\mathcal{C})$ generated by all the instances of a and of a^{-1} . By Theorem 2.1,

any iterate of a may be written as a composite of expansions of instances of a . It follows that each iterate of a has a two-sided inverse.

Now call the associativity isomorphism a *coherent* if to each pair $F, F': \mathcal{C}^p \rightarrow \mathcal{C}$ of iterates of the functor \otimes there is at most one iterate of a which is a natural isomorphism $t: F \cong F'$. Thus coherence requires that any two formally different iterates between the same pair of functors be equal. Coherence can also be defined in terms of diagrams: it requires that any diagram with vertices iterates of \otimes and edges expansions of instances of a be commutative. This means, for example, that the diagram

$$(3.5) \quad \begin{array}{ccc} A \otimes (B \otimes (C \otimes D)) & \xrightarrow{a_1} & (A \otimes B) \otimes (C \otimes D) \xrightarrow{a_2} & ((A \otimes B) \otimes C) \otimes D \\ & \downarrow 1 \otimes a_3 & & \uparrow a_5 \otimes 1 \\ & A \otimes ((B \otimes C) \otimes D) & \xrightarrow{a_4} & (A \otimes (B \otimes C)) \otimes D \end{array}$$

must be commutative. Here each a_i is an evident instance of a ; thus

$$a_1 = a(A, B, C \otimes D), \quad a_2 = a(A \otimes B, C, D), \quad a_3 = a(B, C, D)$$

and so on (and similarly in subsequent diagrams, where we will omit the subscripts i in a_i and the corresponding specification as to which instance is involved). Now the one condition (3.5) suffices to insure coherence:

THEOREM 3.1. *In a category \mathcal{C} with a multiplication \otimes , an associativity isomorphism a is coherent if and only if the pentagonal diagram (3.5) is commutative for every quadruple A, B, C, D of objects.*

The proof will be by induction on a suitably defined *rank* ρ for the iterates of \otimes . The rank of the identity functor is defined to be 0, while for functors $F: \mathcal{C}^m \rightarrow \mathcal{C}$ and $G: \mathcal{C}^n \rightarrow \mathcal{C}$ the rank of $F \otimes G$ is defined in terms of the ranks of F and G as

$$\rho(F \otimes G) = \rho(F) + \rho(G) + n - 1.$$

The presence of $n-1$ in this formula insures that a functor F has rank zero precisely then when it is expressed by a formula in which all pairs of parentheses start "in front."

Now consider all iterates F of fixed multiplicity n ; these are given exactly by the different arrangements of parentheses in an n -fold product. Draw the graph with vertices all these iterates and with edges all expansions $a_i, 1 \otimes a_i, 1 \otimes (a_i \otimes 1), (1 \otimes a_i) \otimes 1, \dots$ of instances of a . For $n = 4$, the graph is exactly the figure displayed in (3.5). In general, any path along successive edges in this graph from vertex F to vertex G represents a natural isomorphism $t: F \cong G$; namely, that isomorphism given as a composite of the instances on its edges.

If the edges of a path, taken in order, involve only instances of a (and none of a^{-1}), call the path *directed*. Let $H_{(n)}: \mathcal{C}^n \rightarrow \mathcal{C}$ be that iterate of \otimes which has all pairs of parentheses starting in front, so that $H_{(2)} = \otimes$, $H_{(n+1)} = H_{(n)} \otimes 1$. There is a directed path from any F to $H_{(n)}$; indeed, we may choose such a path in a canonical way, say by moving an outermost parenthesis toward the front. This proves, for every F and G , that there is at least one natural isomorphism $t: F \cong G$; observe that this proof is really just the known proof of the "general associative law" from $a(bc) = (ab)c$.

The essential point is that remaining: A proof that any two paths from F to G yield the same natural isomorphism $F \cong G$. Along an arbitrary path from F to G , join each "vertex" F_i to the "bottom" vertex $H_{(n)}$ by the canonical directed path. A glance at the diagram

$$\begin{array}{ccccccccc} F & \rightarrow & F_1 & \leftarrow & F_2 & \rightarrow & F_3 & \leftarrow & G \\ \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ H_{(n)} & = & H_{(n)} & = & H_{(n)} & = & H_{(n)} & = & H_{(n)} \end{array}$$

indicates that it will suffice to show that any two *directed* paths from an F_i to $H_{(n)}$ will yield the same isomorphism. This we prove by induction on the rank of $F = F_i$, it being immediate for rank 0. Suppose it true for all F_i of smaller rank, and consider two different directed paths starting at F with the two expansions e and f , as in the following figure:

$$\begin{array}{ccc} & F = G \otimes H & \\ e \swarrow & & \searrow f \\ F' & & F'' \\ \vdots \swarrow & & \nwarrow \vdots \\ & M & \\ \downarrow & \vdots & \downarrow \\ H_{(n)} & = & H_{(n)} = H_{(n)}. \end{array}$$

Both e and f decrease the rank. Hence it will suffice to show that one can "rejoin" e and f by directed paths to some common vertex M in such a way that the diamond from F to M is commutative. If $e = f$, take $F' = M = F''$. If $e \neq f$, the functor F , as an iterate of \otimes , can be expressed uniquely as $F = G \otimes H$. Now the edge e represents an expanded instance of a ; it has one of three forms:

- e acts "inside" G ; that is, $e = e' \otimes 1_H$ for some e' ;
- e acts "inside" H ; that is, $e = 1_G \otimes e''$ for some e'' ;
- e is an instance of a ; that is, $e = a(G, K, L)$ and $H = K \otimes L$.

For f there are the same three choices.

Now compare e with f . If both act inside the same G , use the induction assumption on G . If e acts inside G and f inside H , say as $f = 1_G \otimes f''$, use the diamond

$$\begin{array}{ccc}
 & F = G \otimes H & \\
 e = e' \otimes 1 & \swarrow & \searrow f = 1 \otimes f'' \\
 G' \otimes H & & G \otimes H' \\
 1' \otimes f'' & \searrow & \swarrow e' \otimes 1' \\
 & G' \otimes H' &
 \end{array}$$

which commutes in view of (2.4). There remains the case when e , say, is an instance of a , so that F has the form

$$F = G \otimes H = G \otimes (K \otimes L).$$

Since $e \neq f$, f must act inside G or inside H . If f acts inside G as $f' : G \rightarrow G'$, use the diamond

$$\begin{array}{ccc}
 & G \otimes (K \otimes L) & \\
 a & \swarrow & \searrow f' \otimes 1 \\
 (G \otimes K) \otimes L & & G' \otimes (K \otimes L) \\
 (f' \otimes 1) \otimes 1 & \searrow & \swarrow a \\
 & (G \otimes K) \otimes L &
 \end{array}$$

This commutes because a is natural. We are left with f acting inside $H = K \otimes L$ as $f = 1_G \otimes f''$. If it is actually inside K or inside L , we are again done, by naturality. If it is inside neither, f'' must be an instance of a , say with $L = P \otimes Q$ and

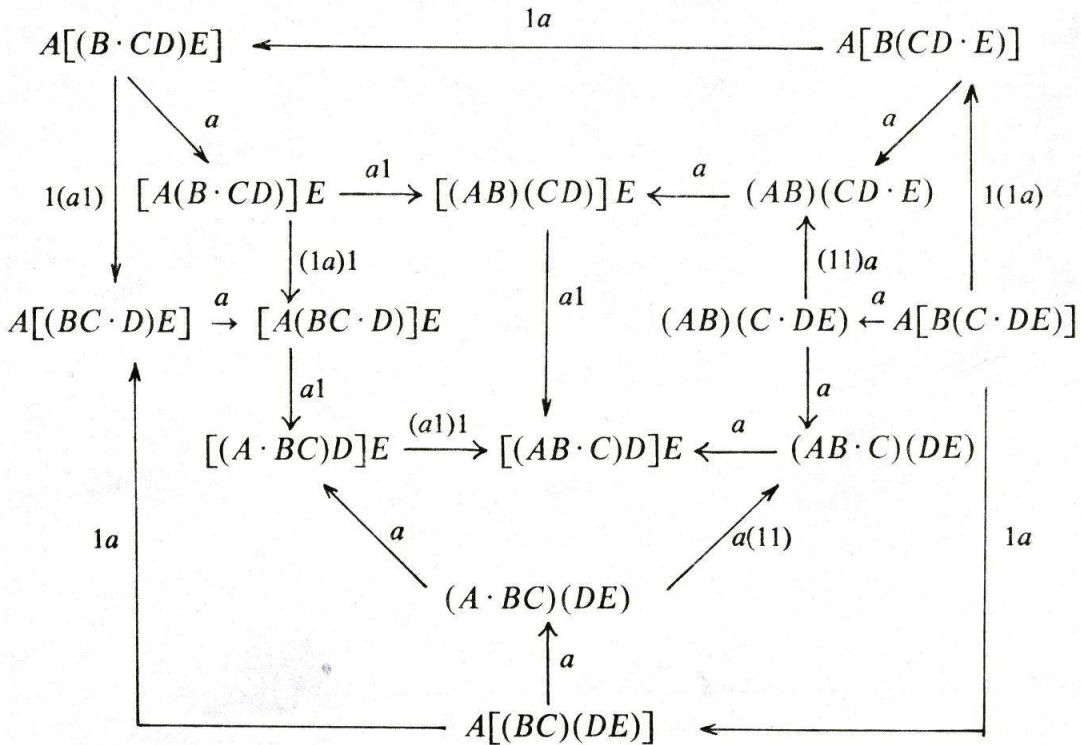
$$f'' = a : K \otimes (P \otimes Q) \rightarrow (K \otimes P) \otimes Q.$$

Here the diamond starts

$$\begin{array}{ccc}
 & G \otimes (K \otimes (P \otimes Q)) & \\
 a & \swarrow & \searrow f = 1 \otimes a \\
 (G \otimes K) \otimes (P \otimes Q) & & G \otimes ((K \otimes P) \otimes Q);
 \end{array}$$

it may be completed by the commutative pentagon of our condition (3.5). This is the final case of the proof. Observe that the diagrams used have been quite analogous to those appearing in one of the familiar proofs of the Jordan-Hölder theorem for groups (cf. [9], 295).

For $n = 5$ the proof may be visualized in the following graph, where AB is short for $A \otimes B$ and $A \cdot BC$ is short for $A(BC)$:



Each complementary region for the graph is either a pentagon (an instance of [3.5]) or a square (an instance of naturality). The whole graph may be regarded as the skeleton of a 3-cell; the regions are its faces and correspond to products $ABCDE$ with two pairs of parentheses omitted, while the edges correspond to products with one pair of parentheses omitted. Stasheff [10] has shown that the corresponding graph for every n gives an $(n - 2)$ -cell.

This coherence result applies in the example of categories with multiplications which we have cited in §2. In the cases of cartesian or tensor products the usual rules

$$(a, (b, c)) \rightarrow ((a, b), c), \quad a \otimes (b \otimes c) \rightarrow (a \otimes b) \otimes c$$

yield associativity isomorphisms which evidently satisfy the pentagon condition, hence are coherent. However, not every associativity isomorphism need satisfy the pentagon condition. A counter example (suggested by Fred Richman) is the isomorphism

$$a \otimes (b \otimes c) \longrightarrow - (a \otimes b) \otimes c$$

in the category of all K -modules.

4. Higher Commutativity Laws. Let \mathcal{C} be a category with a \otimes -multiplication and with a natural homomorphism

$$(4.1) \quad c = c(A, B): A \otimes B \rightarrow B \otimes A$$

such that

$$(4.2) \quad c(B, A) c(A, B) = 1: A \otimes B \rightarrow A \otimes B$$

for every A and B . This insures that $c(B, A)$ is a two-sided inverse for $c(A, B)$. Call c the *commutativity isomorphism*, and assume that \mathcal{C} also has an associativity isomorphism a . Our problem is to describe the ‘‘coherence’’ property for \mathcal{C} under iterates of a , a^{-1} , and c . To this end, we must arrange to permute the arguments of our functor.

To each functor F of multiplicity n and each permutation α in the symmetric group $S(n)$ on n letters we construct a new functor αF of the same multiplicity with the object function

$$(\alpha F)(A_1, \dots, A_n) = F(A_{\alpha_1}, \dots, A_{\alpha_n}), \quad A_i \in \mathcal{C}$$

and the evident corresponding mapping function. Similarly, if $t: F \rightarrow F'$ is a natural transformation between two functors of the same multiplicity n , define $\alpha t: \alpha F \rightarrow \alpha F'$ by

$$(\alpha t)(A_1, \dots, A_n) = t(A_{\alpha_1}, \dots, A_{\alpha_n}), \quad A_i \in \mathcal{C}.$$

If β is also in $S(n)$, then $(\beta\alpha)F = \beta(\alpha F)$ and $(\beta\alpha)t = \beta(\alpha t)$. Moreover, if the composite transformation $t't$ is defined, then

$$(4.3) \quad \alpha(t't) = (\alpha t')(\alpha t).$$

Again, consider natural transformations $t: F \rightarrow F'$ and $u: G \rightarrow G'$ of multiplicities n and m , respectively, so that $t \otimes u$ is defined. Let $\alpha \in S'(n)$ and $\gamma \in S(m)$. Define $\alpha \times \gamma$ to be that permutation of $S(n+m)$ which acts on the first n letters as does α and on the remaining m letters, in order, as does γ . Then the definitions show that $\alpha F \otimes \gamma G = (\alpha \times \gamma)(F \otimes G)$ and that

$$(4.4) \quad \alpha t \otimes \gamma u = (\alpha \times \gamma)(t \otimes u): \alpha F \otimes \gamma G \rightarrow \alpha F' \otimes \gamma G'.$$

A *permuted iterate* (‘‘pit’’ for short) of \otimes will be any functor of the form αF , for $\alpha \in S(n)$ and F an iterate of \otimes of multiplicity n . An *instance* of c is (as before), any natural transformation $c(F, G)$, where the functors F and G are any iterates of \otimes .

DEFINITION. The isomorphisms a and c are *coherent* if every diagram of the following form is commutative: Vertices, permuted instances of \otimes ; edges, permuted expansions of instances of a , a^{-1} , and c .

For example, a special case of this coherence is the commutativity of the following "hexagonal" diagram

$$(4.5) \quad \begin{array}{ccccc} A \otimes (B \otimes C) & \xrightarrow{a} & (A \otimes B) \otimes C & \xrightarrow{c} & C \otimes (A \otimes B) \\ \downarrow 1 \otimes c & & & & \downarrow a \\ A \otimes (C \otimes B) & \xrightarrow{a} & (A \otimes C) \otimes B & \xrightarrow{c \otimes 1} & (C \otimes A) \otimes B. \end{array}$$

Coherence can also be described in more categorical terms. Let $\text{Pit}_{\otimes}(\mathcal{C})$ be the functor category with objects all permuted iterates $H: \mathcal{C}^n \rightarrow \mathcal{C}$ of the functor \otimes and morphisms all natural transformations t between such functors H . Then Pit is a category with a multiplication—and with an action of the groups $S(n)$. Call a subcategory of Pit *symmetric* if it contains with each $t: H \rightarrow H'$ of multiplicity n all αt , for $\alpha \in S(n)$. Consider the symmetric multiplicative subcategory of $\text{Pit}_{\otimes}(\mathcal{C})$ generated by all instances of a , a^{-1} , and c . Then the isomorphisms a and c are coherent if and only if this subcategory contains, for objects H and H' , at most one morphism $t: H \rightarrow H'$. The equivalence of this description of coherence to the previous definition follows from

THEOREM 4.1. *The symmetric multiplicative subcategory of $\text{Pit}_{\otimes}(\mathcal{C})$ generated by any set S of morphisms has as its morphisms all composites of permutations of expansions of morphisms of S .*

Proof. The indicated set of morphisms is clearly closed under composition; we must also show it closed under \otimes -multiplication and under permutation. The latter follows by (4.3), the former by the previous argument for Theorem 2.1 together with the observation that (4.4) turns permutations followed by \otimes into \otimes followed by a suitable permutation.

In the case at hand, when S consists of instances of a , a^{-1} , and c , this theorem shows that every morphism in the corresponding subcategory has a two-sided inverse. In this case we will show that this subcategory contains with any two objects H and H' at least one morphism $t: H \rightarrow H'$. Moreover, the pentagonal and hexagonal conditions suffice for coherence:

THEOREM 4.2. *In a category with a \otimes -multiplication, the associativity and commutativity isomorphisms a and c will be coherent if and only if they satisfy the commutativity conditions given by the pentagon of (3.5) and the hexagon (4.5).*

The result may be restated by saying that three conditions suffice for coherence: the condition $c^2 = 1$ of (4.2) on c alone, the pentagon condition on a alone, and the hexagon condition on a and c together.

For the proof, consider all the permuted iterates of \otimes of some multiplicity n . Write each as αF , for α in $S(n)$ and F an iterate of \otimes , and arrange these αF as “vertices” in $n!$ boxes corresponding to the $n!$ choices of α . Consider the graph formed by the functors as vertices and with edges

$$\square \rightarrow \square \leftarrow \square \rightarrow \dots \leftarrow \square \quad (n! \text{ boxes})$$

all permuted expansions of instances of a and c . Each directed path in this graph corresponds to a (composite) natural transformation between the functors represented by the vertices at the ends of the path. We need show only that each closed path in this graph yields the identity natural transformation. By the pentagon condition and the previous treatment for associativity, we know that the result holds for a closed path staying inside any *one* box. Hence we may concentrate on the portions of the path between the boxes, filling in the path within each box as may be convenient.

The vertices are functors of n arguments, say A_1, \dots, A_n . In the box corresponding to the permutation α , the arguments appear in the order $A_{\alpha_1}, \dots, A_{\alpha_n}$. Each edge between boxes comes from a permuted and expanded instance of the commutativity isomorphism c . Thus the edge $1_{A_2} \otimes c(A_3 \otimes A_5, A_4 \otimes A_1)$ must start at the permutation (23541) in the association $2[(35)(41)]$; it must then exchange the block 35 with the block 41. In general, any such edge will interchange two successive blocks of letters in $A_{\alpha_1}, \dots, A_{\alpha_n}$. But now consider the hexagon condition (4.5). In the top row the instance $c(A \otimes B, C)$ interchanges the block AB with the single letter C ; the hexagon condition states that this interchange may be replaced by two instances of c which interchange single letters with C . Repeated such replacement using instances of the hexagon shows that any interchange of successive blocks may be replaced by interchanges of successive letters, say A_i and A_{i+1} . This means that any closed path can be replaced by one in which each edge between boxes corresponds to the application of one of the transpositions $\sigma_i = (i, i + 1)$, $i = 1, \dots, n - 1$, of successive letters.

Now the symmetric group $S(n)$ is generated by the transpositions σ_i ; hence there exists at least one path between any two vertices (and hence at least one natural isomorphism generated by a and c between any two permuted iterates of \otimes of the same multiplicity). More important, any closed path will correspond to a relation between these generators σ_i of $S(n)$. Now all relations can be represented as products of conjugates of suitable “defining relations,” and the closed paths can be reduced correspondingly. Hence our proof will be complete if we show that each defining relation in these generators gives a closed path which corresponds

to a natural transformation equal to the identity. A classical result ([8], [2]) asserts that a set of defining relations on the generators $\sigma_1, \dots, \sigma_{n-1}$ of $S(n)$ is

$$\begin{aligned} \sigma_i^2 &= 1, & i &= 1, \dots, n-1; \\ (\sigma_i \sigma_{i+1})^3 &= 1, & i &= 1, \dots, n-2; \\ \sigma_i \sigma_j &= \sigma_j \sigma_i, & 1 \leq i < j-1 \leq n-2. \end{aligned}$$

For the first relation $\sigma_i^2 = 1$ we quote the assumed property $c^2 = 1$ of (4.2). For the third relation it suffices to observe the diagram

$$\begin{array}{ccc} (A \otimes B) \otimes (C \otimes D) & \xrightarrow{1 \otimes c'} & (A \otimes B) \otimes (D \otimes C) \\ \downarrow c \otimes 1 & & \downarrow c \otimes 1 \\ (B \otimes A) \otimes (C \otimes D) & \xrightarrow{1 \otimes c'} & (B \otimes A) \otimes (D \otimes C), \end{array}$$

which commutes in view of the property (2.3) of \otimes -multiplication. Finally, consider $(\sigma_1 \sigma_2)^3 = \sigma_1 \sigma_2 \sigma_1 \sigma_2 \sigma_1 \sigma_2$. To draw the corresponding closed path, we must after each transposition σ_i insert an associativity so that the next transposition can indeed be accomplished by the commutativity c . We get the 12-sided polygon written below (with \otimes omitted)

$$\begin{array}{ccccccc} A(BC) & \xrightarrow{1 \cdot c} & A(CB) & \rightarrow & (AC)B & \xrightarrow{c \cdot 1} & (CA)B \\ \downarrow & & & & & & \uparrow \\ (AB)C & \dots & & & & & C(AB) \\ c \cdot 1 \downarrow & & & & & & \downarrow 1 \cdot c \\ (BA)C & \dots & & & & & C(BA) \\ \uparrow & & & & & & \downarrow \\ B(AC) & \xrightarrow{1 \cdot c} & B(CA) & \rightarrow & (BC)A & \xrightarrow{c \cdot 1} & (CB)A \end{array}$$

(the headless arrows are expanded instances of c ; those with heads, of a). We must show this diagram commutative. Insert the two dotted horizontal lines, each an instance of c . The middle rectangle commutes because this c is natural, while the top and bottom are hexagons, both instances of our basic hexagon (4.5). The proof is complete.

This theorem (like the previous one) depends upon a suitable description of the set of all natural transformations $t: F \rightarrow F'$ "generated" by the given isomorphism a , a^{-1} , and c . We have described this set in two ways; first (for the proof) as all compositions of permutations of expansions

of instances of a , a^{-1} , and c , and second (more conceptually) as the morphisms of the symmetric multiplicative subcategory generated by these same instances. There is a third description of this set of transformations; namely, as the smallest set of natural transformations containing a , a^{-1} , and c and closed under composition, tensor product, permutation, and substitution. Here substitution is the process which leads from a to one of its instances. More generally, following Godement [4], we define substitution as follows. Let $t: F \rightarrow F'$ be a natural transformation between functors of multiplicity n . Let G_1, \dots, G_n be n functors (permuted iterates of \otimes) of multiplicities p_1, \dots, p_n . Then *substitution* of the G_i in t yields the natural transformation

$$t*(G_1, \dots, G_n); F(G_1, \dots, G_n) \rightarrow F'(G_1, \dots, G_n)$$

between functors of multiplicity $p_1 + \dots + p_n$, and defined for arguments C_{ij}, \dots , by,

$$t(\dots C_{ij}, \dots) = t(G_1(C_{11}, \dots, C_{1p_1}), \dots, G_n(C_{n1}, \dots, C_{np_n})).$$

5. Higher Identity Laws. Let \mathcal{C} be a category with a \otimes -multiplication and a distinguished object K , called the *ground object*. (This notation is intended to suggest the case when K is a commutative ring and \mathcal{C} the category of all K -modules, so that K is a left and right identity for the usual tensor product).

Assume that there is a natural transformation

$$(5.1) \quad e = e(A): K \otimes A \rightarrow A$$

such that each $e(A)$ has a two-sided inverse in \mathcal{C} ; call e a (left) *identity isomorphism*. An *instance* of e is a natural transformation $e(F)$, where F is any iterate of \otimes . If \mathcal{C} also has associativity and commutativity isomorphisms a and c , we again pose the question of coherence, where coherence means that commutativity holds in every diagram with edges permuted expanded instances of a , a^{-1} , c , e , and e^{-1} . In more detail, the set of functors at issue is the smallest set of functors $\mathcal{C}^n \rightarrow \mathcal{C}$ which contains both the identity functor and the functor $K \otimes A$ of one variable and which is closed under permutation and tensor multiplication of functors. In the category with objects all such functors we consider the symmetric multiplicative subcategory generated by all instances of a , a^{-1} , c , e , and e^{-1} . Just as in Theorem 4.1, this subcategory consists of all composites of permutations of expansions of such instances. Therefore, a , c , and e are coherent if the subcategory contains to any two of its objects at most one morphism $t: H \rightarrow H'$ (it follows readily that it always contains at least one such morphism).

One special case of this coherence condition is

$$(5.2) \quad e(K) = e(K)c(K, K): K \otimes K \rightarrow K;$$

two others are the commutativity of the diagrams (with AB short for $A \otimes B$)

$$(5.3) \quad \begin{array}{ccc} K(BC) & \xrightarrow{a} & (KB)C \\ \downarrow e & & \downarrow e1 \\ BC & = & BC \end{array} \quad \begin{array}{ccc} A(KC) & \xrightarrow{a} & (AK)C \\ \downarrow 1e & & \downarrow c1 \\ AC & \xleftarrow{e1} & (KA)C. \end{array}$$

THEOREM 5.1. *In a category with a \otimes -multiplication and a ground object K , the associativity, commutativity, and identity isomorphisms a , c , and e will be coherent if and only if they satisfy the commutativity conditions given by the three diagrams (5.2) and (5.3), the pentagon (3.5), the hexagon (4.5), and the condition $c^2 = 1$ of (4.2).*

All these necessary conditions hold in the examples of §2 for the usual choices of a , c , and e .

For the proof it is convenient to introduce the "right identity" isomorphism

$$(5.4) \quad e'(A) = e(A)c(A, K): A \otimes K \rightarrow K \otimes A \rightarrow A.$$

Now consider the graph with vertices all functors F at issue, edges all permutations of expansions of instances of a , a^{-1} , c , e , and e^{-1} . In this graph any two vertices can be joined by at least one path, for suitable instances of e and e' can be used to successively remove all factors K . We must prove that any two paths from F to F' correspond to the same natural isomorphism $F \cong F'$. By applying e or e' to F' , we can assume that F' has no factors K . The problem will be reduced to the previous theorem if we show that any path is equivalent to a path which starts by first removing all factors K , and if we show that two different factors K can be removed in either order. This last result is an immediate consequence of (5.2) and naturality. It remains only to show, in effect, that an application of e or of e' after an application of a or of c can be replaced by an application of e or e' first. In many cases this is a consequence of naturality. In other cases, when c is applied first, this is a consequence of the definition (5.4) of e' in terms of e . There remains the case when $a: A(BC) \rightarrow (AB)C$ is followed by e or e' ; this case happens only when one of A , B , or C is K . The result in the first two cases is given by the two cases of commutativity of (5.3). The third case (when $C = K$) requires that the outside rectangle in the following diagram be commutative

$$\begin{array}{ccccc}
 A(BK) & \xrightarrow{a} & & & (AB)K \\
 \downarrow 1c & & & & \downarrow c \\
 A(KB) & \xrightarrow{a} & (AK)B & \cdots \xrightarrow{c1} & (KA)B & \xleftarrow{a} & K(AB) \\
 \downarrow 1e & & & & \downarrow \cdots e1 & & \downarrow e \\
 AB = & = & & = & AB & = & AB.
 \end{array}$$

Fill in the inside dotted arrows as indicated. The top rectangle is then an instance of the hexagon (4.5), while the two bottom squares are the two assumed conditions (5.3). With this argument, the theorem is reduced to the previous cases.

This theorem refers to functors $F: \mathcal{C}^n \rightarrow \mathcal{C}$ of positive multiplicity n . If we include also the constant K as a functor of zero multiplicity, the condition (5.3) must be replaced by the condition that $c(K, K) = 1$.

If commutativity is absent, we must assume two natural isomorphisms

$$(5.5) \quad e(A): K \otimes A \rightarrow A, \quad e'(A): A \otimes K \rightarrow A.$$

In this case, replace (5.2) by

$$(5.6) \quad e(K) = e'(K): K \otimes K \rightarrow K$$

and consider also the commutativities

$$(5.7) \quad \begin{array}{ccc}
 A(KC) & \xrightarrow{1e} & AC \\
 \downarrow a & & \parallel \\
 (AK)C & \xrightarrow{e'1} & AC
 \end{array}, \quad \begin{array}{ccc}
 A(BK) & \xrightarrow{1e'} & AB \\
 \downarrow a & & \parallel \\
 (AB)K & \xrightarrow{e'} & AB
 \end{array}$$

THEOREM 5.2. *In a category with a \otimes -multiplication and a ground object K , the associativity and identity isomorphisms a , e , and e' will be coherent if and only if they satisfy the following five commutativity conditions: The condition (5.6), the pentagon condition (3.5), the first condition of (5.3), and the two conditions (5.7).*

This can be proved by reducing any path to one which first removes all K 's and then applying Theorem 3.1, all much as in the proof of the previous theorem.

6. Tensorred categories. By a *bicategory* we mean a sextuple

$$(\mathcal{C}, \otimes, K, a, c, e),$$

where \mathcal{C} is a category with a multiplication \otimes , K is a selected object of \mathcal{C} , and a, c , and e are associativity, commutativity, and identity isomorphisms

which are coherent; that is, which satisfy the diagrammatic conditions listed in Theorem 5.1. This coherence allows us to "identify" $A \otimes (B \otimes C)$ with $(A \otimes B) \otimes C$, $A \otimes B$ with $B \otimes A$, and $K \otimes A$ with A according to the given isomorphisms a , c , and e and in the fashion familiar, say, for modules. For example, any category with products and with a terminal object T can be regarded as a bicategory (an object T is *terminal* in \mathcal{C} if to each object C of \mathcal{C} there is a unique morphism $C \rightarrow T$ of \mathcal{C}). For, take the product \times as the multiplication \otimes in \mathcal{C} ; this product is naturally associative and commutative, while T satisfies $T \times A \cong A$, and these isomorphisms are readily seen to be coherent.

Bicategories have been introduced independently by several authors. They are in Bénabou [1], with a different but equivalent definition of "coherence," but without any finite list of conditions sufficient for the coherence. In [6] the bicategories are introduced for several purposes: to formulate the notion of a category with a hom functor to some other category, to give a general theory of algebras, and to treat categories of operators and higher homotopies. We will describe briefly each of these objectives.

First, for an arbitrary category \mathcal{C} , hom is a bifunctor

$$\text{hom}: \mathcal{C}^{op} \times \mathcal{C} \rightarrow \mathcal{S}$$

where \mathcal{C}^{op} is the "opposite" or "dual" of the category \mathcal{C} , and \mathcal{S} is the category of sets. Here the composition of homomorphisms is a map of (the product of) sets

$$(6.1) \quad \text{hom}(B, C) \times \text{hom}(A, B) \rightarrow \text{hom}(A, C).$$

In an additive category (defined as usual; see for example [7]), hom is a functor to the category of abelian groups, and composition is a morphism

$$(6.2) \quad \text{hom}(B, C) \otimes \text{hom}(A, B) \rightarrow \text{hom}(A, C)$$

of abelian groups (this is the condition which states that composition and addition of morphisms satisfy the distributive law). In a differential category, as used by Eilenberg-Moore (in unpublished work), hom is a functor to a category of differential graded modules, and composition is a morphism of such modules, much as in (6.2) except that now \otimes denotes the tensor product of such modules. In all three of these cases we have to do with a category whose hom-functor has values in a bicategory (e.g., sets with $\otimes = \times$ or abelian groups with the usual \otimes -multiplication) and whose composition is a morphism as in (6.2).

Second, define a *tensoried category* (cf. [6]) to be a bicategory \mathcal{C} in which

\mathcal{C} is abelian and the functor \otimes is right exact in each of its variables separately. An algebra in \mathcal{C} may thus be defined to be an object Λ of \mathcal{C} together with two morphisms

$$p = p_\Lambda: \Lambda \otimes \Lambda \rightarrow \Lambda, \quad u = u_\Lambda: K \rightarrow \Lambda$$

of \mathcal{C} which represents the usual “product” and “identity element” of Λ . Associativity and the other axioms for an algebra can now be expressed via diagrams in the tensored category \mathcal{C} , and the usual formal properties of Λ -modules and tensor products of Λ -modules and of algebras can be developed (cf. [6]) so as to include all the familiar cases (rings, graded algebras, bigraded algebras, differential graded algebras, and the like).

Finally, consider a bicategory \mathcal{C} whose objects are generated under \otimes -multiplication by a single object B . We can then take the powers of B , in canonical form, to be

$$B^0 = K, \quad B^1 = B, \quad B^{n+1} = B^n \otimes B$$

By coherence, every other object of \mathcal{C} will have a unique isomorphism to some one B^n . Hence we may without loss restrict the objects of the category to the powers B^n ; indeed, we may say that the objects of the category are just the natural numbers n . However, each permutation α in $S(n)$ induces by coherence a unique morphism $B^n \rightarrow B^n$. Thus the bicategory with one generator B may be presented as follows. It is a category \mathcal{H} with objects the natural numbers $\{0, 1, 2, \dots\}$. For each n , the symmetric group $S(n)$ is given as a subgroup of the group of all invertible elements in $\text{hom}(n, n)$; in particular, the identity permutation in $S(n)$ is the identity morphism $n \rightarrow n$. There is also a given bifunctor $\otimes: \mathcal{H} \times \mathcal{H} \rightarrow \mathcal{H}$ with object function

$$m \otimes n = m + n$$

and with mapping function which assigns to $f: m \rightarrow n$ and $g: m' \rightarrow n'$ a morphism $f \otimes g: m + m' \rightarrow n + n'$ satisfying the usual conditions (2.2) and (2.3) for a functor. These structures $S(n)$ and \otimes on \mathcal{H} satisfy the following three axioms. The associative law for \otimes asserts that

$$f \otimes (g \otimes h) = (f \otimes g) \otimes h;$$

the permutation law requires that $\alpha \in S(n)$ and $\gamma \in S(m)$ give

$$\alpha \otimes \gamma = \alpha \times \gamma \in S(n + m)$$

for $\alpha \times \gamma$ defined as in (4.4). Finally, for any m and m' , let $\tau_{m, m'}$ be that permutation in $S(m + m')$ which interchanges the first block of m letters

and the second block of m' letters. For any $f : m \rightarrow n$ and $f' : m' \rightarrow n'$ in \mathcal{H} , the third axiom requires that

$$\tau_{(n,n')}(f \otimes f') = (f' \otimes f) \tau_{(m,m')}.$$

A category with these structures satisfying these axioms is called a **PROP** (short for product and permutation category). Categories of this type have arisen in current studies by J. F. Adams and this author on higher homotopies for cohomology operations. Similar types of categories have arisen in Lawvere's studies in functorial semantics [5]. A number of further results on such categories will be available: for example, it can be demonstrated that there exists a "free" PROP on a given set $g_i : m_i \rightarrow n_i$ of morphisms as free generators. And, as noted above, each bicategory generated by a single object can be represented as a PROP.

To summarize: in a category, the functor hom is part of the formal structure: a *tensoring category* formalizes both of the basic functors hom and \otimes of homological algebra.

REFERENCES

1. Bénabou, Jean, *Catégories avec multiplication*, Comptes Rendue Acad. Sci. Paris 256 (1963), 1887-1890.
2. Burnside, W., *Theory of Groups of Finite Order*, Dover, 1955, Note C.
3. Epstein, D. B. A., *Steenrod operations in abelian categories*, forthcoming.
4. Godement, R., *Théorie des Faisceaux*, Hermann, Paris, 1958.
5. Lawvere, W. V., *Functorial Semantics*, Proc. Nat. Acad. Sci. 50 (1963), 869-872.
6. Mac Lane, S., *Categorical Algebra*, Colloquium Lectures, Am. Math. Soc., 1963.
7. Mac Lane, S., *Homology*, Springer and Academic Press, Heidelberg and New York, 1963.
8. Moore, E. H., *Concerning the abstract group of order $k!$... isomorphic with the symmetric ... substitution group on k letters*, Proc. London Math. Soc. (Ser. 1) 28 (1897), 357-366.
9. Mostow, G. D., Sampson, J. H., and Meyer, J.-P., *Fundamental Structures of Algebra*, McGraw Hill, New York, 1963.
10. Stasheff, J. D., *Homotopy associativity of H-spaces, I*, Trans. Am. Math. Soc. 108 (1963), 275-292.