Graduate Texts in Mathematics 5

Managing Editor: P. R. Halmos

S. Mac Lane

Categories for the Working Mathematician

Springer-Verlag New York Heidelberg Berlin
World Publishing Corporation, Beijing, China

Saunders Mac Lane
Max Mason Distinguished Service Professor of Mathematics, The University of Chicago

AMS Subject Classification (1971)
Primary: 18-02, 18 A XX, 18 C 15, 18 D 10, 18 D 15, 18 E 10, 18 G 30
Secondary: 06-02, 08-02, 08 A 05, 08 A 10, 08 A 15, 08 A 25

Library of Congress Catalog Card Number 78-166080.
Reprinted in China by World Publishing Corporation
For distribution and sale in the People's Republic of China only
只限在中华人民共和国发行
ISBN 0-387-90036-5 Springer-Verlag New York Heidelberg Berlin (soft cover)
ISBN 0-387-90035-7 Springer-Verlag New York Heidelberg Berlin (hard cover)
ISBN 3-540-90036-5 Springer-Verlag Berlin Heidelberg New York (soft cover)
ISBN 7-5062-0035-x World Publishing Corporation China

Preface

Category Theory has developed rapidly. This book aims to present those ideas and methods which can now be effectively used by Mathematicians working in a variety of other fields of Mathematical research. This occurs at several levels. On the first level, categories provide a convenient conceptual language, based on the notions of category, functor, natural transformation, contravariance, and functor category. These notions are presented, with appropriate examples, in Chapters I and II. Next comes the fundamental idea of an adjoint pair of functors. This appears in many substantially equivalent forms: That of universal construction, that of direct and inverse limit, and that of pairs of functors with a natural isomorphism between corresponding sets of arrows. All these forms, with their interrelations, are examined in Chapters III to V. The slogan is "Adjoint functors arise everywhere".

Alternatively, the fundamental notion of category theory is that of a monoid – a set with a binary operation of multiplication which is associative and which has a unit; a category itself can be regarded as a sort of generalized monoid. Chapters VI and VII explore this notion and its generalizations. Its close connection to pairs of adjoint functors illuminates the ideas of universal algebra and culminates in Beck's theorem characterizing categories of algebras; on the other hand, categories with a monoidal structure (given by a tensor product) lead *inter alia* to the study of more convenient categories of topological spaces.

Since a category consists of arrows, our subject could also be described as learning how to live without elements, using arrows instead. This line of thought, present from the start, comes to a focus in Chapter VIII, which covers the elementary theory of abelian categories and the means to prove all the diagram lemmas without ever chasing an element around a diagram.

Finally, the basic notions of category theory are assembled in the last two chapters: More exigent properties of limits, especially of filtered limits, a calculus of "ends", and the notion of Kan extensions. This is the deeper form of the basic constructions of adjoints. We end with the observations that all concepts of category theory are Kan extensions (§ 7 of Chapter X).

I have had many opportunities to lecture on the materials of these chapters: At Chicago; at Boulder, in a series of Colloquium lectures to the American Mathematical Society; at St. Andrews, thanks to the Edinburgh Mathematical Society; at Zurich, thanks to Beno Eckmann and the Forschungsinstitut für Mathematik; at London, thanks to A. Fröhlich and Kings and Queens Colleges; at Heidelberg, thanks to H. Seifert and Albrecht Dold; at Canberra, thanks to Neumann, Neumann, and a Fulbright grant; at Bowdoin, thanks to Dan Christie and the National Science Foundation; at Tulane, thanks to Paul Mostert and the Ford Foundation, and again at Chicago, thanks ultimately to Robert Maynard Hutchins and Marshall Harvey Stone.

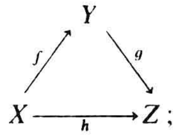
Many colleagues have helped my studies. I have profited much from a succession of visitors to Chicago (made possible by effective support from the Air Force Office of Scientific Research, the Office of Naval Research, and the National Science Foundation): M. André, J. Bénabou, E. Dubuc, F. W. Lawvere, and F. E. J. Linton. I have had good counsel from Michael Barr, John Gray, Myles Tierney, and Fritz Ulmer, and sage advice from Brian Abrahamson, Ronald Brown, W. H. Cockcroft, and Paul Halmos. Daniel Feigin and Geoffrey Phillips both managed to bring some of my lectures into effective written form. My old friend, A. H. Clifford, and others at Tulane were of great assistance. John MacDonald and Ross Street gave pertinent advice on several chapters; Spencer Dickson, S. A. Huq, and Miguel La Plaza gave a critical reading of other material. Peter May's trenchant advice vitally improved the emphasis and arrangement, and Max Kelly's eagle eye caught many soft spots in the final manuscript. I am grateful to Dorothy Mac Lane and Tere Shuman for typing, to Dorothy Mac Lane for preparing the index and to M. K. Kwong for careful proof reading – but the errors which remain, and the choice of emphasis and arrangement, are mine.

Dune Acres, March 27, 1971 Saunders Mac Lane

Table of Contents

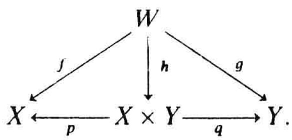
Introduction

Category theory starts with the observation that many properties of mathematical systems can be unified and simplified by a presentation with diagrams of arrows. Each arrow $f : X \to Y$ represents a function; that is, a set X, a set Y, and a rule $x \mapsto fx$ which assigns to each element $x \in X$ an element $fx \in Y$; whenever possible we write fx and not $f(x)$, omitting unnecessary parentheses. A typical diagram of sets and functions is



it is commutative when h is $h = g \circ f$, where $g \circ f$ is the usual composite function $g \circ f : X \to Z$, defined by $x \mapsto g(fx)$. The same diagrams apply in other mathematical contexts; thus in the "category" of all topological spaces, the letters X, Y, and Z represent topological spaces while f, g, and h stand for continuous maps. Again, in the "category" of all groups, X, Y, and Z stand for groups, f, g, and h for homomorphisms.

Many properties of mathematical constructions may be represented by universal properties of diagrams. Consider the cartesian product $X \times Y$ of two sets, consisting as usual of all ordered pairs $\langle x, y \rangle$ of elements $x \in X$ and $y \in Y$. The projections $\langle x, y \rangle \mapsto x$, $\langle x, y \rangle \mapsto y$ of the product on its "axes" X and Y are functions $p : X \times Y \to X$, $q : X \times Y \to Y$. Any function $h : W \to X \times Y$ from a third set W is uniquely determined by its composites $p \circ h$ and $q \circ h$. Conversely, given W and two functions f and g as in the diagram below, there is a unique function h which makes the diagram commute; namely, $hw = \langle fw, gw \rangle$:



Thus, given X and Y, $\langle p, q\rangle$ is "universal" among pairs of functions from some set to X and Y, because any other such pair $\langle f, g\rangle$ factors uniquely (via h) through the pair $\langle p, q\rangle$. This property describes the cartesian product $X \times Y$ uniquely (up to a bijection); the same diagram, read in the category of topological spaces or of groups, describes uniquely the cartesian product of spaces or of the direct product of groups.

Adjointness is another expression for these universal properties. If we write $\hom(W, X)$ for the set of all functions $f : W \to X$ and $\hom(\langle U, V\rangle, \langle X, Y\rangle)$ for the set of all pairs of functions $f : U \to X$, $g : V \to Y$, the correspondence $h \mapsto \langle ph, qh\rangle = \langle f, g\rangle$ indicated in the diagram above is a bijection

$$\hom(W, X \times Y) \cong \hom(\langle W, W\rangle, \langle X, Y\rangle).$$

This bijection is "natural" in the sense (to be made more precise later) that it is defined in "the same way" for all sets W and for all pairs of sets $\langle X, Y\rangle$ (and it is likewise "natural" when interpreted for topological spaces or for groups). This natural bijection involves two constructions on sets: The construction $W \mapsto \langle W, W\rangle$ which sends each set to the diagonal pair $\Delta W = \langle W, W\rangle$, and the construction $\langle X, Y\rangle \mapsto X \times Y$ which sends each pair of sets to its cartesian product. Given the bijection above, we say that the construction $X \times Y$ is a *right adjoint* to the construction Δ, and that Δ is left adjoint to the product. Adjoints, as we shall see, occur throughout mathematics.

The construction "cartesian product" is called a "functor" because it applies suitably to sets *and* to the functions between them; two functions $k : X \to X'$ and $t : Y \to Y'$ have a function $k \times t$ as their cartesian product:

$$k \times t : X \times Y \to X' \times Y', \qquad \langle x, y\rangle \mapsto \langle kx, ty\rangle.$$

Observe also that the one-point set $1 = \{0\}$ serves as an identity under the operation "cartesian product", in view of the bijections

$$1 \times X \xrightarrow{\lambda} X \xleftarrow{\varrho} X \times 1 \tag{1}$$

given by $\lambda\langle 0, x\rangle = x$, $\varrho\langle x, 0\rangle = x$.

The notion of a monoid (a semigroup with identity) plays a central role in category theory. A monoid M may be described as a set M together with two functions

$$\mu : M \times M \to M, \qquad \eta : 1 \to M \tag{2}$$

such that the following two diagrams commute

$$\begin{array}{ccc} M \times M \times M & \xrightarrow{1 \times \mu} & M \times M \\ \downarrow{\scriptstyle \mu \times 1} & & \downarrow{\scriptstyle \mu} \\ M \times M & \xrightarrow{\mu} & M, \end{array} \qquad \begin{array}{ccccc} 1 \times M & \xrightarrow{\eta \times 1} & M \times M & \xleftarrow{1 \times \eta} & M \times 1 \\ \downarrow{\scriptstyle \lambda} & & \downarrow{\scriptstyle \mu} & & \downarrow{\scriptstyle \varrho} \\ M & = & M & = & M\,; \end{array} \tag{3}$$

here 1 in $1 \times \mu$ is the identity function $M \to M$, and 1 in $1 \times M$ is the one-point set $1 = \{0\}$, while λ and ϱ are the bijections of (1) above. To say that these diagrams commute means that the following composites are equal:

$$\mu \circ (1 \times \mu) = \mu \circ (\mu \times 1), \qquad \mu \circ (\eta \times 1) = \lambda, \qquad \mu \circ (1 \times \eta) = \varrho .$$

These diagrams may be rewritten with elements, writing the function μ (say) as a product $\mu(x, y) = xy$ for $x, y \in M$ and replacing the function η on the one-point set $1 = \{0\}$ by its (only) value, an element $\eta(0) = u \in M$. The diagrams above then become

$$\begin{array}{ccc} \langle x, y, z\rangle & \longmapsto & \langle x, yz\rangle \\ \downarrow & & \downarrow \\ \langle xy, z\rangle & \longmapsto & (xy)z = x(yz), \end{array} \qquad \begin{array}{ccc} \langle 0, x\rangle & \longmapsto & \langle u, x\rangle \\ \downarrow & & \downarrow \\ x & = & ux, \end{array} \qquad \begin{array}{ccc} \langle x, u\rangle & \longleftarrow & \langle x, 0\rangle \\ \downarrow & & \downarrow \\ xu & = & x . \end{array}$$

They are exactly the familiar axioms on a monoid, that the multiplication be associative and have an element u as left and right identity. This indicates, conversely, how algebraic identities may be expressed by commutative diagrams. The same process applies to other identities; for example, one may describe a group as a monoid M equipped with a function $\zeta : M \to M$ (of course, the function $x \mapsto x^{-1}$) such that the following diagram commutes

$$\begin{array}{ccccc} M & \xrightarrow{\delta} & M \times M & \xrightarrow{1 \times \zeta} & M \times M \\ \downarrow & & & & \downarrow \mu \\ 1 & & \longrightarrow & & M \end{array} \qquad \begin{array}{ccccc} x & \longmapsto & \langle x, x\rangle & \longmapsto & \langle x, x^{-1}\rangle \\ \downarrow & & & & \downarrow \\ 0 & \longmapsto & u & = & xx^{-1}, \end{array} \qquad (4)$$

here $\delta : M \to M \times M$ is the diagonal function $x \mapsto \langle x, x\rangle$ for $x \in M$, while the unnamed vertical arrow $M \to 1 = \{0\}$ is the evident (and unique) function from M to the one-point set. As indicated just to the right, this diagram does state that ζ assigns to each element $x \in M$ an element x^{-1} which is a right inverse to x.

This definition of a group by arrows μ, η, and ζ in such commutative diagrams makes no explicit mention of group elements, so applies to other circumstances. If the letter M stands for a topological space (not just a set) and the arrows are continuous maps (not just functions), then the conditions (3) and (4) define a topological group – for they specify that M is a topological space with a binary operation μ of multiplication which is continuous (simultaneously in its arguments) and which has a continuous right inverse, all satisfying the usual group axioms. Again, if the letter M stands for a differentiable manifold (of class C^∞) while 1 is the one-point manifold and the arrows μ, η, and ζ are smooth mappings of manifolds, then the diagrams (3) and (4) become the definition of a Lie group. Thus groups, topological groups, and Lie groups can all be described as "diagrammatic" groups in the respective categories of sets, of topological spaces, and of differentiable manifolds.

This definition of a group in a category depended (for the inverse in (4)) on the diagonal map $\delta : M \to M \times M$ to the cartesian square $M \times M$. The definition of a monoid is more general, because the cartesian product $\times$ in $M \times M$ may be replaced by any other operation $\Box$ on two objects which is associative and which has a unit 1 in the sense prescribed by the isomorphisms (1). We can then speak of a monoid in the system $(C, \Box, 1)$, where C is the category, $\Box$ is such an operation, and 1 is its unit. Consider, for example, a monoid M in (**Ab**, $\otimes$, **Z**), where **Ab** is the category of abelian groups, $\times$ is replaced by the usual tensor product of abelian groups, and 1 is replaced by **Z**, the usual group of integers; then (1) is replaced by the familiar isomorphism

$$\mathbf{Z} \otimes X \cong X \cong X \otimes \mathbf{Z}, \qquad X \text{ an abelian group}.$$

Then a monoid M in (**Ab**, $\otimes$, **Z**) is, we claim, simply a ring. For the given morphism $\mu : M \otimes M \to M$ is, by the definition of $\otimes$, just a function $M \times M \to M$, call it multiplication, which is bilinear; i.e., distributive over addition on the left and on the right, while the morphism $\eta : \mathbf{Z} \to M$ of abelian groups is completely determined by picking out one element of M; namely, the image u of the generator 1 of **Z**. The commutative diagrams (3) now assert that the multiplication μ in the abelian group M is associative and has u as left and right unit: – in other words, that M is indeed a ring (with identity = unit).

The (homo)-morphisms of an algebraic system can also be described by diagrams. If $\langle M, \mu, \eta \rangle$ and $\langle M', \mu', \eta' \rangle$ are two monoids, each described by diagrams as above, then a morphism of the first to the second may be defined as a function $f : M \to M'$ such that the following diagrams commute

$$\begin{array}{c} M \\ \downarrow f \\ M', \end{array} \qquad \begin{array}{ccc} M \times M & \xrightarrow{\mu} & M \\ \downarrow f \times f & & \downarrow f \\ M' \times M' & \xrightarrow{\mu'} & M', \end{array} \qquad \begin{array}{ccc} 1 & \xrightarrow{\eta} & M \\ \| & & \downarrow f \\ 1 & \xrightarrow{\eta'} & M'. \end{array} \tag{5}$$

In terms of elements, this asserts that $f(xy) = (fx)(fy)$ and $fu = u'$, with u and u' the unit elements; thus a homomorphism is, as usual, just a function preserving composite and units. If M and M' are monoids in (**Ab**, $\otimes$, **Z**); that is, rings, then a homomorphism f as here defined is just a morphism of rings (preserving the units).

Finally, an *action* of a monoid $\langle M, \mu, \eta\rangle$ on a set S is defined to be a function $\nu : M \times S \to S$ such that the following two diagrams commute

$$\begin{array}{ccc} M \times M \times S & \xrightarrow{1 \times \nu} & M \times S \\ {\scriptstyle \mu \times 1}\downarrow & & \downarrow{\scriptstyle \nu} \\ M \times S & \xrightarrow{\nu} & S\,, \end{array} \qquad \begin{array}{ccc} 1 \times S & \xrightarrow{\eta \times 1} & M \times S \\ & \searrow^{\lambda} & \downarrow{\scriptstyle \nu} \\ & & S\,. \end{array}$$

If we write $\nu(x, s) = x \cdot s$ to denote the result of the action of the monoid element x on the element $s \in S$, these diagrams state just that

$$x \cdot (y \cdot s) = (x\,y) \cdot s\,, \qquad u \cdot s = s$$

for all $x, y \in M$ and all $s \in S$. These are the usual conditions for the action of a monoid on a set, familiar especially in the case of a group acting on a set as a group of transformations. If we shift from the category of sets to the category of topological spaces, we get the usual continuous action of a topological monoid M on a topological space S. If we take $\langle M, \mu, \eta\rangle$ to be a monoid in (**Ab**, $\otimes$, **Z**), then an action of M on an object S of **Ab** is just a left module S over the ring M.

I. Categories, Functors, and Natural Transformations

1. Axioms for Categories

First we describe categories directly by means of axioms, without using any set theory, and calling them "metacategories". Actually, we begin with a simpler notion, a (meta)graph.

A *metagraph* consists of *objects* $a, b, c, \ldots$, *arrows* $f, g, h, \ldots$, and two operations, as follows:

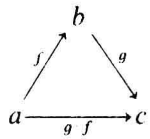
Domain, which assigns to each arrow f an object $a = \operatorname{dom} f$;
Codomain, which assigns to each arrow f an object $b = \operatorname{cod} f$.

These operations on f are best indicated by displaying f as an actual arrow starting at its domain (or "source") and ending at its codomain (or "target"):

$$f: a \to b \quad \text{or} \quad a \xrightarrow{f} b\,.$$

A finite graph may be readily exhibited: Thus $\cdot \to \cdot \to \cdot$ or $\cdot \rightrightarrows \cdot$.

A *metacategory* is a metagraph with two additional operations:
Identity, which assigns to each object a an arrow $\mathrm{id}_a = 1_a : a \to a$;
Composition, which assigns to each pair $\langle g, f \rangle$ of arrows with $\operatorname{dom} g = \operatorname{cod} f$ an arrow $g \circ f$, called their *composite*, with $g \circ f : \operatorname{dom} f \to \operatorname{cod} g$. This operation may be pictured by the diagram



which exhibits all domains and codomains involved. These operations in a metacategory are subject to the two following axioms:

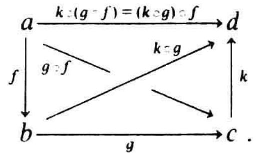
Associativity. For given objects and arrows in the configuration

$$a \xrightarrow{f} b \xrightarrow{g} c \xrightarrow{k} d$$

one always has the equality

$$k \circ (g \circ f) = (k \circ g) \circ f. \tag{1}$$

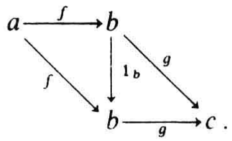
This axiom asserts that the associative law holds for the operation of composition whenever it makes sense (i.e., whenever the composites on either side of (1) are defined). This equation is represented pictorially by the statement that the following diagram is commutative



Unit law. For all arrows $f : a \to b$ and $g : b \to c$ composition with the identity arrow 1_b gives

$$1_b \circ f = f \quad \text{and} \quad g \circ 1_b = g . \tag{2}$$

This axiom asserts that the identity arrow 1_b of each object b acts as an identity for the operation of composition, whenever this makes sense. The Eqs. (2) may be represented pictorially by the statement that the following diagram is commutative:



We use many such diagrams consisting of vertices (labelled by objects of a category) and edges (labelled by arrows of the same category). Such a diagram is *commutative* when, for each pair of vertices c and c', any two paths formed from directed edges leading from c to c' yield, by composition of labels, equal arrows from c to c'. A considerable part of the effectiveness of categorical methods rests on the fact that such diagrams in each situation vividly represent the actions of the arrows at hand.

If b is any object of a metacategory C, the corresponding identity arrow 1_b is uniquely determined by the properties (2). For this reason, it is sometimes convenient to identify the identity arrow 1_b with the object b itself, writing $b : b \to b$. Thus $1_b = b = \mathrm{id}_b$, as may be convenient.

A metacategory is to be any interpretation which satisfies all these axioms. An example is the *metacategory of sets*, which has objects all sets and arrows all functions, with the usual identity functions and the usual composition of functions. Here "function" means a function with specified domain and specified codomain. Thus a function $f : X \to Y$ consists of a set X, its domain, a set Y, its codomain, and a rule $x \mapsto fx$ (i.e., a suitable set of ordered pairs $\langle x, fx \rangle$) which assigns, to each element $x \in X$, an element $fx \in Y$. These values will be written as fx, f_x, or $f(x)$, as may be convenient. For example, for any set S, the assignment $s \mapsto s$ for all $s \in S$ describes the *identity function* $1_S : S \to S$; if S is a subset of Y, the assignment $s \mapsto s$ also describes the *inclusion* or *insertion function* $S \to Y$; these functions are *different* unless $S = Y$. Given functions $f : X \to Y$ and $g : Y \to Z$, the *composite* function $g \circ f : X \to Z$ is defined by $(g \circ f)x = g(fx)$ for all $x \in X$. Observe that $g \circ f$ will mean first apply f, then g – in keeping with the practice of writing each function f to the left of its argument. Note, however, that many authors use the opposite convention.

To summarize, the metacategory of all sets has as objects, all sets, as arrows, all functions with the usual composition. The metacategory of all groups is described similarly: Objects are all groups G, H, K; arrows are all those functions f from the set G to the set H for which $f : G \to H$ is a homomorphism of groups. There are many other metacategories: All topological spaces with continuous functions as arrows; all compact Hausdorff spaces with the same arrows; all ringed spaces with their morphisms, etc. The arrows of any metacategory are often called its *morphisms*.

Since the objects of a metacategory correspond exactly to its identity arrows, it is technically possible to dispense altogether with the objects and deal only with arrows. The data for an *arrows-only metacategory* C consist of arrows, certain ordered pairs $\langle g, f \rangle$, called the composable pairs of arrows, and an operation assigning to each composable pair $\langle g, f \rangle$ an arrow $g \circ f$, called their composite. We say "$g \circ f$ is defined" for "$\langle g, f \rangle$ is a composable pair".

With these data one *defines* an identity of C to be an arrow u such that $f \circ u = f$ whenever the composite $f \circ u$ is defined *and* $u \circ g = g$ whenever $u \circ g$ is defined. The data are then required to satisfy the following three axioms:

(i) The composite $(k \circ g) \circ f$ is defined if and only if the composite $k \circ (g \circ f)$ is defined. When either is defined, they are equal (and this *triple composite* is written as kgf).

(ii) The triple composite kgf is defined whenever both composites kg and gf are defined.

(iii) For each arrow g of C there exist identity arrows u and u' of C such that $u' \circ g$ and $g \circ u$ are defined.

In view of the explicit definition given above for identity arrows, the last axiom is a quite powerful one; it implies that u' and u are unique in (iii), and it gives for each arrow g a codomain u' and a domain u. These axioms are equivalent to the preceding ones. More explicitly, given a metacategory of objects and arrows, its arrows, with the given composition, satisfy the "arrows-only" axioms; conversely, an arrows-only metacategory satisfies the objects-and-arrows axioms when the identity arrows, defined as above, are taken as the objects (Proof as exercise).