I

Introduction

How can you do "new math" problems with an "old math" mind?
——Charlie Brown†

Much of the beauty of mathematics is derived from the fact that it affords abstraction. Not only does it allow one to see the forest rather than the individual trees, but it offers the possibility for study of the structure of the entire forest, in preparation for the next stage of abstraction—comparing forests.

Consider the development of modern group theory and topology. Many different groups and topological spaces were studied individually over long periods of time before the general and abstract concepts of "group" or "topological space" were defined. In these cases, by properly abstracting the "essence" of what was common to objects under consideration and by making the proper definitions, new, wider, and in a sense more beautiful, theories emerged. Category theory involves the next level of abstraction—i.e., comparing forests. It allows the comparison of the class of all groups and homomorphisms with the class of all topological spaces and continuous functions, and further, the comparison of these with other classes of structured sets and structure-preserving functions.

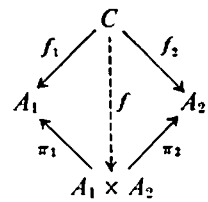
Below we will present four main reasons for the study of categories. In so doing, it is hoped that motivation will also be provided for some of the important abstract notions to be encountered later in the book.

The first reason for studying categories is that, like other mathematical abstractions, category theory provides a new language—a language that affords economy of thought and expression as well as allowing easier communication among investigators in different areas; a language that brings to the surface the common basic ideas underlying various ostensibly unrelated theorems and constructions; and, hence, a language that provides a new context in which to

† From PEANUTS ® by Charles M. Schulz © 1964 United Feature Syndicate, Inc.

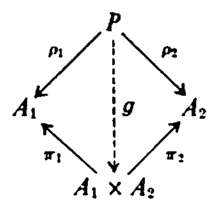
view old problems. Thus, it helps to determine and delineate what the deep, powerful, classical results really are. The need for such a new language can readily be seen after considering the similarities among the following statements:

1. (a) The **cartesian product** $A_1 \times A_2$ of **sets** A_1 and A_2 (together with projection **functions** $\pi_1: A_1 \times A_2 \to A_1$ and $\pi_2: A_1 \times A_2 \to A_2$) has the property that if C is any **set** and $f_1: C \to A_1$ and $f_2: C \to A_2$ are **functions**, then there exists a unique **function** $f: C \to A_1 \times A_2$ such that $\pi_1 \circ f = f_1$ and $\pi_2 \circ f = f_2$, that is, such that the diagram



commutes.

(b) Furthermore, if P is any **set** together with **functions** $\rho_1: P \to A_1$ and $\rho_2: P \to A_2$ having the same universal property as that described in (a) for $A_1 \times A_2$ together with π_1 and π_2, then there exists a **bijection** $g: P \to A_1 \times A_2$ such that $\pi_1 \circ g = \rho_1$ and $\pi_2 \circ g = \rho_2$, that is, such that the diagram



commutes.

2. (a) The **direct product** $A_1 \times A_2$ of **groups** A_1 and A_2 (together with projection **homomorphisms** $\pi_1: A_1 \times A_2 \to A_1$ and $\pi_2: A_1 \times A_2 \to A_2$) has the property that if C is any **group** and $f_1: C \to A_1$ and $f_2: C \to A_2$ are **homomorphisms**, then there exists a unique **homomorphism** $f: C \to A_1 \times A_2$ such that $\pi_1 \circ f = f_1$ and $\pi_2 \circ f = f_2$.

(b) Furthermore, if P is any **group** together with **homomorphisms** $\rho_1: P \to A_1$ and $\rho_2: P \to A_2$ having the same universal property as that described in (a) for $A_1 \times A_2$ together with π_1 and π_2, then there exists an **isomorphism** $g: P \to A_1 \times A_2$ such that $\pi_1 \circ g = \rho_1$ and $\pi_2 \circ g = \rho_2$.

3. (a) The **topological product** $A_1 \times A_2$ of **topological spaces** A_1 and A_2 (with projection **continuous functions** $\pi_1: A_1 \times A_2 \to A_1$ and $\pi_2: A_1 \times A_2 \to A_2$) has the property that if C is any **topological space** and $f_1: C \to A_1$ and $f_2: C \to A_2$ are **continuous functions**, then there exists a unique **continuous function** $f: C \to A_1 \times A_2$ such that $\pi_1 \circ f = f_1$ and $\pi_2 \circ f = f_2$.

(b) Furthermore, if P is any **topological space** together with **continuous functions** $\rho_1: P \to A_1$ and $\rho_2: P \to A_2$ having the same universal property as that described in (a) for $A_1 \times A_2$ together with π_1 and π_2, then there exists a **homeomorphism** $g: P \to A_1 \times A_2$ such that $\pi_1 \circ g = \rho_1$ and $\pi_2 \circ g = \rho_2$.

One immediately notices the following correspondences:

set	$\leftrightarrow$	group	$\leftrightarrow$	topological space
function	$\leftrightarrow$	homomorphism	$\leftrightarrow$	continuous function
bijection	$\leftrightarrow$	isomorphism	$\leftrightarrow$	homeomorphism
cartesian product	$\leftrightarrow$	direct product	$\leftrightarrow$	topological product

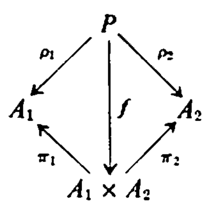
and is tempted to replace the above columns by a single "general" column—

object
morphism
isomorphism
product

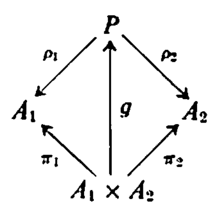
—which is what we are able to do after providing the proper foundations and definitions. Thus, an adequate simultaneous description of all three types of products (and, of course, many more) is made possible.

Moreover, we will be able not only to describe various products simultaneously, but will be able to prove things about them simultaneously as well. For example, the "uniqueness" of the "product" (part (b) of each statement above) is a categorical theorem, the proof of which goes somewhat as follows:

Since each of $(A_1 \times A_2, \pi_1, \pi_2)$ and (P, ρ_1, ρ_2) is "universal", there exist morphisms $f: P \to A_1 \times A_2$ and $g: A_1 \times A_2 \to P$ such that the diagrams

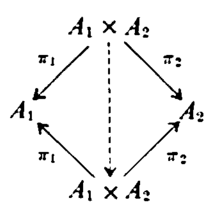


and



commute.

Now because composition of morphisms is associative, each of the morphisms $f \circ g$ and the identity on $A_1 \times A_2$ makes the diagram



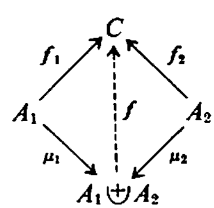
commute $[\pi_i \circ (f \circ g) = (\pi_i \circ f) \circ g = \rho_i \circ g = \pi_i]$. Hence, by the uniqueness property in part (a), $f \circ g$ *is* the identity on $A_1 \times A_2$. Similarly, $g \circ f$ can be shown to be the identity on P. Thus, g is an isomorphism (that is, a bijection for sets, an isomorphism for groups, and a homeomorphism for topological spaces).

Note that this "proof" depends upon two properties that we will later require in the definition of category—the associativity of morphism composition and the existence of an identity morphism associated with every object (see 3.1 and 3.8). It should also be noted that in the definition of the categorical product of two objects, not only the new object, but the attached projection morphisms must be considered. In fact, the projections have a dominant role—how they are composed with other morphisms is the essence of the term "product". The elements of the product objects have an even worse fate than the objects they comprise. They are not considered at all—nor need they be. Similarly in general category theory the main consideration is the morphisms and how they are composed; the objects serve little purpose other than to remind us of the domain and range of the morphisms; and elements of objects are not mentioned at all.

Statements 1, 2, and 3 above can also serve to illustrate the second reason for the study of category theory—the "two for the price of one" or "duality" principle (which will be discussed in more detail in §4). Briefly, it is this: for every categorical concept, there is a dual concept that is obtained by reversing the direction of all morphisms in the description of the original one. As a consequence, every categorical statement or theorem has a dual which (because of symmetry in the definition of category) is true, provided that the original is true. Hence, for categories, every concept is two concepts; every theorem, two theorems; and every proof, two proofs.

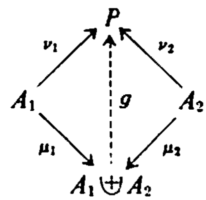
For example, the dual of the notion of "product" discussed above is "coproduct", and the concept of coproduct in the case of the category of sets and functions is "disjoint union". Thus, we have the following:

4. (a) The disjoint union $A_1 \uplus A_2$ of sets A_1 and A_2 (together with injection functions $\mu_1: A_1 \to A_1 \uplus A_2$ and $\mu_2: A_2 \to A_1 \uplus A_2$) has the property that if C is any set and $f_1: A_1 \to C$ and $f_2: A_2 \to C$ are functions, then there exists a unique function $f: A_1 \uplus A_2 \to C$ such that $f \circ \mu_1 = f_1$ and $f \circ \mu_2 = f_2$, that is, such that the diagram



commutes.

(b) Furthermore, if P is any set together with functions $\nu_1: A_1 \to P$ and $\nu_2: A_2 \to P$ having the same universal property as that described in (a) for $A_1 \uplus A_2$ together with μ_1 and μ_2, then there exists a bijection $g: A_1 \uplus A_2 \to P$ such that $g \circ \mu_1 = \nu_1$ and $g \circ \mu_2 = \nu_2$, that is, such that the diagram



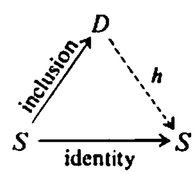
commutes.

Statement 4(b) is a special case of the dual of the uniqueness theorem for the categorical product, which was proved above. Thus, by the duality principle, we know that 4(b) is true. If we specialize the notion of coproduct to groups, modules, or topological spaces, we obtain the concepts "free product", "direct sum", and "topological sum", respectively. All of these are likewise essentially unique by the dual of the categorical theorem given above.

The third reason for studying categories is that by using categorical techniques, difficult problems in some areas of mathematics can in certain cases be translated into easy problems in other areas. Consider, for example, the Brouwer fixed-point theorem, which states that every continuous function from the unit disc into itself has a fixed point. The essential lemma for this theorem—and a result that is difficult to prove in a purely topological setting—is the following:

There exists no continuous function from the unit disc D onto the unit circle S that leaves each point of the circle fixed. Proving this lemma is equivalent to answering the following question in the negative:

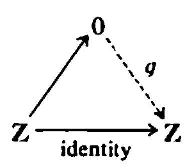
Does there exist a continuous function h: $D \to S$ such that the diagram



commutes?

There exist suitable translation processes from the category of topological spaces to the category of groups. These translations are special cases of what we shall later call functors (see Chapter V). For a particular functor, the translation of the above becomes the following:

Does there exist a group homomorphism g from 0 into **Z** such that the diagram



commutes?

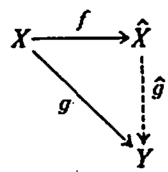
(Where 0 is the trivial group and **Z** is the group of integers.) The answer to the latter question is clearly no, from which we can conclude that the answer to the former question is also no. Thus, the topological lemma can be established via group theory.

Our fourth reason for the study of the theory of categories is that with it one can make precise certain hitherto vague notions—such as the concepts of "universality" and "naturality".

Consider, for example, the following four similar "universal" type theorems:

THEOREM A

For any set X, there exists a group $\hat{X}$ (called the free group generated by X) and a function $f: X \to \hat{X}$ such that for any group Y and any function $g: X \to Y$, there exists a unique group homomorphism $\hat{g}: \hat{X} \to Y$ such that $\hat{g} \circ f = g$, that is, such that the diagram



commutes.

THEOREM B

For any completely regular Hausdorff space X, there exists a compact Hausdorff space $\hat{X}$ (called the Stone-Čech compactification of X) and a continuous function $f: X \to \hat{X}$ such that for any compact Hausdorff space Y and any continuous function $g: X \to Y$, there exists a unique continuous function $\hat{g}: \hat{X} \to Y$ such that $\hat{g} \circ f = g$, that is, such that the diagram of Theorem A commutes.

THEOREM C

For any group X, there exists an abelian group $\hat{X}$ (which is the factor group of X by its commutator subgroup) and a homomorphism $f: X \to \hat{X}$ such that for any abelian group Y and any homomorphism $g: X \to Y$, there exists a unique homomorphism $\hat{g}: \hat{X} \to Y$ such that $\hat{g} \circ f = g$, that is, such that the diagram of Theorem A commutes.

THEOREM D

Let A and B be modules over a commutative ring R and let X denote their set cartesian product. Then there exists an R-module $\hat{X}$ (called the tensor product of A and B) and a bilinear function $f: X \to \hat{X}$ such that for any R-module Y and any bilinear function $g: X \to Y$ there exists a unique linear transformation $\hat{g}: \hat{X} \to Y$ such that $\hat{g} \circ f = g$, that is, such that the diagram of Theorem A commutes.

To explain exactly what these theorems have in common requires a precise definition of universality. Category theory allows this (see Chapter VII); in fact, we will see that all of the above theorems are merely special cases of *one* very general, yet very powerful, theorem—the adjoint functor theorem.

Making precise the notion of "naturality" was actually the original reason for the definition of a category by Eilenberg and Mac Lane.

Consider the following example that motivated their work:

Let V be a finite dimensional vector space over the real numbers **R**. From V we can form the set $\hat{V}$ of all linear functionals from V into **R**. Using the usual pointwise definition of addition of functionals and multiplication by constant functionals, $\hat{V}$ also becomes a vector space with the same dimension as V. Thus, from the theory of finite dimensional vector spaces, we know that V and $\hat{V}$ are isomorphic. Likewise, since $\hat{\hat{V}}$ (that is, the vector space of R-linear functionals over $\hat{V}$) has the same dimension as $\hat{V}$ and V, we know that V and $\hat{\hat{V}}$ are isomorphic. However, there is a fundamental difference between these two situations. There is a "natural" isomorphism between V and $\hat{\hat{V}}$, but there is no natural isomorphism between V and $\hat{V}$. What is meant by natural? Recall that if V and V' are finite dimensional vector spaces over **R** and $F\colon V \to V'$ is a linear transformation between them, then there is an induced linear transformation $\hat{F}\colon \hat{V}' \to \hat{V}$ defined by $\hat{F}(g) = g \circ F$; likewise, there is an induced linear transformation $\hat{\hat{F}}\colon \hat{\hat{V}} \to \hat{\hat{V}}'$. We will now define the natural isomorphism $\eta_V\colon V \to \hat{\hat{V}}$: for each $x \in V$, let $\eta_V(x)$ be that linear functional on $\hat{V}$ whose value at any $g \in \hat{V}$ is $g(x)$; that is, $(\eta_V(x))(g) = g(x)$. Notice the following:

(1) η_V is defined without resorting to choosing a basis for V.

(2) There is simultaneously defined an entire class of isomorphisms, one for each finite dimensional vector space, which are "connected" in the following way: if $F\colon V \to V'$ is any linear transformation, then the diagram

$$\begin{array}{ccc} V & \xrightarrow{F} & V' \\ {\scriptstyle \eta_V}\downarrow & & \downarrow{\scriptstyle \eta_{V'}} \\ \hat{\hat{V}} & \xrightarrow{\hat{\hat{F}}} & \hat{\hat{V}}' \end{array}$$

commutes.

For vector spaces, conditions (1) or (2) or both could be chosen as the definition of "natural". However, it is evident that (1) is a special condition for vector spaces, which cannot be easily abstracted to a categorical situation, whereas (2) is a condition that can be generalized without difficulty. The abstraction of (2) is what is actually used, then, as the definition of "natural" (see §13).

It is now easy to show that there is no "natural" isomorphism between each finite dimensional vector space and its space of linear functionals; for if $\varepsilon_V\colon V \to \hat{V}$ were such an isomorphism (for each space V) and $F\colon U \to U'$ a linear transformation taking all of the vector space U onto the 0 in U', then the commutativity of the diagram

$$\begin{array}{ccc} U & \xrightarrow{F} & U' \\ {\scriptstyle \varepsilon_U}\downarrow & & \downarrow{\scriptstyle \varepsilon_{U'}} \\ \hat{U} & \xleftarrow{\hat{F}} & \hat{U}' \end{array}$$

would force the isomorphism ε_U to take all of U onto the 0 of $\hat{U}$.

A situation analogous to that which has been described above for vector spaces occurs in the case of finite abelian groups. Each such group is isomorphic in a "non-natural" way with its group of characters, but is "naturally" isomorphic with the character group of its character group.

"Natural" isomorphisms also arise in analyzing each of the "universal" Theorems A, B, C, and D above. Properly defining and describing "natural" and "universal" and the relationship between these concepts is one major accomplishment of category theory.