# An Introduction
# to
# Category Theory

by Steven Roman
www.sroman.com

To Donna

# Preface

The purpose of this book is to provide an introduction to the *basic language* of category theory. It is intended for the graduate student, advanced undergraduate student, non specialist mathematician or scientist working in a need-to-know area. The treatment is abstract in nature, with examples drawn mainly from abstract algebra.

## Motivation

Category theory is a relatively young subject, founded in the mid 1940's, with the lofty goals of *unification*, *clarification* and *efficiency* in mathematics.

Indeed, Saunders Mac Lane, one of the founding fathers of category theory (along with Samuel Eilenberg), says in the first sentence of his book *Categories for the Working Mathematician*: “Category theory starts with the observation that many properties of mathematical systems can be unified and simplified [clarified!] by a presentation with diagrams of arrows.” Of course, unification and simplification are common themes throughout mathematics.

To illustrate these concepts, consider the set $\mathbb{R}^*$ of nonzero real number under multiplication, the set $\mathcal{M} = \mathcal{M}(n,k)$ of $n \times k$ matrices over the complex numbers under addition and the set $\mathcal{B}$ of bijections of the integers under composition. Very few mathematicians would take the time to prove that inverses in each of these sets are unique—They would simply note that each of these is an example of a *group* and prove in one quick line that the inverse of any “element” in a group is unique, to wit, if $\alpha$ and $\beta$ are inverses for the group element $a$, then

$$\alpha = \alpha 1 = \alpha(a\beta) = (\alpha a)\beta = 1\beta = \beta$$

This at once *clarifies* the role of uniqueness of inverses by showing that this property has *nothing whatever to do with real numbers, matrices or bijections*. It has to do only with associativity and the identity property itself. This also *unifies* the concept of uniqueness of inverses because it shows that uniqueness of inverses in each of these three cases is really a single concept. Finally, it makes life more *efficient* for mathematicians because they can prove uniqueness of inverse for *all* examples of groups *in one fell swoop*, as it were.

Now, this author knows from over 40 years of experience teaching mathematics that the clarifying, unifying, economizing concept of a group is far too abstract for most lay persons (non mathematicians) as well as for many undergraduate students of mathematics (and alas even some graduate students). However, at the same time, the concept of a group is a most natural, hardly-abstract-at-all concept for most mathematicians and a great many others, such as many physicists, for example.

Now, category theory attempts to do the same for *all* of mathematics (perhaps a bit of a hyperbole) as group theory does for the cases described above. However, for various reasons, even a great many mathematicians find category theory to be too abstract for general comprehension. Perhaps one reason for this is that category theory is not introduced to students in any natural way (pardon the pun). To be more specific, a natural way to introduce category theory is slowly, in small bites, in beginning graduate classes in algebra, logic, topology, geometry and so on. For it would seem that plunging most students into a full-fledged course in category theory designed to be as comprehensive as our common courses in algebra, logic, topology and so on is simply too much abstraction at one time for all but those who are ordained by the gods to be among our most abstract thinkers. The motto for teaching category theory should be "easy does it at first."

Hence this book.

## The Five Concepts of Category Theory

It can be said that there are five *major* concepts in category theory, namely,

- *Categories*
- *Functors*
- *Natural transformations*
- *Universality*
- *Adjoints*

Some would argue that each of these concepts was "invented" or "discovered" in order to produce the next concept in this list. For example, Saunders MacLane himself is reported to have said: "I did not invent category theory to talk about functors. I invented it to talk about natural transformations."

Whether this be true or not, many students of mathematics are finding that the language of category theory is popping up in many of their classes in abstract algebra, lattice theory, number theory, differential geometry, algebraic topology and more. Also, category theory has become an important topic of study for many computer scientists and even for some mathematical physicists. Hopefully, this book will fill a need for those who require an understanding of the *basic* concepts of the subject. If the need or desire should arise, one can then turn to more lengthy and advanced treatments of the subject.

This author believes that one of the major stumbling blocks to gaining a basic understanding of category theory lies in the *notation* and the *terminology* that is most commonly used by authors of the subject, both of which can quickly overwhelm the uninitiated. Accordingly, in this book, both the terminology and the notation are "relaxed" somewhat in an effort to let the reader focus more on the concepts than the language and notation.

## Coverage

The first chapter of the book introduces the definitions of category and functor and discusses diagrams, duality, initial and terminal objects, special types of morphisms and some special types of categories, particularly comma categories and hom-set categories. Chapter 2 is devoted to functors and natural transformations, concluding with Yoneda's lemma.

Chapter 3 introduces the concept of universality and Chapter 4 continues the discussion by introducing cones, limits and the most common categorical constructions: products, equalizers, pullbacks and exponentials (and their duals). The chapter concludes with a theorem on the existence of limits. Chapter 5 is devoted to adjoints and adjunctions.

## Thanks

I would like to thank my students Phong Le, Sunil Chetty, Timothy Choi, Josh Chan, Tim Tran and Zachary Faubion, who attended my lectures on a much expanded version of this book and offered many helpful suggestions.

# Contents

# Chapter 1
# Categories

## Foundations

Before giving the definition of a category, we must briefly (and somewhat informally) discuss a notion from the foundations of mathematics. In category theory, one often wishes to speak of "the category of (all) sets" or "the category of (all) groups." However, it is well known that these descriptions cannot be made precise within the context of sets alone.

In particular, not all "collections" that one can define informally though the use of the English language, or even formally through the use of the language of set theory, can be considered sets without producing some well-known logical paradoxes, such as the Russell paradox of 1901 (discovered by Zermelo a year earlier). More specifically, if $\phi(x)$ is a well-formed formula of set theory, then the collection

$$X = \{\text{sets } x \mid \phi(x) \text{ is true}\}$$

cannot always be viewed as a set. For example, the family of all sets, or of all groups, cannot be considered a set. Nonetheless, it is desirable to be able to apply some of the operations of sets, such as union and cartesian product, to such families. One way to achieve this goal is through the notion of a **class**. Every set is a class and the classes that are not sets are called **proper classes**. Now we can safely speak of the *class* of all sets, or the *class* of all groups. Classes have many of the properties of sets. However, while every set of a set is an element of another set, no class can be an element of another class. We can now state that the family $X$ defined above is a class without apparent contradiction.

Another way to avoid the problems posed by the logical paradoxes is to use the concept of a set $\mathcal{U}$ called a **universe**. The elements of $\mathcal{U}$ are called **small sets**. Some authors refer to the *subsets* of $\mathcal{U}$ as *sets* and some use the term *classes*. In order to carry out "ordinary mathematics" within the universe $\mathcal{U}$, it is assumed to be closed under the basic operations of set theory, such as the taking of ordered pairs, power sets and unions.

These two approaches to the problem of avoiding the logical paradoxes result in essentially the same theory and so we will generally use the language of sets and classes, rather than universes.

## The Definition

We can now give the definition of a category.

**Definition** *A* **category** $\mathcal{C}$ *consists of the following:*

*1*) (**Objects**) *A class* **Obj**$(\mathcal{C})$ *whose elements are called the* **objects**. *It is customary to write* $A \in \mathcal{C}$ *in place of* $A \in$ **Obj**$(\mathcal{C})$.

*2*) (**Morphisms**) *For each (not necessarily distinct) pair of objects* $A, B \in \mathcal{C}$*, a set* $\hom_{\mathcal{C}}(A, B)$*, called the* **hom-set** *for the pair* $(A, B)$. *The elements of* $\hom_{\mathcal{C}}(A, B)$ *are called* **morphisms**, **maps** *or* **arrows** *from* $A$ *to* $B$. *If* $f \in \hom_{\mathcal{C}}(A, B)$*, we also write*

$$f\colon A \to B \quad \text{or} \quad f_{AB}$$

*The object* $A = \mathrm{dom}(f)$ *is called the* **domain** *of* $f$ *and the object* $B = \mathrm{codom}(f)$ *is called the* **codomain** *of* $f$.

3) *Distinct hom-sets are disjoint.*
4) (**Composition**) *For* $f \in \hom_{\mathcal{C}}(A,B)$ *and* $g \in \hom_{\mathcal{C}}(B,C)$ *there is a morphism* $g \circ f \in \hom_{\mathcal{C}}(A,C)$, *called the* **composition** *of* $g$ *with* $f$. *Moreover, composition is associative:*

$$f \circ (g \circ h) = (f \circ g) \circ h$$

*whenever the compositions are defined.*

5) (**Identity morphisms**) *For each object* $A \in \mathcal{C}$ *there is a morphism* $1_A \in \hom_{\mathcal{C}}(A,A)$, *called the* **identity morphism** *for* $A$, *with the property that if* $f_{AB} \in \hom_{\mathcal{C}}(A,B)$ *then*

$$1_B \circ f_{AB} = f_{AB} \quad \textit{and} \quad f_{AB} \circ 1_A = f_{AB}$$

*The class of all morphisms of* $\mathcal{C}$ *is denoted by* **Mor**$(\mathcal{C})$.□

A variety of notations are used in the literature for hom-sets, including

$$(A,B), \quad [A,B], \quad \mathcal{C}(A,B) \quad \text{and} \quad \mathrm{Mor}(A,B)$$

(We will drop the subscript $\mathcal{C}$ in $\hom_{\mathcal{C}}$ when no confusion will arise.)

We should mention that not all authors require property 3) in the definition of a category. Also, some authors permit the hom-sets to be classes. In this case, the categories for which the hom-classes are sets is called a **locally small category**. Thus, all of our categories are locally small. A category $\mathcal{C}$ for which both the class **Obj**$(\mathcal{C})$ and the class **Mor**$(\mathcal{C})$ are sets is called a **small category**. Otherwise, $\mathcal{C}$ is called a **large category**.

Two arrows belonging to the same hom-set $\hom(A,B)$ are said to be **parallel**. We use the phrase "$f$ is a morphism **leaving** $A$" to mean that the domain of $f$ is $A$ and "$f$ is a morphism **entering** $B$" to mean that the codomain of $f$ is $B$.

When we speak of a composition $g \circ f$, it is with the tacit understanding that the morphisms are **compatible**, that is, $\mathrm{dom}(g) = \mathrm{codom}(f)$.

The concept of a category is *very general*. Here are some examples of categories. In most cases, composition is the "obvious" one.

### Example 1

The Category **Set** of Sets
 **Obj** is the class of all sets.
 $\hom(A,B)$ is the set of all functions from $A$ to $B$.

The Category **Mon** of Monoids
 **Obj** is the class of all monoids.
 $\hom(A,B)$ is the set of all monoid homomorphisms from $A$ to $B$.

The Category **Grp** of Groups
 **Obj** is the class of all groups.
 $\hom(A,B)$ is the set of all group homomorphisms from $A$ to $B$.

The Category **AbGrp** of Abelian Groups
 **Obj** is the class of all abelian groups.
 $\hom(A,B)$ is the set of all group homomorphisms from $A$ to $B$.

The Category **Mod**$_R$ of $R$-modules, where $R$ is a ring
 **Obj** is the class of all $R$-modules.

hom$(A, B)$ is the set of all $R$-maps from $A$ to $B$.

The Category **Vect**$_F$ of Vector Spaces over a Field $F$
**Obj** is the class of all vector spaces over $F$.
hom$(A, B)$ is the set of all linear transformations from $A$ to $B$.

The Category **Rng** of Rings
**Obj** is the class of all rings (with unit).
hom$(A, B)$ is the set of all ring homomorphisms from $A$ to $B$.

The Category **CRng** of Commutative Rings with identity
**Obj** is the class of all commutative rings with identity.
hom$(A, B)$ is the set of all ring homomorphisms from $A$ to $B$.

The Category **Field** of Fields
**Obj** is the class of all fields.
hom$(A, B)$ is the set of all ring embeddings from $A$ to $B$.

The Category **Poset** of Partially Ordered Sets
**Obj** is the class of all partially ordered sets.
hom$(A, B)$ is the set of all **monotone** functions from $A$ to $B$, that is, functions $f\colon P \to Q$ satisfying

$$p \leq q \quad \Rightarrow \quad f(p) \leq f(q)$$

The Category **Rel** of relations
**Obj** is the class of all sets.
hom$(A, B)$ is the set of all binary relations from $A$ to $B$, that is, subsets of the cartesian product $A \times B$.

The Category **Top** of Topological Spaces
**Obj** is the class of all topological spaces.
hom$(A, B)$ is the set of all continuous functions from $A$ to $B$.

The Category **SmoothMan** of Manifolds with Smooth Maps
**Obj** is the class of all manifolds.
hom$(A, B)$ is the set of all smooth maps from $A$ to $B$. □

**Example 2** The class $\mathcal{A}$ of *all* categories does not form the class of objects of a category, since otherwise $\mathcal{A}$ would be an element of **Obj**$(\mathcal{A})$, but no class is a member of another class. On the other hand, the class $\mathcal{S}$ of all *small* categories does form the objects of another category, whose morphisms are the *functors*, to be defined a bit later in the chapter. This does not present the same problem as the class of all categories because $\mathcal{S}$ is not small and therefore not a member of $\mathcal{S}$.□

Here are some slightly more unusual categories.

**Example 3** Let $F$ be a field. The category **Matr**$_F$ of matrices over $F$ has objects equal to the set $\mathbb{Z}^+$ of positive integers. For $m, n \in \mathbb{Z}^+$, the hom-set hom$(m, n)$ is the set of all $n \times m$ matrices over $F$, composition being matrix multiplication. Why do we reverse the roles of $m$ and $n$? Well, if $M \in$ hom$(m, n)$ and $N \in$ hom$(n, k)$, then $M$ has size $n \times m$ and $N$ has size $k \times n$ and so the product $NM$ makes sense and has size $k \times m$, that is, it belongs to hom$(m, k)$, as required. Incidentally, this is a case in which the category is named after its morphisms, rather than its objects.□

**Example 4** A single monoid $M$ defines a category with a single object $M$, where each element is a morphism. We define the composition $a \circ b$ to be the product $ab$. This example applies to other algebraic structures, such as groups. All that is required is that there be an identity element and that the operation be associative.□

**Example 5** Let $(P, \leq)$ be a partially ordered set. The objects of the category **Poset**$(P, \leq)$ are the elements of $P$. Also, $\hom(a, b)$ is empty unless $a \leq b$, in which case $\hom(a, b)$ contains a single element, denoted by $ab$. Note that the hom-sets specify the relation $\leq$ on $P$. As to composition, there is really only one choice: If $ab: a \to b$ and $bc: b \to c$ then it follows that $a \leq b \leq c$ and so $a \leq c$, which implies that $\hom(a, c) \neq \emptyset$. Thus, we set $bc \circ ab = ac$. The hom-set $\hom(a, a)$ contains only the identity morphism for the object $a$.

As a specific example, you may recall that each positive natural number $n \in \mathbb{N}$ is defined to be the set of all natural numbers that precede it:

$$n = \{0, 1, \ldots, n - 1\}$$

and the natural number $0$ is defined to be the empty set. Thus, natural numbers are ordered by membership, that is, $m < n$ if and only if $m \in n$ and so $n$ is the set of all natural numbers *less than* $n$. Each natural number $n$ defines a category whose objects are its elements and whose morphisms describe this order relation. The category $n$ is sometimes denoted by bold face $\boldsymbol{n}$.□

**Example 6** A category for which there is *at most one* morphism between every pair of (not necessarily distinct) objects is called a **preordered category** or a **thin category**. If $\mathcal{C}$ is a thin category, then we can use the *existence* of a morphism to define a binary relation on the objects of $\mathcal{C}$, namely, $A \preceq B$ if there exists a morphism from $A$ to $B$. It is clear that this relation is reflexive and transitive. Such relations are called **preorders**. (The term *preorder* is used in a different sense in combinatorics.)

Conversely, any preordered class $(P, \preceq)$ is a category, where the objects are the elements of $P$ and there is a morphism $f_{AB}$ from $A$ to $B$ if and only if $A \preceq B$ (and there are no other morphisms). Reflexivity provides the identity morphisms and transitivity provides the composition.

More generally, if $\mathcal{C}$ is any category, then we can use the *existence* of a morphism to define a preorder on the objects of $\mathcal{C}$, namely, $A \preceq B$ if there is at least one morphism from $A$ to $B$.□

**Example 7** Consider a deductive logic system, such as the propositional calculus. We can define two different categories as follows. In both cases, the well-formed formulas (wffs) of the system are the objects of the category. In one case, there is one morphism from the wff $\alpha$ to the wff $\beta$ if and only if we can deduce $\beta$ given $\alpha$. In the other case, we define a morphism from $\alpha$ to $\beta$ to be a *specific deduction* of $\beta$ from $\alpha$, that is, a specific ordered list of wffs starting with $\alpha$ and ending with $\beta$ for which each wff in the list is either an axiom of the system or is deducible from the previous wffs in the list using the rules of deduction of the system.□

## *The Categorical Perspective*

The notion of a category is extremely general. However, the definition is *precisely* what is needed to set the correct stage for the following two key tenets of mathematics:

1) Morphisms (e.g. linear transformations, group homomorphisms, monotone maps) play an essentially equal role alongside the mathematical structures that they morph (e.g. vector spaces, groups, partially ordered sets).
2) Many mathematical notions are best described in terms of morphisms between structures rather than in terms of the individual elements of these structures.

In order to implement the second tenet, one must grow accustomed to the idea of focusing on the appropriate *maps* between mathematical structures and not on the *elements* of these structures. For example, as we will see in due course, such important notions as a basis for a vector space, a direct product of vector spaces, the field of fractions of an integral domain and the quotient of a group by a normal subgroup can be described using maps rather than elements. In fact, many of the most important properties of these notions follow from their morphism-based descriptions.

Note also that one of the consequences of the second tenet is that important mathematical notions tend to be defined *only up to isomorphism*, rather than uniquely.

An immediate example seems in order, even though it may take some time (and further reading) to place in perspective.

**Example 8** Let $V$ and $W$ be vector spaces over a field $F$. The external direct product of $V$ and $W$ is usually defined in elementary linear algebra books as the set of ordered pairs

$$V \times W = \{(v, w) \mid v \in V, w \in W\}$$

with componentwise operations:

$$(v, w) + (v', w') = (v + v', w + w')$$

and

$$r(v, w) = (rv, rw)$$

for $r \in F$. One then defines the **projection maps**

$$\rho_1\colon V \times W \to V \quad \text{and} \quad \rho_2\colon V \times W \to W$$

by

$$\rho_1(v, w) = v \quad \text{and} \quad \rho_2(v, w) = w$$

However, the importance of these projection maps is not always made clear, so let us do this now.

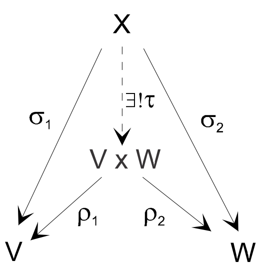


*Figure 1*

As shown in Figure 1, the ordered triple $(V \times W, \rho_V, \rho_W)$ has the following **universal property**: Given any vector space $X$ over $F$ and any "projection-like" pair of linear transformations

$$\sigma_1\colon X \to V \quad \text{and} \quad \sigma_2\colon X \to W$$

from $X$ to $V$ and $W$, respectively, there is a *unique* linear transformation $\tau\colon X \to V \times W$ for which

$$\rho_1 \circ \tau = \sigma_1 \quad \text{and} \quad \rho_2 \circ \tau = \sigma_2$$

Indeed, these two equations uniquely determine $\tau(x)$ for any $x \in X$ because

$$\tau(x) = (\rho_1(\tau(x)), \rho_2(\tau(x))) = (\sigma_1(x), \sigma_2(x))$$

It remains only to show that $\tau$ is linear, which follows easily from the fact that $\sigma_1$ and $\sigma_2$ are linear.

Now, the categorical perspective is that this universal property is the essence of the direct product, at least up to isomorphism. In fact, it is not hard to show that if an ordered triple

$$(U, \lambda_1\colon U \to V, \lambda_2\colon U \to W)$$

has the universal property described above, that is, if for any vector space $X$ over $F$ and any pair of linear transformations

$$\sigma_1 : X \to V \quad \text{and} \quad \sigma_2 : X \to W$$

there is a *unique* linear transformation $\tau: X \to U$ for which

$$\lambda_1 \circ \tau = \sigma_1 \quad \text{and} \quad \lambda_2 \circ \tau = \sigma_2$$

then $U$ and $V \times W$ are isomorphic as vector spaces. Indeed, in some more advanced treatments of linear algebra, the direct product of vector spaces is *defined* as *any* triple that satisfies this universal property. Note that, using this definition, *the direct product is defined only up to isomorphism*.

If this example seems to be a bit overwhelming now, don't be discouraged. It can take a while to get accustomed to the categorical way of thinking. It might help to redraw Figure 1 a few times without looking at the book.□

## Functors

If we are going to live by the two main tenets of category theory described above, we should immediately discuss morphisms between categories! Structure-preserving maps between categories are called *functors*. At this time, however, there is much to say about categories as individual entities, so we will briefly describe functors now and return to them in detail in a later chapter.

The unabridged dictionary defines the term *functor*, from the New Latin *functus* (past participle of *fungi*: to perform) as “something that performs a function or operation.” The term *functor* was apparently first used by the German philosopher Rudolf Carnap (1891–1970) to represent a special type of function sign. In category theory, the term *functor* was introduced by Samuel Eilenberg and Saunders Mac Lane in their paper *Natural Isomorphisms in Group Theory* [8].

Since the structure of a category consists of *both* its objects and its morphisms, a functor should map objects to objects and morphisms to morphisms. This requires two different maps. Also, there are two versions of functors: *covariant* and *contravariant*.

**Definition** *Let $\mathcal{C}$ and $\mathcal{D}$ be categories. A* **functor** $F:\mathcal{C} \Rightarrow \mathcal{D}$ *is a pair of functions (as is customary, we use the same symbol $F$ for both functions):*

*1) The* **object part** *of the functor*

$$F: \mathbf{Obj}(\mathcal{C}) \to \mathbf{Obj}(\mathcal{D})$$

*maps objects in $\mathcal{C}$ to objects in $\mathcal{D}$*

*2) The* **arrow part**

$$F: \mathbf{Mor}(\mathcal{C}) \to \mathbf{Mor}(\mathcal{D})$$

*maps morphisms in $\mathcal{C}$ to morphisms in $\mathcal{D}$ as follows:*

*a) For a* **covariant functor**,

$$F: \hom_{\mathcal{C}}(A, B) \to \hom_{\mathcal{D}}(FA, FB)$$

*for all $A, B \in \mathcal{C}$, that is, $F$ maps a morphism $f: A \to B$ in $\mathcal{C}$ to a morphism $Ff: FA \to FB$ in $\mathcal{D}$.*

*b) For a* **contravariant functor**,

$$F: \hom_{\mathcal{C}}(A, B) \to \hom_{\mathcal{D}}(FB, FA)$$

*for all $A, B \in \mathcal{C}$, that is, $F$ maps a morphism $f: A \to B$ in $\mathcal{C}$ to a morphism $Ff: FB \to FA$ in $\mathcal{D}$. (Note the reversal of direction).*

*We will refer to the restriction of $F$ to $\hom_{\mathcal{C}}(A, B)$ as a* **local arrow part** *of $F$.*

3) *Identity and composition are preserved, that is,*

$$F1_A = 1_{FA}$$

*and for a covariant functor,*

$$F(g \circ f) = Fg \circ Ff$$

*and for a contravariant functor,*

$$F(g \circ f) = Ff \circ Fg$$

*whenever all compositions are defined.*□

As is customary, we use the same symbol $F$ for both the object part and the arrow part of a functor. We will also use a double arrow notation for functors. Thus, the expression $F:\mathcal{C} \Rightarrow \mathcal{D}$ implies that $\mathcal{C}$ and $\mathcal{D}$ are categories and is read "$F$ is a functor from $\mathcal{C}$ to $\mathcal{D}$." (For readability sake in figures, we use a thick arrow to denote functors.)

A functor $F:\mathcal{C} \Rightarrow \mathcal{C}$ from $\mathcal{C}$ to itself is referred to as a **functor on** $\mathcal{C}$. A functor $F:\mathcal{C} \Rightarrow$ **Set** is called a **set-valued functor**. We say that functors $F, G:\mathcal{C} \Rightarrow \mathcal{D}$ with the same domain and the same codomain are **parallel** and functors of the form $F:\mathcal{C} \Rightarrow \mathcal{D}$ and $G:\mathcal{D} \Rightarrow \mathcal{C}$ are **antiparallel**.

The term *covariant* appears to have been first used in 1853 by James Joseph Sylvester (who was quite fond of coining new terms) as follows: "Covariant, a function which stands in the same relation to the primitive function from which it is derived as any of its linear transforms do to a similarly derived transform of its primitive." In plainer terms, an operation is covariant if it varies in a way that preserves some related structure or operation. In the present context, a covariant functor preserves the direction of arrows and a *contravariant* functor reverses the direction of arrows.

One way to view the concept of a functor is to think of a (covariant) functor $F:\mathcal{C} \Rightarrow \mathcal{D}$ as a mapping of one-arrow diagrams in $\mathcal{C}$,

$$A \xrightarrow{f} B$$

to one-arrow diagrams in $\mathcal{D}$,

$$FA \xrightarrow{Ff} FB$$

with the property that "identity loops" and "triangles" are preserved, as shown in Figure 2.

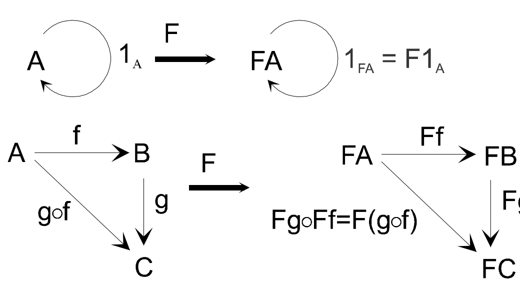


*Figure 2*

A similar statement holds for contravariant functors.

## *Composition of Functors*

Functors can be composed in the "obvious" way. Specifically, if $F:\mathcal{C} \Rightarrow \mathcal{D}$ and $G:\mathcal{D} \Rightarrow \mathcal{E}$ are functors, then $G \circ F:\mathcal{C} \Rightarrow \mathcal{E}$ is defined by

$$(G \circ F)(A) = G(FA)$$

for $A \in \mathcal{C}$ and

$$(G \circ F)(f) = G(Ff)$$

for $f \in \hom_{\mathcal{C}}(A, B)$. We will often write the composition $G \circ F$ as $GF$.

### *Special Types of Functors*

**Definition** *Let $F\colon \mathcal{C} \Rightarrow \mathcal{D}$ be a functor.*

1) *$F$ is* **full** *if all of its the local arrow parts are surjective.*
2) *$F$ is* **faithful** *if all of its local arrow parts are injective.*
3) *$F$ is* **fully faithful** *(i.e., full and faithful) if all of its local arrow parts are bijective.*
4) *$F$ is an* **embedding** *of $\mathcal{C}$ in $\mathcal{D}$ if it is fully faithful and the object part of $F$ is injective.*$\square$

We should note that the term *embedding*, as applied to functors, is defined differently by different authors. Some authors define an embedding simply as a full and faithful functor. Other authors define an embedding to be a faithful functor whose object part is injective. We have adopted the strongest definition, since it applies directly to the important Yoneda lemma (coming later in the book).

Note that a faithful functor $F\colon \mathcal{C} \Rightarrow \mathcal{D}$ need not be an embedding, for it can send two morphisms from *different* hom sets to the same morphism in $\mathcal{D}$. For instance, if $FA = FA'$ and $FB = FB'$ then it may happen that

$$Ff_{AB} = Fg_{A'B'}$$

which does not violate the condition of faithfulness. Also, a full functor need not be surjective on $\mathbf{Mor}(\mathcal{C})$.

### *A Couple of Examples*

Here are a couple of examples of functors. We will give more examples in the next chapter.

**Example 9** The **power set functor** $\wp\colon \mathbf{Set} \Rightarrow \mathbf{Set}$ sends a set $A$ to its power set $\wp(A)$ and sends each set function $f\colon A \to B$ to the induced function $f\colon \wp(A) \to \wp(B)$ that sends $X$ to $fX$. (It is customary to use the same notation for the function and its induced version.) It is easy to see that this defines a faithful but not full covariant functor.

Similarly, the **contravariant power set functor** $F\colon \mathbf{Set} \Rightarrow \mathbf{Set}$ sends a set $A$ to its power set $\wp(A)$ and a set function $f\colon A \to B$ to the induced *inverse* function $f\colon \wp(B) \to \wp(A)$ that sends $X \subseteq B$ to $f^{-1}X \subseteq A$. The fact that $F$ is contravariant follows from the well known fact that

$$(f \circ g)^{-1} = g^{-1} \circ f^{-1} \qquad \square$$

**Example 10** The following situation is quite common. Let $\mathcal{C}$ be a category. Suppose that $\mathcal{D}$ is another category with the property that every object in $\mathcal{C}$ is an object in $\mathcal{D}$ and every morphism $f\colon A \to B$ of $\mathcal{C}$ is a morphism $f\colon A \to B$ of $\mathcal{D}$.

For instance, every object in **Grp** is also an object in **Set**: we simply ignore the group operation. Also, every group homomorphism is a set function. Similarly, every ring can be thought of as an abelian group by ignoring the ring multiplication and every ring map can be thought of as a group homomorphism.

We can then define a functor $F\colon \mathcal{C} \Rightarrow \mathcal{D}$ by sending an object $A \in \mathcal{C}$ to itself, thought of as an object in $\mathcal{D}$ and a morphism $f\colon A \to B$ in $\mathcal{C}$ to itself, thought of as a morphism in $\mathcal{D}$.

Functors such as these that “forget” some structure are called **forgetful functors**. In general, these functors are faithful but not full. For example, distinct group homomorphisms $f, g\colon A \to B$ are also distinct as functions, but not every set function between groups is a group homomorphism.

For any category $\mathcal{C}$ whose objects are sets, perhaps with additional structure and whose morphisms are set functions, also perhaps with additional structure, the “most forgetful” functor is the one that forgets all structure and thinks of an object simply as a set and a morphism simply as a set function. This functor is called the **underlying-set functor** $U: \mathcal{C} \Rightarrow$ **Set** on $\mathcal{C}$.□

## The Category of All Categories

As mentioned earlier, it is tempting to define the category of all categories, but this does not exist. For the collection of all categories must surely be a proper class, being too large to be a set. If this collection formed the objects of a category $\mathcal{C}$, then $\mathcal{C}$ would belong to itself, which is not allowed for a class. In fact, even if $\mathcal{C}$ was a set, then $\mathcal{C} \in \mathcal{C}$ would violate the axiom of regularity, which implies that no set can be a member of itself.

On the other hand, the category **SmCat** of all *small* categories does exist. Its objects are the small categories and its morphisms are the covariant functors between categories. Of course, **SmCat** is a *large* category and so does not belong to itself.

## Concrete Categories

Despite the two main tenets of category theory described earlier, most common categories do have the property that their objects are sets whose elements are "important" and whose morphisms are ordinary set functions on these elements, usually with some additional structure (such as being group homomorphisms or linear transformations). This leads to the following definition.

**Definition** *A category $\mathcal{C}$ is* **concrete** *if there is a faithful functor $F: \mathcal{C} \Rightarrow$* **Set**. *Put more colloquially, $\mathcal{C}$ is concrete if the following hold:*

1) *Each object $A$ of $\mathcal{C}$ can be thought of as a set $FA$ (which is often $A$ itself). Note that distinct objects may be thought of as the same set.*
2) *Each distinct morphism $f: A \to B$ in $\mathcal{C}$ can be thought of as a distinct set function $Ff: FA \to FB$ (which is often $f$ itself).*
3) *The identity $1_A$ morphism can be thought of as the identity set function $F1: FA \to FA$ and the composition $f \circ g$ in $\mathcal{C}$ can be thought of as the composition $Ff \circ Fg$ of the corresponding set functions.*□

Categories that are not concrete are called **abstract categories**. Many concrete categories have the property that $FA$ is $A$ and $Ff$ is $f$. This applies, for example, to most of the previously defined categories, such as **Grp**, **Rng**, **Vect** and **Poset**. The category **Rel** is an example of a category that is not concrete.

In fact, the subject of which categories are concrete and which are abstract can be rather involved and we will not go into it in this introductory book, except to remark that all small categories are concrete, a fact which follows from Yoneda's lemma, to be proved later in the book.

## Subcategories

Subcategories are defined as follows.

**Definition** *Let $\mathcal{C}$ be a category. A* **subcategory** *$\mathcal{D}$ of $\mathcal{C}$ is a category for which the following properties hold:*

1) **Obj**$(\mathcal{D}) \subseteq$ **Obj**$(\mathcal{C})$, *as classes.*
2) *For every $A, B \in \mathcal{D}$,*

$$\hom_{\mathcal{D}}(A, B) \subseteq \hom_{\mathcal{C}}(A, B)$$

*and the identity map $1_A$ in $\mathcal{D}$ is the identity map $1_A$ in $\mathcal{C}$, that is,*

$$(1_A)_{\mathcal{D}} = (1_A)_{\mathcal{C}}$$

2) *Composition in $\mathcal{D}$ is the composition from $\mathcal{C}$, that is, if*

$$f\colon A \to B \quad \textit{and} \quad g\colon B \to C$$

*are morphisms in $\mathcal{D}$, then the $\mathcal{C}$-composite $g \circ f$ is the $\mathcal{D}$-composite $g \circ f$.*
*If equality holds in part 2) for all $A, B \in \mathcal{D}$, then the subcategory $\mathcal{D}$ is* **full**.□

**Example 11** The category **AbGrp** of abelian groups is a full subcategory of the category **Grp**, since the definition of group morphism is independent of whether or not the groups involved are abelian. Put another way, a group homomorphism between abelian groups is just a group homomorphism.

However, the category **Rng** of rings is a *nonfull* subcategory of the category **AbGrp** of abelian groups, since every ring is an additive abelian group but not all additive group homomorphisms $f\colon R \to S$ between rings are ring maps. Similarly, the category of differential manifolds with smooth maps is a nonfull subcategory of the category **Top**, since not all continuous maps are smooth.□

However, the category **AbGrp** of abelian groups is a *nonfull* subcategory of the category **Rng** of rings, since not all additive group homomorphisms $f\colon R \to S$ between rings are ring maps. Similarly, the category of differential manifolds with smooth maps is a nonfull subcategory of the category **Top**, since not all continuous maps are smooth.□

## *The Image of a Functor*

Note that if $F\colon \mathcal{C} \Rightarrow \mathcal{D}$, then the image $F\mathcal{C}$ of $\mathcal{C}$ under the functor $F$, that is, the set

$$\{FA \mid A \in \mathcal{C}\}$$

of objects and the set

$$\{Ff \mid f \in \hom_{\mathcal{C}}(A, B)\}$$

of morphisms need *not* form a subcategory of $\mathcal{D}$. The problem is illustrated in Figure 3.

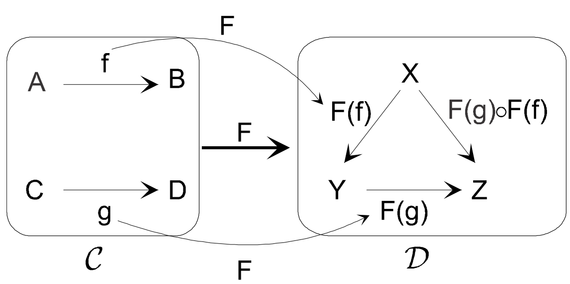


*Figure 3*

In this case, the composition $F(g) \circ F(f)$ is not in the image $F\mathcal{C}$. The only way that this can happen is if the composition $g \circ f$ does not exist because $f$ and $g$ are not compatible for composition. For if $g \circ f$ exists, then

$$F(g) \circ F(f) = F(g \circ f) \in F\mathcal{C}$$

Note that in this example, the object part of $F$ is not injective, since $F(A) = F(C) = X$. This is no coincidence.

**Theorem 12** *If the object part of a functor $F\colon \mathcal{C} \Rightarrow \mathcal{D}$ is injective, then $F\mathcal{C}$ is a subcategory of $\mathcal{D}$, under the composition inherited from $\mathcal{D}$.*
**Proof.** The only real issue is whether the $\mathcal{D}$-composite $Fg \circ Ff$ of two morphisms in $F\mathcal{C}$, when it exists, is also in $F\mathcal{C}$. But this composite exists if and only if

$$Ff\colon FA \to FB \quad \text{and} \quad Fg\colon FB \to FC$$

and so the injectivity of $F$ on objects implies that

$$f: A \to B \quad \text{and} \quad g: B \to C$$

Hence, $g \circ f$ exists in $\mathcal{C}$ and so

$$F(g) \circ F(f) = F(g \circ f) \in F\mathcal{C}$$

□

## Diagrams

The purpose of a *diagram* is to describe a portion of a category $\mathcal{C}$. By "portion" we mean one or more objects of $\mathcal{C}$ along with some of the arrows connecting these objects. let us begin by describing an informal definition of a diagram in a category.

As you may know, a **directed graph** or **digraph** is a set of points, called **nodes**, together with a set of directed line segments, called **arcs**, between (not necessarily distinct) pairs of nodes. An arc from a node to itself is called a **loop**.

As shown in Figure 4, a *diagram* in $\mathcal{C}$ consists of a digraph whose nodes are labeled with objects form $\mathcal{C}$ and whose arcs from the node labeled $A$ to the node labeled $B$ are labeled with morphisms from $A$ to $B$. (In the figure, the nodes are not drawn—only their labels are drawn.)

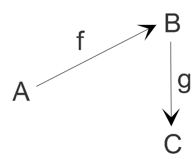


*Figure 4*

Now, this informal definition of a diagram suffices for many purposes. However, we will find it lacking when we define the category of all diagrams of a category $\mathcal{C}$, and for this important purpose, a more formal definition is required. We will give that formal definition now and then connect the formal and informal definitions.

**Definition** *Let $\mathcal{J}$ and $\mathcal{C}$ be categories. A* **diagram** *in $\mathcal{C}$ with* **index category** *$\mathcal{J}$ is a functor $J: \mathcal{J} \Rightarrow \mathcal{C}$.*□

Often, the index category is a finite category. Since the image $J(\mathcal{J})$ is "indexed" by the objects and morphisms of the index category $\mathcal{J}$, the objects in $\mathcal{J}$ are often denoted by lower case letters such as $m$, $n$, $p$, $q$. Figure 5 illustrates this definition.

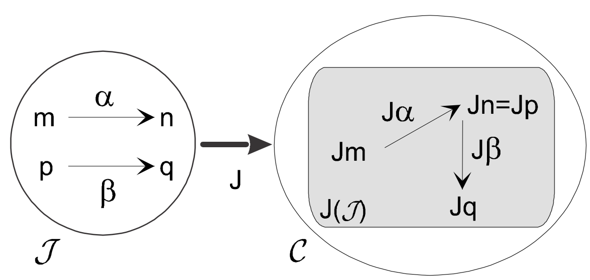


*Figure 5*

As we remarked earlier, $J(\mathcal{J})$ need not be a subcategory of $\mathcal{C}$. In this example, $J$ sends $n$ and $p$ to the same object in $\mathcal{C}$ but since $\alpha$ and $\beta$ are not compatible for composition, the image of $J$ need not contain the composition $J\beta \circ J\alpha$. Thus, the image of a functor simply contains *some* objects of $\mathcal{C}$ as well as *some* morphisms between these objects.

## *The Digraph-Based Version of a Diagram*

To connect this formal definition of a diagram with the informal definition given earlier, let us slowly morph the formal definition. First, we give the formal definition of a labeled digraph, along with some terminology that we will need later in the book.

**Definition**

1) *A* **directed graph** (*or* **digraph**) $\mathcal{D}$ *consists of a nonempty class* $\mathcal{V}(\mathcal{D})$ *of* **vertices** *or* **nodes** *and for every ordered pair* $(v,w)$ *of nodes, a (possibly empty) set* $\mathcal{A}(v,w)$ *of* **arcs** *from* $v$ *to* $w$*. We say that an arc in* $\mathcal{A}(v,w)$ **leaves** $v$ *and* **enters** $w$*. Two arcs from* $v$ *to* $w$ *are said to be* **parallel**. *The arcs from* $v$ *to itself are called* **loops**.
2) *The cardinal number of arcs entering a node is called the* **in-degree** *of the node and the cardinal number of arcs leaving a node is called the* **out-degree** *of the node. The sum of the in-degree and the out-degree is called the* **degree** *of the node.*
3) *A* **labeled digraph** $\mathcal{D}$ *is a digraph for which each node is labeled by elements of a labeling class and each arc is labeled by elements of a labeling class. We require that parallel arcs have distinct labels. A labeled digraph is* **uniquely labeled** *if no two distinct nodes have the same label.*□

A **directed path** (or just **path**) in a labeled digraph $\mathcal{D}$ is a sequence of arcs of the form

$$e_1 \in \mathcal{A}(v_1,v_2), e_2 \in \mathcal{A}(v_2,v_3), \ldots, e_{n-1} \in \mathcal{A}(v_{n-1},v_n)$$

where the ending node of one arc is the starting node of the next arc. The **length** of a path is the number of arcs in the path.

To create what we will call the **digraph version** of a diagram $J\colon \mathcal{J} \Rightarrow \mathcal{C}$, we first draw a digraph whose nodes are labeled with the distinct objects of the index category $\mathcal{J}$ and whose arcs are labeled with the distinct morphisms of $\mathcal{J}$, subject to the obvious condition that the morphism $f\colon A \to B$ labels an arc from the node labeled $A$ to the node labeled $B$. This is referred to as the **underlying digraph** for the diagram. This is shown on the left in Figure 6.

Then, as shown on the right in Figure 6, we further label the nodes and arcs of the digraph with the image of the functor $J$. Note that the labels from the index category $\mathcal{J}$ are distinct, but the labels from $\mathcal{C}$ are not necessarily distinct (in this example, $Jn = Jp$). It is clear that the original diagram $J$ is fully recoverable from the digraph version of the diagram and so the two versions are equivalent. The digraph view of a diagram will be useful when we define morphisms between diagrams.

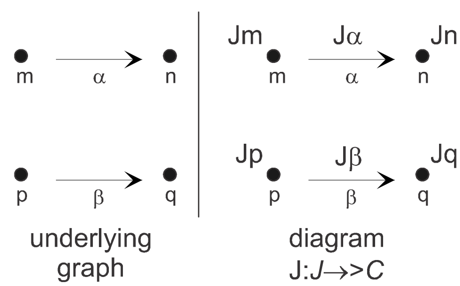


*Figure 6*

Note that if the object part of the diagram functor $J$ is not injective, then two distinct nodes of the underlying graph will be labeled with the same object in $\mathcal{C}$. Although this is useful on occasion (we will use it precisely once), for most applications of diagrams (at least in this book), $J$ is an embedding and so the nodes and arcs are *uniquely* labeled.

Now, since the purpose of the objects and morphisms of the index category is to *uniquely identify* the nodes and arcs of the underlying digraph, once the digraph is drawn on paper, the nodes and arcs are uniquely identified by their location and so the labels from $\mathcal{J}$ are no longer needed. For this reason, they are typically omitted and we arrive at the informal definition of a diagram given earlier. This is why diagrams are often drawn simply as in Figure 4.

We will use blackboard letters $\mathbb{D}, \mathbb{E}, \mathbb{F}, \ldots$ to denote diagrams and if we need to emphasize the functor, we will write

$$\mathbb{D}(J : \mathcal{J} \Rightarrow \mathcal{C})$$

### *Commutative Diagrams*

We consider that any directed path in a diagram is labeled by the *composition* of the morphisms that label the arcs of the path, taken in the reverse order of appearance in the path. For example, the label of the path

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

in Figure 4 is $g \circ f$.

A diagram $\mathbb{D}$ in a category $\mathcal{C}$ is said to **commute** if for every pair $(A, B)$ of objects in $\mathbb{D}$ and any pair of directed paths from $A$ to $B$, *one of which has length at least two*, the corresponding path labels are equal. A diagram that commutes is called a **commuting diagram** or **commutative diagram**.

For example, the diagram in Figure 1 commutes since

$$\rho_1 \circ \tau = \sigma_1 \quad \text{and} \quad \rho_2 \circ \tau = \sigma_2$$

Note that we exempt the case of two parallel paths both having length one so that a diagram such as the one in Figure 7 can be commutative without forcing $f$ and $g$ to be the same morphism. The commutativity condition for this diagram is thus $f \circ e = g \circ e$.

$$\begin{tikzcd} E \arrow[r, "e"] & A \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & B \end{tikzcd}$$

*Figure 7*

## Special Types of Morphisms

Now let us briefly discuss a topic that may not be *de rigueur* among category theorists these days, but seems to this author to be somewhat enlightening for a beginning course in the subject.

For functions, the familiar concepts of *invertibility* (both one-sided and two-sided) and *cancelability* (both one-sided and two-sided) are both categorical concepts. However, the familiar concepts of injectivity and surjectivity are *not* categorical because they involve the *elements* of a set.

In the category **Set**, morphisms are just set functions. For this particular category, the concepts of right-invertibility, right-cancelability and surjectivity are equivalent, as are the concepts of left-invertibility, left-cancelability and injectivity. However, things fall apart totally in arbitrary categories. As mentioned, the concepts of injectivity and surjectivity are not even categorical concepts and so must go away. Moreover, the concepts of invertibility and cancelability are not equivalent in arbitrary categories!

We will explore the relationship between invertibility and cancelability for morphisms in an arbitrary category. In the exercises, we will ask you to explore the relationship between these concepts and the noncategorical concepts of injectivity and surjectivity, when they exist in the context of a particular category.

Let us begin with the formal definitions.

**Definition** *Let $\mathcal{C}$ be a category.*

1) *A morphism* $f\colon A \to B$ *is* **right-invertible** *if there is a morphism* $f_R\colon B \to A$*, called a* **right inverse** *of* $f$*, for which*

$$f \circ f_R = 1_B$$

2) *A morphism* $f\colon A \to B$ *is* **left-invertible** *if there is a morphism* $f_L\colon A \to B$*, called a* **left inverse** *of* $f$*, for which*

$$f_L \circ f = 1_A$$

3) *A morphism* $f\colon A \to B$ *is* **invertible** *or an* **isomorphism** *if there is a morphism* $f^{-1}\colon B \to A$*, called the* (**two-sided**) **inverse** *of* $f$*, for which*

$$f^{-1} \circ f = 1_A \quad \textit{and} \quad f^{-1} \circ f = 1_A$$

*In this case, the objects* $A$ *and* $B$ *are* **isomorphic** *and we write* $A \approx B$.□

Note that the *categorical* term *isomorphism* says nothing about injectivity or surjectivity, for it must be defined in terms of morphisms only!

In fact, this leads to an interesting observation. For categories whose objects are sets and whose morphisms are set functions, we can define an isomorphism in two ways:

1) (Categorical definition) An isomorphism is a morphism with a two-sided inverse.
2) (Non categorical definition) An isomorphism is a bijective morphism.

In most cases of algebraic structures, such as groups, rings or vector spaces, these definitions are equivalent. However, there are cases where only the categorical definition is correct.

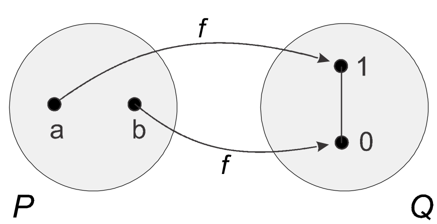


*Figure 8*

For example, as shown in Figure 8, let $P = \{a, b\}$ be a poset in which $a$ and $b$ are incomparable and let $Q = \{0, 1\}$ be the poset with $0 < 1$. Let $f\colon P \to Q$ be defined by $fa = 0$ and $fb = 1$. Then $f$ is a bijective morphism of posets, that is, a bijective monotone map. However, it is not an isomorphism of posets!

Proof of the following familiar facts about inverses is left to the reader.

**Theorem 13**
1) *Two-sided inverses, when they exist, are unique.*
2) *If a morphism is both left and right-invertible, then the left and right inverses are equal and are a* (*two-sided*) *inverse.*
3) *If the composition* $f \circ g$ *of two isomorphisms is defined, then it is an isomorphism as well and*

$$(f \circ g)^{-1} = g^{-1} \circ f^{-1}$$

□

**Definition** *Let* $\mathcal{C}$ *be a category.*
1) *A morphism* $f\colon A \to B$ *is* **right-cancellable** *if*

$$g \circ f = h \circ f \quad \Rightarrow \quad g = h$$

*for any parallel morphisms* $g, h\colon B \to C$*. A right-cancellable morphism is called an* **epic** (*or* **epi**).

2) *A morphism* $f\colon A \to B$ *is* **left-cancellable**, *if*

$$f \circ g = f \circ h \quad \Rightarrow \quad g = h$$

*for any parallel morphisms* $g, h\colon C \to A$. *A left-cancellable morphism is called a* **monic** (*or a* **mono**).□

In general, invertibility is stronger than cancellability. We also leave proof of the following to the reader.

**Theorem 14** *Let* $f, g$ *be morphisms in a category* $\mathcal{C}$.
1) $f$ *left-invertible* $\Rightarrow$ $f$ *left-cancellable* (*monic*)
2) $f$ *right-invertible* $\Rightarrow$ $f$ *right-cancellable* (*epic*)
3) $f$ *invertible* $\Rightarrow$ $f$ *monic and epic.*
*Moreover, the converse implications fail in general.*□

It is also true that a morphism can be both monic and epic (both right and left cancellable) but fail to be an isomorphism. (*Hint*: Think about the more unusual examples of categories.) On the other hand, one-sided cancelability together with one-sided invertibility (on the other side, of course) do imply an isomorphism.

**Theorem 15** *Let* $f\colon A \to B$ *be a morphism in a category* $\mathcal{C}$.
1) *If* $f$ *is monic* (*left-cancellable*) *and right-invertible, then it is an isomorphism.*
2) *If* $f$ *is epic* (*right-cancellable*) *and left-invertible, then it is an isomorphism.*□

## Initial, Terminal and Zero Objects

Anyone who has studied abstract algebra knows that the trivial object (the trivial vector space $\{0\}$, the trivial group $\{1\}$, etc.) often plays a key role in the theory, if only to the point of constantly needing to be excluded from consideration. In general categories, there are actually two concepts related to these trivial or "zero" objects.

**Definition** *Let* $\mathcal{C}$ *be a category.*
1) *An object* $I \in \mathcal{C}$ *is* **initial** *if for every* $A \in \mathcal{C}$, *there is exactly one morphism from* $I$ *to* $A$.
2) *An object* $T$ *is* **terminal** *if for every* $A \in \mathcal{C}$, *there is exactly one morphism from* $A$ *to* $T$.
3) *An object that is both initial and terminal is called a* **zero object**.□

Note that if $C$ is either initial or terminal then $\hom(C, C) = \{1_C\}$. The following simple result is key.

**Theorem 16** *Let* $\mathcal{C}$ *be a category. Any two initial objects in* $\mathcal{C}$ *are isomorphic and any two terminal objects in* $\mathcal{C}$ *are isomorphic.*
**Proof.** If $A$ and $B$ are initial, then there are unique morphisms $f\colon A \to B$ and $g\colon B \to A$ and so $g \circ f \in \hom(A, A) = \{1_A\}$. Similarly, $f \circ g = 1_B$ and so $A \approx B$. A similar proof holds for terminal objects.□

**Example 17** In the category **Set**, the empty set is the only initial object and each singleton-set is terminal. Hence, **Set** has no zero object. In **Grp**, the trivial group $\{1\}$ is a zero object.□

### *Zero Morphisms*

In the study of algebraic structures, one also encounters "zero" functions, such as the zero linear transformation and the map that sends each element of a group $G$ to the identity element of another group $H$. Here is the subsuming categorical concept.

**Definition** *Let* $\mathcal{C}$ *be a category with a zero object* $0$. *Any morphism* $f\colon A \to B$ *that can be factored through the zero object, that is, for which*

$$f = h_{0B} \circ g_{A0}$$

*for morphisms* $h\colon 0 \to B$ *and* $g\colon A \to 0$ *is called a* **zero morphism**.□

To explain this rather strange looking concept, let us take the case of linear algebra, where the zero linear transformation $z: V \to W$ between vector spaces is usually defined to be the map that sends any vector in $V$ to the zero vector in $W$. This definition is not categorical because it involves the zero *element* in $W$. To make it categorical, we interpose the *zero vector space* $\{0\}$. Indeed, the zero transformation $z$ can be written as the composition $z = h \circ g$, where

$$g: V \to \{0\} \quad \text{and} \quad h: \{0\} \to W$$

Here, both $g$ and $h$ are uniquely defined by their domains and ranges, without mention of any elements. The point is that $g$ has no choice but to send every vector in $V$ to the zero vector in $\{0\}$ and $h$ must send the zero vector in $\{0\}$ to the zero vector in $W$. Using $g$ and $h$, we can avoid having to explicitly mention any individual vectors!

In the category of groups, the zero morphisms are precisely the group homomorphisms that map every element of the domain to the identity element of the range. Similar maps exist in **CRng** and **Mod**.

It is clear that any morphism entering or leaving $0$ is a zero morphism.

**Theorem 18** *Let $\mathcal{C}$ be a category with a zero object $0$.*
1) *There is exactly one zero morphism between any two objects in $\mathcal{C}$.*
2) *Zero morphisms "absorb" other morphisms, that is, if $z: A \to B$ is a zero morphism, then so are $f \circ z$ and $z \circ g$, whenever the compositions make sense.* □

## Duality

The concept of duality is prevalent in category theory.

### *Dual or Opposite Categories*

For every category $\mathcal{C}$, we may form a new category $\mathcal{C}^{\text{op}}$, called the **opposite category** or the **dual category** whose objects are the same as those of $\mathcal{C}$, but whose morphisms are "reversed", that is,

$$\hom_{\mathcal{C}^{\text{op}}}(A, B) = \hom_{\mathcal{C}}(B, A)$$

For example, in the category $\mathbf{Set}^{\text{op}}$ the morphisms from $A$ to $B$ are the set functions from $B$ to $A$. This may seem a bit strange at first, but one must bear in mind that morphisms are not necessarily functions in the traditional sense: By definition, they are simply elements of the hom-sets of the category. Therefore, there is no reason why a morphism from $A$ to $B$ cannot be a function from $B$ to $A$.

The rule of composition in $\mathcal{C}^{\text{op}}$, which we denote by $\circ_{\text{op}}$, is defined as follows: If $f \in \hom_{\mathcal{C}^{\text{op}}}(A, B)$ and $g \in \hom_{\mathcal{C}^{\text{op}}}(B, C)$, then

$$g \circ_{\text{op}} f \in \hom_{\mathcal{C}^{\text{op}}}(A, C)$$

is the morphism $f \circ g \in \hom_{\mathcal{C}}(C, A)$. In short,

$$g \circ_{\text{op}} f = f \circ g$$

Note that $(\mathcal{C}^{\text{op}})^{\text{op}} = \mathcal{C}$ and so every category is a dual category.

It might occur to you that we have not really introduced anything *new*, and this is true. Indeed, every category is a dual category (and conversely), since it is dual to its own dual. But we have introduced a new way to look at old things and this will prove fruitful. Stay tuned.

### *The Duality Principle*

Let $p$ be a property that a category $\mathcal{C}$ may possess, for example, $p$ might be the property that $\mathcal{C}$ has an initial object. We say that a property $p^{\text{op}}$ is a **dual property** to $p$ if for all categories $\mathcal{C}$,

$$\mathcal{C} \text{ has } p^{\mathrm{op}} \quad \text{iff} \quad \mathcal{C}^{\mathrm{op}} \text{ has } p$$

Note that this is a symmetric definition and so we can say that two properties are dual (or not dual) to one another. For instance, since the initial objects in $\mathcal{C}^{\mathrm{op}}$ are precisely the terminal objects in $\mathcal{C}$, the properties of having an initial object and having a terminal object are dual. The property of being isomorphic is *self-dual*, that is, $A \approx B$ in $\mathcal{C}$ if and only if $A \approx B$ in $\mathcal{C}^{\mathrm{op}}$.

In general, if $s$ is a statement about a category $\mathcal{C}$, then the **dual statement** is the same statement stated for the dual category $\mathcal{C}^{\mathrm{op}}$, but expressed in terms of the original category. For example, consider the statement

the category $\mathcal{C}$ has an initial object

Stated for the dual category $\mathcal{C}^{\mathrm{op}}$, this is

the category $\mathcal{C}^{\mathrm{op}}$ has an initial object

Since the initial objects in $\mathcal{C}^{\mathrm{op}}$ are precisely the terminal objects in $\mathcal{C}$, this is equivalent to the statement

the category $\mathcal{C}$ has a terminal object

which is therefore the dual of the original statement.

A statement and its dual are not, in general, logically equivalent. For instance, there are categories that have initial objects but not terminal objects. However, for a special and very common type of conditional statement, things are different.

Let $\Pi = \{q_i \mid i \in I\}$ be a set of properties and let $\Pi^{\mathrm{op}} = \{q_i^{\mathrm{op}} \mid i \in I\}$ be the set of dual properties. Let $p$ be a single property. Consider the statement

1) If a category $\mathcal{C}$ has $\Pi$, then it also has $p$ (abbreviated $\Pi \Rightarrow p$).

Since all categories have the form $\mathcal{C}^{\mathrm{op}}$ for some category $\mathcal{C}$, this statement is logically equivalent to the statement

2) If a category $\mathcal{C}^{\mathrm{op}}$ has $\Pi$, then it also has $p$.

and this is logically equivalent to

3) If a category $\mathcal{C}$ has $\Pi^{\mathrm{op}}$, then it also has $p^{\mathrm{op}}$ (abbreviated $\Pi^{\mathrm{op}} \Rightarrow p^{\mathrm{op}}$).

The fact that

$$\Pi \Rightarrow p \quad \text{iff} \quad \Pi^{\mathrm{op}} \Rightarrow p^{\mathrm{op}}$$

is called the **principle of duality** for categories. Note that if $\Pi$ is **self-dual**, that is, if $\Pi = \Pi^{\mathrm{op}}$, then the principle of duality becomes

$$\Pi \Rightarrow p \quad \text{iff} \quad \Pi \Rightarrow p^{\mathrm{op}}$$

Of course, the empty set of properties is self-dual. Moreover, the condition $\emptyset \Rightarrow p$ means that all categories possess property $p$. Hence, we deduce that

> if all categories possess a property $p$, then all categories also possess any dual property $p^{\mathrm{op}}$

For example, all categories possess the property that initial objects (when they exist) are isomorphic. Hence, the principle of duality implies that all terminal objects (when they exist) are isomorphic.

## New Categories From Old Categories

There are many ways to define new categories from old categories. One of the simplest ways is to take the Cartesian product of the objects in two categories. There are also several important ways to turn the morphisms of one category into the objects of another category.

### *The Product of Categories*

If $\mathcal{B}$ and $\mathcal{C}$ are categories, we may form the **product category** $\mathcal{B} \times \mathcal{C}$, in the expected way. Namely, the objects of $\mathcal{B} \times \mathcal{C}$ are the ordered pairs $(B, C)$, where $B$ is an object of $\mathcal{B}$ and $C$ is an object of $\mathcal{C}$. A morphism from $B \times C$ to $B' \times C'$ is a pair $(f, g)$ of morphisms, where $f: B \to B'$ and $g: C \to C'$. Composition is done componentwise:

$$(f, g) \circ (h, k) = (f \circ h, g \circ k)$$

A functor $F: \mathcal{A} \times \mathcal{B} \Rightarrow \mathcal{C}$ from a product category $\mathcal{A} \times \mathcal{B}$ to another category is called a **bifunctor**.

### *The Category of Arrows*

Given a category $\mathcal{C}$, we can form the **category of arrows** $\mathcal{C}^{\to}$ of $\mathcal{C}$ by taking the objects to be the morphisms of $\mathcal{C}$.

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

*Figure 9*

A morphism in $\mathcal{C}^{\to}$, that is, a *morphism between arrows* is defined as follows. A morphism from $f: A \to B$ to $g: A' \to B'$ is a *pair* of arrows

$$(\alpha: A \to A', \beta: B \to B')$$

in $\mathcal{C}$ for which the diagram in Figure 9 commutes, that is, for which

$$g \circ \alpha = \beta \circ f$$

We leave it to the reader to verify that $\mathcal{C}^{\to}$ is a category, with composition defined pairwise:

$$(\gamma, \delta) \circ (\alpha, \beta) = (\gamma \circ \alpha, \delta \circ \beta)$$

and with identity morphisms $(1_A, 1_B)$.

### *Comma Categories*

Comma categories form one of the most important classes of categories. We will define the simplest form of comma category first and then generalize twice.

#### Arrows Entering (or Leaving) an Object

The simplest form of comma category is defined as follows. Let $\mathcal{C}$ be a category and let $A \in \mathcal{C}$. We will refer to $A$ as the **anchor object**. The category of **arrows leaving** $A$, denoted by $(A \to \mathcal{C})$ has for its objects the set of all pairs

$$\{(B, f: A \to B) \mid B \in \mathcal{C}\}$$

Note that since a morphism uniquely determines its codomain, we could define the objects of $(A \to \mathcal{C})$ to be just the morphisms $f: A \to B$ themselves but the present definition, which includes the codomains explicitly, is more traditional.

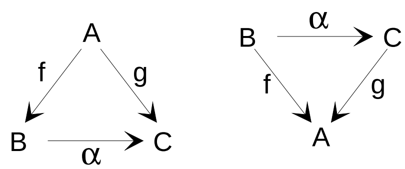


*Figure 10*

As shown on the left in Figure 10, a morphism $\alpha: (B, f) \to (C, g)$ in $(A \to \mathcal{C})$ is just a morphism $\alpha: B \to C$ in $\mathcal{C}$ between the codomains for which the triangle commutes, that is, for which

$$\alpha \circ f = g$$

The category of arrows leaving $A$ is also called a **coslice category**.

Similarly, the category $(\mathcal{C} \to A)$ of **arrows entering** the anchor object $A$ has for its objects the pairs

$$\{(B, f: B \to A) \mid B \in \mathcal{C}\}$$

and as shown on the right in Figure 10, a morphism $\alpha: (B, f) \to (C, g)$ in $(\mathcal{C} \to A)$ is a morphism $\alpha: B \to C$ in $\mathcal{C}$ between the domains for which

$$g \circ \alpha = f$$

The category of arrows entering $A$ is also called a **slice category**.

## The First Generalization

To generalize this one step (see Figure 11), let $F: \mathcal{C} \Rightarrow \mathcal{D}$ be a functor and let $A \in \mathcal{D}$ be the **anchor object**.

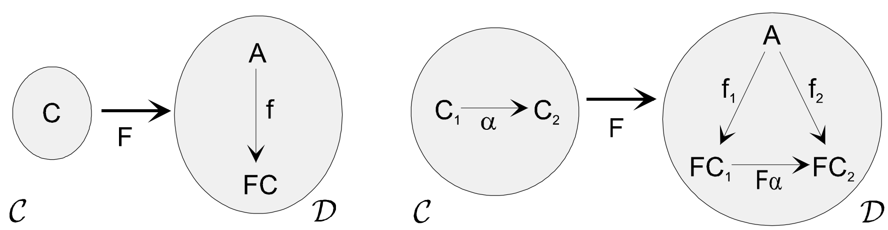


*Figure 11*

As shown on the left in Figure 11, the objects of the comma category $(A \to F)$ are the pairs

$$\{(C, f: A \to FC) \mid C \in \mathcal{C}\}$$

As to morphisms, as shown on the right in Figure 11, if

$$X = (C_1, f_1: A \to FC_1) \quad \text{and} \quad Y = (C_2, f_2: A \to FC_2)$$

are objects in $(A \to F)$, then a morphism $\alpha: X \to Y$ in $(A \to F)$ is a morphism $\alpha: C_1 \to C_2$ in $\mathcal{C}$ with the property that

$$F\alpha \circ f_1 = f_2$$

Note that the comma category $(A \to \mathcal{C})$ defined earlier is just $(A \to I_{\mathcal{C}})$, where where $I_{\mathcal{C}}$ is the identity functor on $\mathcal{C}$. We can also define the comma category $(F \to A)$ by reversing the arrows.

## The Final Generalization

As a final generalization, let $F: \mathcal{B} \Rightarrow \mathcal{D}$ and $G: \mathcal{C} \Rightarrow \mathcal{D}$ be functors with the same codomain. As shown in Figure 12, an object of the **comma category** $(F \to G)$ is a triple

$$(B, C, f\colon FB \to GC)$$

where $B \in \mathcal{B}$, $C \in \mathcal{C}$ and $f$ is a morphism in $\mathcal{D}$.

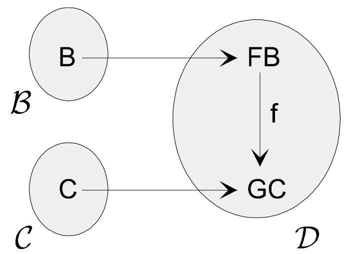


*Figure 12*

As to morphisms, as shown in Figure 13,

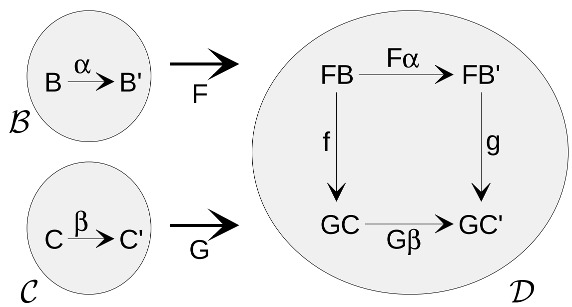


*Figure 13*

a morphism from $(B, C, f\colon FB \to GC)$ to $(B', C', f'\colon FB' \to GC')$ is a pair of morphisms

$$(\alpha\colon B \to B', \beta\colon C \to C')$$

for which the square commutes, that is,

$$G\beta \circ f = f' \circ F\alpha$$

The composition of pairs is done componentwise.

**Example 19** Let $\mathcal{C}$ be a category and let $F\colon \mathcal{C} \Rightarrow$ **Set** be a set-valued functor. The objects of the **category of elements Elts**$(F)$ are ordered pairs $(C, a)$, where $C \in \mathcal{C}$ and $a \in FC$. A morphism $f\colon (C, a) \to (D, b)$ is a morphism $f\colon C \to D$ for which $Ff(a) = b$. We leave it to the reader to show that this is a special type of comma category.□

## *Hom-Set Categories*

Rather than treating individual arrows as the objects of a new category, we can treat entire hom-sets

$$\{\hom_{\mathcal{C}}(A, X) \mid X \in \mathcal{C}\}$$

as the objects of a category $\mathcal{C}(A, -)$. As to the morphisms, referring to the left half of Figure 14, let $\hom_{\mathcal{C}}(A, X)$ and $\hom_{\mathcal{C}}(A, Y)$ be hom-sets. Then for each morphism $f\colon X \to Y$ in $\mathcal{C}$, there is a morphism

$$f^{\leftarrow}\colon \hom_{\mathcal{C}}(A, X) \to \hom_{\mathcal{C}}(A, Y)$$

defined in words as "follow by $f$," that is,

$$f^{\leftarrow}(\alpha) = f \circ \alpha$$

for all $\alpha \in \hom_{\mathcal{C}}(A, X)$.

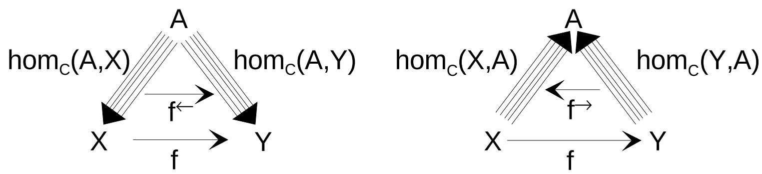


*Figure 14*

We can also define a category $\mathcal{C}(-,A)$ whose objects are

$$\{\hom_{\mathcal{C}}(X,A) \mid X \in \mathcal{C}\}$$

As shown on the right half of Figure 14, for each morphism $f\colon X \to Y$ in $\mathcal{C}$, there is a morphism in $\mathcal{C}(-,A)$ from $\hom_{\mathcal{C}}(Y,A)$ and $\hom_{\mathcal{C}}(X,A)$:

$$f^{\rightarrow}\colon \hom_{\mathcal{C}}(Y,A) \to \hom_{\mathcal{C}}(X,A)$$

defined by "precede by $f$," that is,

$$f^{\rightarrow}(\alpha) = \alpha \circ f$$

Note that any category $\mathcal{C}$ can be viewed as a hom-set category by adjoining a new initial "object" $*$ not in $\mathcal{C}$ and defining a new morphism $f_A\colon * \to A$ from $*$ to each object $A \in \mathcal{C}$. Then each object $A \in \mathcal{C}$ can be identified with its hom-set $\hom(*,A)$. Also, the morphisms $f\colon A \to B$ in $\mathcal{C}$ are identified with the morphisms

$$f^{\leftarrow}\colon \hom(*,A) \to \hom(*,B)$$

of hom-sets.

## Exercises

1. Prove that identity morphisms are unique.
2. If $F\colon \mathcal{C} \Rightarrow \mathcal{D}$ is fully faithful, prove that

$$FC \approx FC' \quad \Rightarrow \quad C \approx C'$$

3. Indicate how one might define a category without mentioning objects.
4. A category with only one object is essentially just a monoid. How?
5. Let $V$ be a real vector space. Define a category $\mathcal{C}$ as follows. The objects of $\mathcal{C}$ are the vectors in $V$. For $u, v \in V$, let

$$\hom(u,v) = \{a \in \mathbb{R} \mid \text{there is } r \geq 1 \text{ such that } rau = v\}$$

   Let composition be ordinary multiplication. Show that $\mathcal{C}$ is a category.
6. a) Prove that the composition of monics is monic.
   b) Prove that if $f \circ g$ is monic, then so is $g$.
   c) Prove that if $f \circ g$ is epic, then so is $f$.
7. Find a category with nonidentity morphisms in which every morphism is monic and epic, but no nonidentity morphism is an isomorphism.
8. Prove that any two initial objects are isomorphic and any two terminal objects are isomorphic.
9. Find the initial, terminal and zero objects in $\mathbf{Mod}_R$ and **CRng**.
10. Find the initial, terminal and zero objects in the following categories:
    a) **Set** $\times$ **Set**
    b) $\mathbf{Set}^{\rightarrow}$
11. In each case, find an example of a category with the given property.
    a) No initial or terminal objects.
    b) An initial object but no terminal objects.
    c) No initial object but a terminal object.
    d) An initial and a terminal object that are not isomorphic.

12. Let $\mathbb{D}$ be a diagram in a category $\mathcal{C}$. Show that there is a smallest subcategory $\mathcal{D}$ of $\mathcal{C}$ for which $\mathbb{D}$ is a diagram in $\mathcal{D}$.
13. Let $\mathcal{C}$ and $\mathcal{D}$ be categories. Prove that the product category $\mathcal{C} \times \mathcal{D}$ is indeed a category.
14. Let $F: \mathcal{B} \Rightarrow \mathcal{D}$ and $G: \mathcal{C} \Rightarrow \mathcal{D}$ be functors with the same codomain.
    a) Let $R$ be a commutative ring with unit. Show that the category $(R \to \mathbf{CRng})$ is the category of $R$-algebras.
    b) Let $t$ be a terminal element of a category $\mathcal{C}$. Describe $(\mathcal{C} \to t)$.
15. Show by example that the following do *not* hold in general.
    a) monic $\Rightarrow$ injective
       *Hint*: Let $\mathcal{C}$ be the category whose objects are the subsets of the integers $\mathbb{Z}$ and for which $\hom_{\mathcal{C}}(A, B)$ is the set of all *nonnegative* set functions from $A$ to $B$, along with the identity function when $A = B$. Consider the absolute value function $\alpha: \mathbb{Z} \to \mathbb{N}$.
    b) injective $\Rightarrow$ left-invertible
       *Hint*: Consider the inclusion map $\kappa: \mathbb{Z} \to \mathbb{Q}$ between rings.
    c) epic $\Rightarrow$ surjective
       *Hint*: Consider the inclusion map $\kappa: \mathbb{N} \to \mathbb{Z}$ between monoids.
    d) surjective $\Rightarrow$ right-invertible
       *Hint*: Let $C = \langle a \rangle$ be a cyclic group and let $H = \langle a^2 \rangle$. Consider the canonical projection map $\pi: C \to C/H = \{H, aH\}$.
16. Prove the following:
    a) For morphisms between sets, monoids, groups, rings or modules, any monic is injective. *Hint*: Let $f: A \to X$ be monic. Extend the relevant algebraic structure on $A$ coordinatewise to the cartesian product $A \times A$ and let

$$S = \{(a, b) \in A \times A \mid f(a) = f(b)\}$$

       Let $\rho_1: S \to A$ be projection onto the first coordinate and let $\rho_2: S \to A$ be projection onto the second coordinate. Apply $f \circ \rho_i$ to $(a, b) \in S$.
    b) For morphisms between sets, groups or modules, epic implies surjective. *Hint*: suppose that $f: A \to X$ is not surjective and let $I = \mathrm{im}(f)$. Find two distinct morphisms $p, q: X \to Y$ that agree on $I$, then $p \circ f = q \circ f$ but $p \neq q$, in contradiction to epicness. (For groups, this is a bit hard.)
    c) However, for morphisms between monoids or rings, epic does not imply surjective. *Hint*: Consider the inclusion map $\kappa: \mathbb{N} \to \mathbb{Z}$ between monoids and the inclusion map $\kappa: \mathbb{Z} \to \mathbb{Q}$ between rings.
17. (For those familiar with the tensor product) We want to characterize the epimorphisms in **CRng**, the category of commutative rings with identity. Let $A, B \in$ **CRng** and $f: A \to B$. Then $B$ is an $A$-module with scalar multiplication defined by

$$ab = f(a)b$$

    for $a \in A$ and $b \in B$. Consider the tensor product $B \otimes B$ of the $A$-module $B$ with itself. Show that $f$ is an epic if and only if $1 \otimes b = b \otimes 1$ for all $b \in B$. *Hint*: any ring map $\lambda: A \to R$ defines an $A$-module structure on $R$.
18. Let $\mathcal{C}$ be a category with a zero object. Show that the following are equivalent:
    1) $C$ is an initial object.
    2) $C$ is a terminal object.
    3) $\iota_C = 0_{CC}$
    4) $\hom_{\mathcal{C}}(C, C) = \{0_{CC}\}$

### *Image Factorization Systems*

An **image factorization system** for a category $\mathcal{C}$ is a pair $(\mathcal{E}, \mathcal{M})$ where
a) $\mathcal{E}$ is a nonempty class of epics of $\mathcal{C}$, closed under composition.
b) $\mathcal{M}$ is a nonempty class of monics of $\mathcal{C}$, closed under composition.
c) Any isomorphism of $\mathcal{C}$ belongs to $\mathcal{E}$ and $\mathcal{M}$.

d) Every morphism $f\colon A \to B$ can be factored as $f = m \circ e$ where $m \in \mathcal{M}$ and $e \in \mathcal{E}$. Moreover, this factorization is unique in the following sense: If $f = m' \circ e'$ with $m' \in \mathcal{M}$ and $e' \in E$, then there is an isomorphism $\theta\colon I \to J$ for which the following diagram commutes:

$$\begin{tikzcd} & I \arrow[dr, hook, "m"] \arrow[dd, "\exists\theta"'] & \\ A \arrow[ur, two heads, "e"] \arrow[dr, two heads, "e'"'] & & B \\ & J \arrow[ur, hook, "m'"'] & \end{tikzcd}$$

*Figure 15*

that is, $\theta \circ e = e'$ and $m' \circ \theta = m$.

19. Find an image factorization system for **Set**.
20. Find an image factorization system for **Grp**.
21. Prove the *diagonal fill-in theorem*: Let $(\mathcal{E}, \mathcal{M})$ be an image factorization system. Let $f\colon A \to C$ and $g\colon B \to D$ be morphisms in $\mathcal{C}$ and let $e \in \mathcal{E}$ and $m \in \mathcal{M}$, with the square in Figure 16 commutes.

$$\begin{tikzcd} A \arrow[r, two heads, "e"] \arrow[d, "f"'] & B \arrow[d, "g"] \arrow[dl, dashed, "\exists h"'] \\ C \arrow[r, hook, "m"'] & D \end{tikzcd}$$

*Figure 16*

Then there exists a morphism $h\colon B \to C$ for which the diagram in Figure 16 commutes.

$$\begin{tikzcd} A \arrow[rrr, "e"] \arrow[ddd, "f"'] \arrow[dr, "e_1"] & & & B \arrow[ddd, "g"] \arrow[dl, "e_2"'] \\ & I \arrow[ddl, "m_1"] & J \arrow[l, dashed, "\exists\theta"'] \arrow[ddr, "m_2"'] & \\ & & & \\ C \arrow[rrr, "m"'] & & & D \end{tikzcd}$$