

Chapter 6. Categories and functors.

6.1. What is a category? Let us lead up to the concept of category by first recalling the motivations for some more familiar mathematical concepts:

(a) *Groups.* The definition of a group is motivated by considering the structure on the set $\text{Aut}(X)$ of all automorphisms of a mathematical object X . Given $a, b \in \text{Aut}(X)$, the *composite* map ab lies in $\text{Aut}(X)$; for every $a \in \text{Aut}(X)$, its *inverse* a^{-1} is a member of $\text{Aut}(X)$, and, of course, the *identity map* id_X always belongs to $\text{Aut}(X)$. Thus, $\text{Aut}(X)$ is a set with a binary operation of composition, a unary operation “ $^{-1}$ ”, and a zeroary operation id_X . When one examines the conditions these operations satisfy, one discovers the associative law, the inverse laws, and the neutral-element laws.

These laws and their consequences turn out to be fundamental to a wide class of considerations involving automorphisms, so one makes a general definition: A 4-tuple $G = (|G|, \cdot, ^{-1}, 1)$, where $|G|$ is a set and $\cdot, ^{-1}, 1$ are operations on $|G|$ satisfying the above laws, is called a *group*.

Let me point out something which is obvious today, but took getting used to for the first generation to see the above definition: The definition does not say that G actually consists of automorphisms of an object X – only that it has certain properties we have abstracted from that context. In fact, systems with these properties are also found to arise in other ways:

The additive structures of the sets of integers, rational numbers, and real numbers form groups.

If (X, x_0) is a topological space with basepoint, the set of homotopy classes of closed curves beginning and ending at x_0 forms a group, $\pi_1(X, x_0)$.

And there are groups that are familiar, not because of a particular way they occur, but because of their importance as basic components in the study of groups in general. The finite cyclic groups \mathbb{Z}_n are the simplest examples.

Despite our abstract definition, and the existence of groups arising in these different ways, the original motivation of the group concept should not be forgotten. A natural question is: *Which* abstract groups can be represented *concretely*, that is, are isomorphic to a family of permutations of a set X under the operations of composition, inverse map, and identity permutation? As we learn in undergraduate algebra, the answer is that *every* group has this property (Cayley’s Theorem). Let us rederive the well-known proof.

The idea is to use the simplest nontrivial construction of a G -set X : Introduce a single generating element $x \in X$, and let all the elements gx ($g \in |G|$) be distinct. Formally we may define X to be the set of symbols “ gx ”, where x is a fixed symbol and g ranges over $|G|$. We let G act on X in the appropriate way to make this a G -action, namely by the law

$$h(gx) = (hg)x \quad (g, h \in |G|).$$

The permutations of the set X given by the elements of G are seen to form a “concrete” group isomorphic to G . One then observes that the symbol “ x ” is irrelevant to the proof. Stripping it away, we get the textbook proof: “Let G act on $|G|$ by left multiplication . . .” ([23, p. 62], [26, p.121], [27, p. 9], [29, p. 90], [32, p. 52]).

(b) *Monoids.* Suppose we consider not just the *automorphisms* of a mathematical object X but all its *endomorphisms*, that is, homomorphisms into itself. The set $\text{End}(X)$ is closed under

composition and contains the identity map, but there is no inverse operation. The operations of composition and identity still satisfy associative and neutral-element laws, and one calls any set with a binary operation and a distinguished element 1 satisfying these laws a *monoid*. Like the definition of a group, this definition does *not* require that a monoid actually consist of endomorphisms of an object X .

And indeed, there are again examples which arise in other ways than the one which motivated the definition. The nonnegative integers form a monoid under *multiplication* (with 1 as neutral element), and also under the operation \max (with 0 as neutral element). Isomorphism classes of (say) finitely generated abelian groups form a monoid under the operation induced by “ \oplus ”, or alternatively under the operation induced by “ \otimes ”. (One may remove some set-theoretic difficulties from this example by restricting oneself to a set of finitely generated abelian groups with exactly one member from each isomorphism class.)

One has the precise analog of Cayley’s Theorem: Every monoid S is isomorphic to a monoid of maps of a set into itself, and this is proved the same way, by letting S act on $|S|$ by left multiplication.

(c) *Partially ordered sets.* Again let X be any mathematical object, and now let us consider the set $\text{Sub}(X)$ of all *subobjects* of X .

In general, we do not have a way of defining interesting *operations* on this set. (There are often operations of “least upper bound” and “greatest lower bound”, but not always.) However, $\text{Sub}(X)$ is not structureless; one subobject of X may be *contained in* another, and this inclusion relation is seen to satisfy the conditions of *reflexivity*, *antisymmetry* and *transitivity*.

Again we abstract the situation, calling an arbitrary pair $P = (|P|, \leq)$, where $|P|$ is a set, and \leq is a binary relation on $|P|$ satisfying the above three laws, a *partially ordered set*.

Examples of partial orderings arising in other ways than the above “prototypical” one are the relation “ \leq ” on the integers or the real numbers, and the logical relation “ \Rightarrow ” on a family of inequivalent propositions. Partially ordered sets are also natural models of various hierarchical and genealogical structures in nature, language, and human society.

Given an arbitrary partially ordered set P , will P be isomorphic to a “concrete” partially ordered set – a family of subsets of a set X , ordered by inclusion? Again, let us try to build such an X in as simple-minded a way as possible. We want to associate to every $p \in |P|$ a subset \bar{p} of a set X , so as to duplicate the order relation among elements of P . To make sure all these sets are distinct, let us introduce for each $p \in |P|$ an element $x_p \in X$ belonging to \bar{p} , and hence necessarily to every \bar{q} with $q \geq p$, but not to any of the other sets \bar{q} ($q \not\geq p$). It turns out that this works – if we define X to be the set of symbols $\{x_p \mid p \in |P|\}$, and if for $p \in |P|$ we set $\bar{p} = \{x_q \mid q \leq p\} \subseteq X$, we find that $\{\bar{p} \mid p \in |P|\}$, under the relation “ \subseteq ”, forms a partially ordered set isomorphic to P . Again, the symbol “ x ” is really irrelevant, so we can get a simplified construction by taking $X = |P|$ and $\bar{p} = \{q \mid q \leq p\}$ ($p \in |P|$). Thus we have “Cayley’s Theorem for partially ordered sets”.

(d) “*Bimonoids.*” Let us go back to the idea that led to the definition of a monoid, but make a small change. Suppose that X and Y are two mathematical objects of the same sort (two sets, two rings, etc.), and we consider the family of all homomorphisms among them. What structure does this system have?

First, it is a system of four sets:

$$\text{Hom}(X, X), \quad \text{Hom}(X, Y), \quad \text{Hom}(Y, X), \quad \text{Hom}(Y, Y).$$

Elements of certain of these sets can be composed with elements of others, giving us *eight* composition maps:

$$\begin{aligned} \mu_{XXX} &: \text{Hom}(X, X) \times \text{Hom}(X, X) \rightarrow \text{Hom}(X, X), \\ \mu_{XXY} &: \text{Hom}(X, Y) \times \text{Hom}(X, X) \rightarrow \text{Hom}(X, Y), \\ &\quad \cdot \quad \cdot \quad \cdot \\ \mu_{YYX} &: \text{Hom}(Y, X) \times \text{Hom}(Y, Y) \rightarrow \text{Hom}(Y, X), \\ \mu_{YYY} &: \text{Hom}(Y, Y) \times \text{Hom}(Y, Y) \rightarrow \text{Hom}(Y, Y). \end{aligned}$$

(There is no composition on the remaining eight pairs, e.g., $\text{Hom}(X, Y) \times \text{Hom}(X, Y)$.)

These composition operations are associative – we have *sixteen* associative laws; namely, for every 4-tuple (Z_0, Z_1, Z_2, Z_3) of objects from $\{X, Y\}$ (e.g., (Y, Y, X, Y)) we get the law

$$(6.1.1) \quad (ab)c = a(bc)$$

for maps:

$$Z_0 \xrightarrow{c} Z_1 \xrightarrow{b} Z_2 \xrightarrow{a} Z_3.$$

(We could write (6.1.1) more precisely by specifying the four μ 's involved.) We also have two neutral elements, $\text{id}_X \in \text{Hom}(X, X)$ and $\text{id}_Y \in \text{Hom}(Y, Y)$, satisfying eight neutral element laws, which you can write down.

Cumbersome though this description is, it is clear that we have here a fairly natural mathematical structure, and we might abstract these conditions by defining a *bimonoid* to be any system of sets and operations

$$S = ((|S|_{ij})_{i,j \in \{0,1\}}, (\mu_{ijk})_{i,j,k \in \{0,1\}}, (1_i)_{i \in \{0,1\}})$$

such that the $|S|_{ij}$ are sets, the μ_{ijk} are maps

$$\mu_{ijk}: |S|_{jk} \times |S|_{ij} \rightarrow |S|_{ik},$$

satisfying associative laws $(ab)c = a(bc)$ on 3-tuples $(a, b, c) \in |S|_{jk} \times |S|_{ij} \times |S|_{hi}$ for all $h, i, j, k \in \{0, 1\}$, and such that the 1_i are elements of $|S|_{ii}$ ($i \in \{0, 1\}$) satisfying

$$1_j a = a = a 1_i \quad (a \in |S|_{ij}).$$

Again, these objects can arise in ways other than the one just indicated:

We can get an analog of the “ π_1 ” construction for groups: If X is a topological space and x_0, x_1 are two points of X , then the set of homotopy classes of paths in X whose initial and final points both lie in $\{x_0, x_1\}$ is easily seen to form a “bimonoid” which we might call $\pi_1(X; x_0, x_1)$.

Readers familiar with the ring-theoretic concept of a *Morita context* $(R, S; {}_R P_S, {}_S Q_R; \tau, \tau')$ will see that it also has this form: The underlying sets of the rings R and S play the roles of $|S|_{00}$ and $|S|_{11}$, the underlying sets of the bimodules P and Q give $|S|_{10}$ and $|S|_{01}$, and the required eight multiplication maps are given by the internal multiplication maps of R and S , the bimodule structures of P and Q , and the bilinear maps $\tau: P \times Q \rightarrow R$, and $\tau': Q \times P \rightarrow S$.

Finally, if K is a field and for any two integers i and j we write $M_{ij}(K)$ for the set of $i \times j$ matrices over K , then for any m and n , the four systems of matrices $M_{mm}(K), M_{nm}(K), M_{mn}(K), M_{nn}(K)$, form a “bimonoid” under matrix multiplication. (The astute reader will notice that this is really a disguised case of “two mathematical objects and maps among them”,

since matrix multiplication is designed precisely to encode composition of linear maps between vector spaces K^m and K^n . And the ring-theorist will note that this matrix example is also a Morita context.)

Is there a ‘‘Cayley’s Theorem for bimonoids’’, saying that any bimonoid S is isomorphic to a subbimonoid of the bimonoid of all maps between two sets X and Y ? Following the models of the preceding cases, our approach should be to introduce a small number of elements in X and/or Y , and use them to ‘‘generate’’ the rest of X and Y under the action of elements of S . Will it suffice to introduce a single generator $x \in X$, and let X and Y consist of elements obtained from x by application of the elements of the $|S|_{0j}$? In particular, this would mean taking for Y the set $\{tx \mid t \in |S|_{01}\}$. For some bimonoids S this will work; but in general it will not. For example, one can define a bimonoid S by taking any two monoids for $|S|_{00}$ and $|S|_{11}$, and the empty set for both $|S|_{01}$ and $|S|_{10}$. For such an S , the above construction gives empty Y , though if the monoid $|S|_{11}$ is nontrivial it cannot be represented faithfully by an action on the empty set. In the same way, it will not suffice to take *only* a generator in Y .

Let us, therefore, introduce as generators one element $x \in X$ and one element $y \in Y$, and let X be the set of all symbols of either of the forms sx or ty with $s \in |S|_{00}$, $t \in |S|_{10}$, and Y the set of symbols ux or vy with $u \in |S|_{01}$, $v \in |S|_{11}$. If we let S ‘‘act on’’ this pair of sets by defining

$$a(bz) = (ab)z,$$

whenever $z \in \{x, y\}$, and a and b are members of sets $|S|_{ij}$ such that these symbolic combinations should be meaningful, then we find that this yields an embedding of S in the bimonoid of all maps between X and Y , as desired. The interested reader can work out the details.

(e) *Categories*. We could go on in the same vein, looking at maps among 3, 4, etc., mathematical objects, and define ‘‘trimonoids’’, ‘‘quadrimonoids’’ etc., with larger and larger collections of operations and identities.

But clearly it makes more sense to treat these as cases of one general concept! Let us now, therefore, try to abstract the algebraic structure we find when we look at an arbitrary *family* \mathbf{X} of mathematical objects and the homomorphisms among them.

In the above development of ‘‘bimonoids’’, the index set $\{0, 1\}$ that ran through our considerations was the same for all bimonoids. But in the general situation, the corresponding index set must be specified as part of the object. This is the first component of the 4-tuple described in the next definition.

Definition 6.1.2 (provisional). *A category will mean a 4-tuple*

$$\mathbf{C} = (\text{Ob}(\mathbf{C}), \text{Ar}(\mathbf{C}), \mu(\mathbf{C}), \text{id}(\mathbf{C})),$$

where $\text{Ob}(\mathbf{C})$ is any collection of elements, $\text{Ar}(\mathbf{C})$ is a family of sets $\mathbf{C}(X, Y)$ indexed by the pairs of elements of $\text{Ob}(\mathbf{C})$:

$$\text{Ar}(\mathbf{C}) = (\mathbf{C}(X, Y))_{X, Y \in \text{Ob}(\mathbf{C})},$$

$\mu(\mathbf{C})$ is a family of operations

$$\mu(\mathbf{C}) = (\mu_{XYZ})_{X, Y, Z \in \text{Ob}(\mathbf{C})}$$

$$\mu_{XYZ}: \mathbf{C}(Y, Z) \times \mathbf{C}(X, Y) \rightarrow \mathbf{C}(X, Z),$$

and $\text{id}(\mathbf{C})$ is a family of elements

$$\begin{aligned} \text{id}(\mathbf{C}) &= (\text{id}_X)_{X \in \text{Ob}(\mathbf{C})} \\ \text{id}_X &\in \mathbf{C}(X, X), \end{aligned}$$

such that, using multiplicative notation for the maps μ_{XYZ} , the associative identity

$$a(bc) = (ab)c$$

is satisfied for all elements $a \in \mathbf{C}(Y, Z)$, $b \in \mathbf{C}(X, Y)$, $c \in \mathbf{C}(W, X)$ ($W, X, Y, Z \in \text{Ob}(\mathbf{C})$), and the identity laws

$$a \text{id}_X = a = \text{id}_Y a$$

are satisfied for all $a \in \mathbf{C}(X, Y)$ ($X, Y \in \text{Ob}(\mathbf{C})$).

The above definition is labeled “provisional” because it avoids the question of what we mean by a “collection of elements $\text{Ob}(\mathbf{C})$ ”. If we hope to be able to deal with categories within set theory, we should require $\text{Ob}(\mathbf{C})$ to be a *set*. Yet we will find that the most useful applications of category theory are to cases where $\text{Ob}(\mathbf{C})$ consists of *all* algebraic objects of a certain type (e.g., *all groups*), which calls for larger “collections”. We will deal with this dilemma in §6.4. In the next section, where we will give examples of categories, we will interpret “collection” broadly or narrowly as the example requires.

I mentioned that the concept of an “abstract group” – a group given as a set of elements with certain operations on them, rather than as a concrete family of permutations of a set – was confusing to people when it was first introduced. The “abstract” concept of a category still causes many people problems – there is a great temptation for beginning students to imagine that the members of $\mathbf{C}(X, Y)$ must be actual *maps* between *sets* X and Y .

One reason for this confusion is that the terminology of category theory is set up to closely mimic that of the situation which motivated the concept. The word “category” is suggestive to begin with; “ $\text{Ob}(\mathbf{C})$ ” stands for “objects of \mathbf{C} ”, and this is what elements of $\text{Ob}(\mathbf{C})$ are called; elements $f \in \mathbf{C}(X, Y)$ are called “morphisms” from X to Y , the objects X and Y are called the “domain” and “codomain” of f , these morphisms are often denoted diagrammatically by arrows, $X \xrightarrow{f} Y$, and objects and morphisms are shown together in the sort of diagrams that are used to represent objects and maps in other areas of mathematics. In place of $\mathbf{C}(X, Y)$, the notation $\text{Hom}(X, Y)$ is very common. And $\mu_{XYZ}(f, g)$ is generally written fg or $f \cdot g$ or $f \circ g$, and so looks just like a composite of functions.

So I urge you to note carefully the distinction between the situation that motivated our definition, and the definition itself. Within that definition, the collection $\text{Ob}(\mathbf{C})$ is simply an “index set” for the families of elements on which the composition operation is defined. Hence in discussing an abstract category \mathbf{C} , one cannot give arguments based on considering “an *element* of the object X ”, “the *image* of the morphism a ”, etc.; any more than in considering an abstract group G one can refer to such concepts as “the set of points left fixed by G ”. (However, the latter concept is meaningful for concrete groups of permutations, and the former concepts are likewise meaningful for “concrete categories”, a concept we will define precisely in §6.5.)

Of course, the motivating situation should not be forgotten, and a natural question is: Is every category isomorphic to a system of maps among some sets? We can give a qualified affirmative

Summary of §6.1

(Read across rows, referring to headings at top, then compare downwards)

Consider:	Structure:	Properties:	Abstract definition:	Other examples:	Can be represented by:
All automorphisms of a mathematical object X .	Set with composition, inverse operation, and identity element.	$(ab)c = a(bc)$, $a^{-1}a = \text{id} = a a^{-1}$, $a \text{id} = a = \text{id} a$.	<i>group</i> (same properties, but not assumed to arise as at left)	$(\mathbb{Z}, +, -, 0)$, $\pi_1(X, x_0)$, \mathbb{Z}_n , etc.	permutations of a set (Cayley's Theorem).
All endomorphisms of a mathematical object X .	Set with composition and identity element.	$(ab)c = a(bc)$, $a \text{id} = a = \text{id} a$.	<i>monoid</i> (same properties, but not assumed to arise as at left)	$(\mathbb{N}, \cdot, 1)$, $(\mathbb{N}, \max, 0)$, {f.g. ab. gps.}, \otimes, \mathbb{Z})	maps of a set into itself.
All sub-objects of a mathematical object X .	Set with relation \subseteq .	transitive, antisymmetric, reflexive.	<i>partially ordered set</i> (same properties, but not assumed to arise as at left)	(\mathbb{Z}, \leq) , \Rightarrow , genealogies, etc.	subsets of a set, under \subseteq .
All homomorphisms between two mathematical objects X and Y .	Four sets, $ S _{00}, S _{01}, S _{10}, S _{11}$, with composition maps $ S _{jk} \times S _{ij} \rightarrow S _{ik}$ and identity elements id_0, id_1 .	$(ab)c = a(bc)$ (when defined); $a \text{id}_i = a = \text{id}_j a$ ($a \in S _{ij}$).	"bimonoid" (same properties, but not assumed to arise as at left)	" $\pi_1(X; x_0, x_1)$ ", Morita contexts, matrices.	maps between two sets.
All homomorphisms among a family \mathbf{X} of mathematical objects.	Family of sets $\text{Hom}(X, Y)$ ($X, Y \in \mathbf{X}$) with composition maps $\text{Hom}(X_1, X_2) \times \text{Hom}(X_0, X_1) \rightarrow \text{Hom}(X_0, X_2)$ and identity elements $\text{id}_X \in \text{Hom}(X, X)$.	$(ab)c = a(bc)$ for $X_0 \xrightarrow{c} X_1 \xrightarrow{b} X_2 \xrightarrow{a} X_3$; $a \text{id}_X = a = \text{id}_Y a$ for $X \xrightarrow{a} Y$.	<i>category</i> (same properties, but not assumed to arise as at left)	coming up, in §6.2.	family of sets and maps among them (if $\text{Ob}(\mathbf{C})$ is a set).

answer. The complete answer depends on the set-theoretic matters that we have postponed to §6.4, but if $\text{Ob}(\mathbf{C})$ is actually a *set*, then we can indeed construct sets $(\bar{X})_{X \in \text{Ob}(\mathbf{C})}$, and set maps among these, including the identity map of each of these sets, which form under composition of maps a category isomorphic to \mathbf{C} . The proof is the analog of the one we sketched for “bimonoids”.

Exercise 6.1:1. Write out the argument indicated above – “Cayley’s Theorem” for a category with only a *set* of objects.

Incidentally, we will now discard the term “bimonoid”, since the structure it described was, up to notational adjustment, simply a category having for object-set the two-element set $\{0, 1\}$.

6.2. Examples of categories. To describe a category, one should, strictly, specify the class of *objects*, the *morphism-set* associated with any pair of objects, the *composition* operation on morphisms, and the *identity morphism* of each object. In practice, some of this structure is usually clear from context. When one is dealing with the prototype situation – a family of mathematical objects and all homomorphisms among them – the whole structure is usually clear once the class of objects is named. In other cases the morphism-sets must be specified as well; once this is done the intended composition operation is usually (though not always) obvious. As to the identity elements, these are uniquely determined by the remaining structure (just as in groups or monoids), so the only task is to verify that they exist, which is usually easy.

Categories consisting of families of mathematical objects and the homomorphisms among them are generally denoted by boldface or script names for the type of object (often abbreviated. The particular abbreviations may vary from author to author.) Some important examples are:

Set, the category of all sets and set maps among them. (Another symbol commonly used for this category is **Ens**, from the French word *ensemble*.)

Group, the category whose objects are all groups, and whose morphisms are the group homomorphisms; and similarly **Ab**, the category of abelian groups.

Monoid, **Semigroup**, **AbMonoid** and **AbSemigroup**, the categories of monoids, semigroups, abelian monoids and abelian semigroups.

Ring¹, and **CommRing**¹ the categories of associative, respectively associative commutative, rings with unity. (One can then denote by the same symbols without a superscript “1,” the corresponding categories of nonunital rings – i.e., 5-tuples $R = (|R|, +, \cdot, -, 0)$, where $|R|$ need not contain an element 1 satisfying the neutral law for multiplication, and where, even if rings happen to possess such elements, morphisms are not required to respect them. But we will not refer to those categories often enough in these notes to need to fix names for them.)

If R is a unital associative ring, we will write the category of left R -modules **R -Mod** and the category of right R -modules **Mod- R** . (Other common notations for these are ${}_R\mathbf{Mod}$ and \mathbf{Mod}_R respectively.) Similarly, for G a group, the category of (left) G -sets will be written **G -Set**; here the morphisms are the set maps respecting the actions of all elements of G .

Top denotes the category of all topological spaces and *continuous maps* among them. Topologists often find it useful to work with topological spaces with basepoint, (X, x_0) , so we also define the category **Top**^{Pt} of *pointed* topological spaces, the objects of which are such pairs (X, x_0) , and the morphisms of which are the continuous maps which send basepoint to basepoint. Much of topology is done under the assumption that the space is Hausdorff; thus one considers the subcategory **HausTop** of **Top** whose objects are the Hausdorff spaces.

We shall write **POSet** for the category of partially ordered sets, with isotone maps for morphisms. If we want to allow only strict isotone maps, i.e., maps respecting the relation “<”, we can call the resulting category **POSet**_<.

We have mentioned that our concept of “bimonoid” was a special case of the concept of category. Let us make this precise. The definition of a category requires specification of the object-set, whereas for bimonoids the implicit object-set was always $\{0, 1\}$. So given a bimonoid $S = ((|S|_{ij}), (\mu_{ijk}), (1_i))$, to translate it to a category **C**, we throw in a formal first component $\text{Ob}(\mathbf{C}) = \{0, 1\}$. We can then define $\mathbf{C}(i, j) = |S|_{ij}$, getting the category $(\{0, 1\}, (|S|_{ij}), (\mu_{ijk}), (1_i))$, which we may denote S_{cat} .

This works because the situation from which we abstracted the concept of a bimonoid was a special case of the situation from which we abstracted the concept of a category. Now in fact, the situations from which we abstracted the concepts of *group*, *monoid*, and *partially ordered set* were also special cases of that situation! Can objects of these types similarly be identified with certain kinds of categories?

The objects most similar to bimonoids are the monoids. Since they are modeled after the algebraic structure on the set of endomorphisms of a single algebraic object, let us associate to an arbitrary monoid S a *one-object* category S_{cat} , with object-set $\{0\}$. The only morphism-set to define is $S_{\text{cat}}(0, 0)$, we take this to be $|S|$; for the composition map on pairs of elements of $S_{\text{cat}}(0, 0)$, we use the composition operation of S , and for the identity morphism, the neutral element of S .

Conversely, if **C** is any category with only one object, X , then the unique morphism set $\mathbf{C}(X, X)$ with its identity element will form a monoid S under the composition operation of **C**, such that the category S_{cat} formed as above is isomorphic to our original category **C**, the only difference being the name of the one object (originally X , now 0). Thus, a category with exactly one object is “essentially” a monoid.

If we start with a group G , we can similarly form a category G_{cat} with just one object, 0 , whose morphisms are the elements of G and whose composition operation is the composition of G . We cannot incorporate the inverse operation of G as an operation of the category; in fact, what we are doing is essentially forgetting the inverse operation, i.e., forming from G the monoid G_{md} , and then applying the previous construction; thus $G_{\text{cat}} = (G_{\text{md}})_{\text{cat}}$. We see that via this construction, a *group* is equivalent to a category which has exactly one object, and in which every morphism is invertible.

Note that for G a group, the one member of $\text{Ob}(G_{\text{cat}})$ should not be thought of as the group G ; intuitively it is a fictitious mathematical object on which G acts. Thus, morphisms in this category from that one object to itself do not correspond to endomorphisms of G , as students sometimes think, but to *elements* of G . (One can also define a category with one object whose morphisms *are* the endomorphisms of G ; that is the category $\text{End}(G)_{\text{cat}}$; but G_{cat} is a more elementary construction.)

The case of partially ordered sets is a little different. In the motivating situation, though we started with a single object X , we considered a family of objects obtained from it, namely all its subobjects. Although there might exist many maps among these objects, the structure of partially ordered set only reveals a certain subfamily of these: the inclusion maps. (In fact, since a “homomorphism” means a map which respects the kind of structure being considered, and we are considering these objects as subobjects of X , one could say that a homomorphism *as subobjects* should mean a set map which respects the way the objects are embedded in X , i.e., an inclusion

map; so from this point of view, these really are the only relevant maps.) A composite of inclusion maps is an inclusion map, and identity maps are (trivial) inclusions, so the subobjects of X with the inclusion maps among them form a category. In this category there is a morphism from A to B if and only if $A \subseteq B$, and the morphism is then unique, so the partial ordering of the subobjects determines the structure of the category.

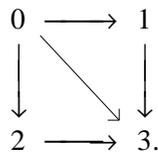
If we start with an abstract partially ordered set $P = (|P|, \leq)$, we can construct from it an abstract category $P_{\mathbf{cat}}$ in the way suggested by this concrete prototype: Take $\text{Ob}(P_{\mathbf{cat}}) = |P|$, and for all $A, B \in |P|$, define there to be one morphism from A to B if $A \leq B$ in P , none otherwise. What should we take this one morphism to be? This is like asking in our construction of $G_{\mathbf{cat}}$ what to call the one object. The choice doesn't really matter. Since we want to associate to each ordered pair (A, B) with $A \leq B$ in P some element, the easiest choice is to take for that element the pair (A, B) itself. Thus, we can define $P_{\mathbf{cat}}$ to have object-set $|P|$, and for $A, B \in |P|$, take $P_{\mathbf{cat}}(A, B)$ to be the singleton $\{(A, B)\}$ if $A \leq B$, the empty set otherwise. The reader can easily describe the composition operation and identity elements of $P_{\mathbf{cat}}$.

Incidentally, we see that this construction works equally well if \leq is a preordering rather than a partial ordering.

Exercise 6.2:1. Let \mathbf{C} be a category.

- (i) Show that \mathbf{C} is isomorphic to $P_{\mathbf{cat}}$ for some partially ordered set P if and only if “there is at most one morphism between any unordered pair of objects”; in the sense that each hom-set $\mathbf{C}(X, Y)$ has cardinality at most 1, and the hom-sets $\mathbf{C}(X, Y)$ and $\mathbf{C}(Y, X)$ do not both have cardinality 1 unless $X = Y$.
- (ii) State a similar condition necessary and sufficient for \mathbf{C} to be isomorphic to $P_{\mathbf{cat}}$ for P a preorder. (No proof required.)

We mentioned that some groups, such as the cyclic groups \mathbb{Z}_n , are of interest as “pieces” in terms of which we look at general groups. Thus, to give an element of order n in a group G is equivalent to displaying an isomorphic copy of \mathbb{Z}_n in G , and to give an element satisfying $x^n = e$ is equivalent to displaying a homomorphic image of \mathbb{Z}_n in G . Various simple categories are of interest for essentially the same reason. For instance a *commutative square* $\begin{matrix} \downarrow & \rightarrow & \downarrow \\ & & \\ \downarrow & \rightarrow & \downarrow \end{matrix}$ of objects and morphisms in a category \mathbf{C} corresponds to an image in \mathbf{C} of a certain category having four objects, which we can name 0, 1, 2 and 3, and, aside from their identity morphisms, five arrows, as shown below:



Here the diagonal arrow is *both* the composite of the morphisms from 0 to 1 to 3 and the composite of the morphisms from 0 to 2 to 3. This “diagram category” might be conveniently named “ $\begin{matrix} \downarrow & \rightarrow & \downarrow \\ & & \\ \downarrow & \rightarrow & \downarrow \end{matrix}$ ”.

A simpler example is the diagram category $\cdot \rightrightarrows \cdot$, with two objects and only two nonidentity morphisms, going in the same direction. Copies of this in a category \mathbf{C} correspond to the type of data one starts with in the definitions of *equalizers* and *coequalizers*. Still simpler is $\cdot \rightarrow \cdot$, which is often called “**2**”; an image of this in a category corresponds to a choice of two objects and one morphism between them. (So the category **2** takes its place in our vocabulary beside the ordinal

2, the Boolean ring $\mathbf{2}$, the lattice $\mathbf{2}$, and the partially ordered set $\mathbf{2}$!) A larger diagram category is

$$\cdot \rightarrow \cdot \rightarrow \cdot \rightarrow \cdot \rightarrow \dots$$

images of which in \mathbf{C} correspond to right-infinite chains of morphisms. The morphisms of this diagram category are the identity morphisms, the arrows shown in the picture, and all composites of these arrows, of which there is exactly one from every object to every object to the right of it. Finally, one might denote by \mathcal{C} a category having one object 0 , and, aside from the identity morphism of 0 , one other morphism x , and all its powers, x^2, x^3 , etc.. An image of this in a category \mathbf{C} will correspond to a choice of an object and a morphism from this object to itself.

(In the above discussion I have been vague about what I meant by an ‘‘image’’ of one category in another. In §6.5 we shall introduce the category-theoretic concept analogous to that of *homomorphism*, in terms of which this can be made precise. At this point, for the sake of giving you some broad classes of examples to think about, I have spoken without having the formal definition at hand.)

The various types of examples we have discussed are by no means disjoint. Three of the above ‘‘diagram categories’’ can be recognized as having the form $P_{\mathbf{cat}}$, where P is respectively, a 4-element partially ordered set, the partially ordered set $\mathbf{2}$, and the partially ordered set of nonnegative integers, while the last example is $S_{\mathbf{cat}}$, for S the free monoid on one generator x .

Many of the other ‘‘nonprototypical’’ ways in which we saw that groups, etc., arise also have generalizations to categories:

If R is any ring, we see that multiplication of rectangular matrices over R satisfies precisely the laws for composition of morphisms in a category. Thus, we get a category \mathbf{Mat}_R by defining the objects to be the nonnegative integers, the morphism-set $\mathbf{Mat}_R(m, n)$ to be the set of all $n \times m$ matrices over R , the composition μ to be matrix multiplication, and the morphisms id_n to be the identity matrices I_n . This is not very novel, since as we observed before, matrix multiplication is defined to encode composition of linear maps among free R -modules. But it is interesting to note that the abstract system of matrices over R is not limited to serving that function; if M is any left R -module, one can use $n \times m$ matrices over R to represent operations which carry m -tuples of elements of M to n -tuples formed from these elements using linear expressions with coefficients in R .

This line of thought suggests similar constructions for other sorts of algebraic objects. For instance, we can define a category \mathbf{C} whose objects are again the nonnegative integers, and such that $\mathbf{C}(m, n)$ represents all ways of getting an n -tuple of elements of an arbitrary *group* from an m -tuple using *group operations*. Precisely, we can define $\mathbf{C}(m, n)$ to be the set of all n -tuples of *derived group-theoretic operations* in m variables. The composition maps

$$(6.2.1) \quad \mathbf{C}(n, p) \times \mathbf{C}(m, n) \rightarrow \mathbf{C}(m, p)$$

can be described in terms of substitution of derived operations into one another.

Generalizing the construction of the fundamental group of a topological space with basepoint (X, p) , one can associate to any topological space X a category $\pi_1(X)$ whose objects are all points of X , and where a morphism from x_0 to x_1 means a homotopy class of paths from one point to the other.

We can also define categories which have familiar mathematical entities for their objects, but put unexpected twists into the definitions of the morphism-sets. Recall that in the category \mathbf{Set} , the morphisms from the set X to the set Y are all functions from X to Y . Now formally, a

function is a relation $f \subseteq X \times Y$ such that for every $x \in X$ there exists a unique $y \in Y$ such that $(x, y) \in f$. Suppose we drop this restriction, and consider arbitrary relations $R \subseteq X \times Y$. One can compose these using the same formula by which one composes functions: If $R \subseteq X \times Y$ and $S \subseteq Y \times Z$, one defines

$$S \circ R = \{(x, z) \in X \times Z \mid (\exists y \in Y) (x, y) \in R, (y, z) \in S\}.$$

This operation of composing relations is associative, and the identity relations satisfy the identity laws; hence one can define a category **RelSet**, whose objects are ordinary sets, but such that **RelSet**(X, Y) is the set of relations in $X \times Y$.

Algebraic topologists work with topological spaces, but instead of individual maps among them, they are concerned with *homotopy classes* of maps. Thus, they use the category **HtpTop** whose objects are topological spaces, and whose morphisms are such homotopy classes. Composition of continuous maps respects homotopy, allowing one to define the composition operation of this category.

In complex variable theory, one often fixes a point z of the complex plane and considers all analytic functions defined in a neighborhood of z . Different functions in this set are defined on different neighborhoods of z , so these functions do not all have any domain of definition in common. Further, functions which are the same in a neighborhood of z may not agree on the full intersection of their domains, if this intersection is not connected. E.g., the natural logarithm function $\ln(z)$ with value zero at $z = 1$ extends to some connected regions of the plane so as to assume the value $+\pi i$ at the point -1 , and to other such regions so as to assume the value $-\pi i$ at that point. To eliminate distinctions which are not relevant to the behavior of functions in the vicinity of the specified point z , one introduces the concept of a *germ of a function* at z . This is an equivalence class of functions defined on neighborhoods of z , under the relation making two functions equivalent if they agree on some common neighborhood of z .

An apparent inconvenience of this concept is that for germs of functions at z , one does not have a well-defined operation of composition. For instance, if f and g are germs of analytic functions at $z = 0$, one cannot generally attach a meaning to gf unless $f(0) = 0$, because g does not have a well-defined “value” at $f(0)$. (This is the analog of the algebraic problem that given formal power series $f(z) = a_0 + a_1 z + \dots$ and $g(z) = b_0 + b_1 z + \dots$, one cannot in general “substitute f into g ” to get another formal power series in z , unless $a_0 = 0$.) But this ceases to be a problem if we define a category **GermAnal**, whose objects are the points of the complex plane, and where a morphism from z to w means a germ of an analytic function at z whose value at z is w . Then for any three points z_0, z_1, z_2 , one sees that one does indeed have a well-defined composition operation

$$\mathbf{GermAnal}(z_1, z_2) \times \mathbf{GermAnal}(z_0, z_1) \rightarrow \mathbf{GermAnal}(z_0, z_2).$$

I.e., the partial operation of composition of germs of analytic functions is defined in exactly those cases needed to make these germs the morphisms of a category.

These examples allow endless modification as needed. A topologist may impose the restriction that the topological spaces considered in a given context be Hausdorff, be locally compact, be given with basepoint, etc., and modify the category he or she uses accordingly. The definition of a *germ of a function* is not limited to complex variable theory, so analogs of **GermAnal** can be set up wherever needed. Here is an interesting case:

Exercise 6.2:2. If G and H are groups, let us define an *almost-homomorphism* from G to H to mean a homomorphism $f: G_f \rightarrow H$, whose domain G_f is a subgroup of *finite index* in G . Given two almost-homomorphisms f and g from G to H , with domains G_f and G_g , let us write $f \approx g$ if the subgroup $\{x \in |G_f| \cap |G_g| \mid f(x) = g(x)\}$ also has finite index in G .

- (i) Show that \approx is an equivalence relation on the set of almost-homomorphisms from G to H .
- (ii) Show how one may define a category \mathbf{C} whose objects are all groups, and whose morphisms are the equivalence classes of almost-homomorphisms, under \approx .
- (iii) Describe the endomorphism-monoid $\mathbf{C}(\mathbb{Z}, \mathbb{Z})$, where \mathbf{C} is the category described above, and \mathbb{Z} is the additive group of integers.

We noted earlier that isomorphism classes of abelian groups formed a monoid under \otimes . The reader with some ring-theoretic background might like the following generalization of this monoid to a category.

Exercise 6.2:3. Show that one can define a category \mathbf{C} such that $\text{Ob}(\mathbf{C})$ is the class of all rings, $\mathbf{C}(R, S)$ is, for each $R, S \in \text{Ob}(\mathbf{C})$, the family of all isomorphism-classes $[P]$ of (S, R) -bimodules P , and the composite $[P][Q]$ is the isomorphism class of the tensor product, $[P \otimes_S Q]$, for $[P] \in \mathbf{C}(S, T)$, $[Q] \in \mathbf{C}(R, S)$. (Either ignore the problem that the classes involved in this definition are not sets, or modify the statement in some reasonable way to avoid this problem.)

If you are familiar with Morita equivalence, verify that two objects are isomorphic in this category if and only if they are Morita equivalent as rings. (If you do this part, state the precise definition of Morita equivalence that you are using, since it has various formulations.)

The following example shows that not every plausible definition works:

Exercise 6.2:4. Suppose one attempts to define a category \mathbf{C} by taking all sets for the objects, and letting $\mathbf{C}(X, Y)$ consist of all equivalence classes of set maps $X \rightarrow Y$, under the relation that makes $f \approx g$ if $\{x \in X \mid f(x) \neq g(x)\}$ is finite.

- (i) Show that this does not work, i.e., that composition of set maps does not induce a composition operation on equivalence classes of set maps.

On the other hand

- (ii) Find the least equivalence relation \sim on set maps which contains the above equivalence relation \approx , and has the property that composition of set maps does induce a composition operation on equivalence classes of set maps under \sim .

(Precisely, \sim will be a family of equivalence relations: a relation $\sim_{X, Y}$ on $\mathbf{C}(X, Y)$ for each pair of sets X and Y . So what you should show is that among such families of equivalence relations, there is a least \sim such that composition of set maps induces composition operations on the factor sets $\mathbf{C}(X, Y)/\sim_{X, Y}$, and such that $f \approx g \Rightarrow f \sim g$ for all f and g ; and describe these relations $\sim_{X, Y}$.)

Here is an interesting variant on the construction S_{cat} , for S a monoid. (For an application, see [43].)

Exercise 6.2:5. Let S be a monoid, and X an S -set. One can define a category whose objects are the elements of X , and such that a morphism $x \rightarrow y$ ($x, y \in |X|$) is an element $s \in |S|$ such that $sx = y$. However, to help remind us of the intended domain and codomain of each morphism, let us, rather, take the morphisms $x \rightarrow y$ to be all 3-tuples (y, s, x) such that $s \in |S|$ and $sx = y$. We define composition by $(z, t, y)(y, s, x) = (z, ts, x)$; the definition of the identity morphisms should be clear.

- (i) Show that the construction S_{cat} is a special case of this construction.
- (ii) In general, can one reconstruct the monoid S and the S -set X from the structure of the

category $X_{\mathbf{cat}}$?

I don't know the answer to the first part of

(iii) Given a category \mathbf{C} , is there a nice necessary and sufficient condition for there to exist a monoid S and an S -set X such that $\mathbf{C} \cong X_{\mathbf{cat}}$? For there to exist a group G and a G -set X such that this isomorphism holds?

6.3. Other notations and viewpoints. The language and notation of category theory are still far from uniform. Let me note some of the commonest variations on the conventions I have presented.

I have mentioned that what we are writing $\mathbf{C}(X, Y)$ is often written $\text{Hom}(X, Y)$; this may be made more explicit as $\text{Hom}_{\mathbf{C}}(X, Y)$; there is also the shorter notation (X, Y) . Even though we shall not use the notation $\text{Hom}(X, Y)$, we shall often call these sets "hom-sets".

More problematically, the order in which the objects are written may be reversed; i.e., some authors write the set of morphisms from X to Y as $\mathbf{C}(Y, X)$, $\text{Hom}(Y, X)$, etc.. There are advantages to each choice: The order we are using matches the conceptual order of going "from X to Y ", and the use of arrows drawn from left to right, $X \xrightarrow{a} Y$, but has the disadvantage that composition of morphisms $X \rightarrow Y \rightarrow Z$ must be described as a map $\mathbf{C}(Y, Z) \times \mathbf{C}(X, Y) \rightarrow \mathbf{C}(X, Z)$, while under the reversed notation, composition goes more nicely, $\mathbf{C}(Z, Y) \times \mathbf{C}(Y, X) \rightarrow \mathbf{C}(Z, X)$. A different cure for the same problem is to continue to think of elements of $\mathbf{C}(X, Y)$ as morphisms from X to Y (as we are doing), but reverse the way composition is written, letting the composite of $X \xrightarrow{a} Y \xrightarrow{b} Z$ be denoted $ab \in \mathbf{C}(X, Z)$, rather than ba . However if one does this, then when writing functions on sets, one is more or less forced to abandon the conventional notation $f(x)$, which leads to the usual order of composition, and write xf instead.

Note that the above difficulties in category-theoretic notation simply mirror conflicts of notation already existing within mathematics!

The elements of $\mathbf{C}(X, Y)$, which we call "morphisms", are called "arrows" by some. Our notation $\text{Ar}(\mathbf{C})$ for the family of morphism-sets is based on that word; some authors write $\text{Fl}(\mathbf{C})$, based on the French *flèche* (arrow). Colloquially they are also called "maps" from X to Y , and I may allow myself to fall into this easy usage at times, hoping that you understand by now that they are *not* maps in the literal sense, i.e., functions.

The identity element in $\mathbf{C}(X, X)$ which we are writing id_X is also written I_X (like an identity matrix) or 1_X (just as the identity element of a group is often written 1).

The student has probably noticed at some point in his or her study of mathematics the petty but vexing question: If X is a subset of Y , is the inclusion map of X into Y the "same" as the identity map of X ? If we follow the convenient formalization of a function as a set of ordered pairs $(x, f(x))$, then they are indeed the same. But this means that a question like "Is f surjective?" is meaningless; one can only ask whether f is surjective as a map from X to Y , whether it is surjective as a map from X to Z , etc.. A formalization more in accord with the way we think about these things might be to define a function $f: X \rightarrow Y$ as a 3-tuple $(X, Y, |f|)$, where $|f|$ is the set of ordered pairs used in the usual definition. Then f is surjective if and only if the set of second components of members of $|f|$ equals the whole set Y . (Since X is determined by $|f|$, our making X a component of the 3-tuple is, strictly, unnecessary; but it seems worth doing for symmetry. Note that if one wants to use a similar notation for general relations $|R| \subseteq X \times Y$, then neither X nor Y will be determined by $|R|$, so one needs both of these in the tuple describing the relation. Having both in the tuple describing a function then allows one to consider the functions from X to Y as a subset of the relations between these sets.)

The same problem arises when we abstract our functions in the definition of a category: Can an element be a member of two different morphism-sets, $\mathbf{C}(X, Y)$ and $\mathbf{C}(X', Y')$, with $(X, Y) \neq (X', Y')$? Under our definitions, yes. However, some authors add to the definition of a category the condition that the sets of morphisms between distinct pairs of objects be disjoint.

Let us note what such a condition would entail. In the category **Group**, as an example, a group homomorphism $f: G \rightarrow H$ would have to determine not merely its set-theoretic domain and codomain $|G|$ and $|H|$, but the full group structures $G = (|G|, \mu_G, \iota_G, e_G)$ and $H = (|H|, \mu_H, \iota_H, e_H)$. When one thinks about it, this makes good sense, not only from the point of view of category theory but from that of group theory; for without knowing the group structures on $|G|$ and $|H|$, one cannot say whether a map $f: |G| \rightarrow |H|$ is a homomorphism, let alone answer such group-theoretic questions as, say, whether its kernel contains all elements of order 2.

Observe that in set theory, even if one does not define a function so as to determine its codomain, certain things remain well-defined; for example, the composite fg of two composable maps can be defined knowing only the set of ordered pairs comprising each of these maps. But there is nothing in the axioms of a category that says that if g lies in both $\mathbf{C}(X, Y)$ and $\mathbf{C}(X', Y')$, while f lies in both $\mathbf{C}(Y, Z)$ and $\mathbf{C}(Y', Z')$, then the composites $\mu_{XYZ}(f, g)$ and $\mu_{X'Y'Z'}(f, g)$ will be the same; so even the symbol “ fg ” is formally ambiguous.

On the whole, I think it desirable to include in the definition of a category the condition that morphism-sets be disjoint. However, we shall not do so in these notes, largely because it would increase the gap between our category theory and ordinary mathematical usage. So the difficulties mentioned above mean that we have to be careful, understanding for instance that in a given context, we are using fg as a shorthand for $\mu_{XYZ}(f, g)$, which is the only really unambiguous expression. Note, however, that given any structure which is a category \mathbf{C} under our definition, we can form a new category \mathbf{C}^{disj} with disjoint morphism-sets, by using the same objects, and letting $\mathbf{C}^{\text{disj}}(X, Y)$ consist of all 3-tuples $f = (X, Y, |f|)$ with $|f| \in \mathbf{C}(X, Y)$, and composition operations obtained in the obvious way from those of \mathbf{C} .

Authors who require morphism-sets to be disjoint can play some interesting variations on the definition of category. Instead of defining $\text{Ar}(\mathbf{C})$ to be a family of sets, $\text{Ar}(\mathbf{C}) = (\mathbf{C}(X, Y))_{X, Y \in \text{Ob}(\mathbf{C})}$, they can take it to be a single set (or class), the union of all the $\mathbf{C}(X, Y)$'s. To recover *domains* and *codomains* of morphisms, they then add to the definition of a category two operations, $\text{dom}, \text{cod}: \text{Ar}(\mathbf{C}) \rightarrow \text{Ob}(\mathbf{C})$. They can then make composition of morphisms a single map

$$\mu: \{(f, g) \in \text{Ar}(\mathbf{C})^2 \mid \text{dom}(f) = \text{cod}(g)\} \rightarrow \text{Ar}(\mathbf{C}).$$

One can be even more radical and eliminate all reference to objects, as sketched in the next exercise:

Exercise 6.3:1. (i) Let \mathbf{C} be a category such that distinct ordered pairs of objects (X, Y) have disjoint morphism-sets. Let $A = \bigcup_{X, Y} \mathbf{C}(X, Y)$. Let μ denote the composition operation in A , considered now as a *partial map* from $A \times A$ to A , i.e., a function from a subset of $A \times A$ to A . Show that the pair (A, μ) determines \mathbf{C} up to isomorphism.

(ii) Find conditions on a pair (A, μ) , where A is a set and μ a partial binary operation on A , which are necessary and sufficient for it to arise, as above, from a category \mathbf{C} .

One gets a nicer structure by combining the above approach with that of giving functions specifying the domain and codomain of each morphism. Namely, given a category \mathbf{C} with disjoint morphism-sets, let A be defined as in (i), let $\text{dom}: A \rightarrow A$ be the map associating to each morphism f the *identity* morphism of its domain, and similarly let $\text{cod}: A \rightarrow A$ associate to each morphism the identity morphism of its codomain. Since the pair (A, μ) determines \mathbf{C}

up to isomorphism, the same will be true, a fortiori, of the 4-tuple $(A, \mu, \text{dom}, \text{cod})$.

(iii) Find necessary and sufficient conditions on a 4-tuple $(A, \mu, \text{dom}, \text{cod})$ for it to arise as above from a category \mathbf{C} .

So one could redefine a category as an ordered pair (A, μ) or 4-tuple $(A, \mu, \text{dom}, \text{cod})$ satisfying appropriate conditions.

However, these differences in definition do not make a great difference in how one actually works with categories. If, for instance, one defines a category as a 5-tuple $\mathbf{C} = (\text{Ob}(\mathbf{C}), \text{Ar}(\mathbf{C}), \text{dom}_{\mathbf{C}}, \text{cod}_{\mathbf{C}}, \text{id}_{\mathbf{C}})$, one then immediately makes the definition

$$\mathbf{C}(X, Y) = \{ f \in \text{Ar}(\mathbf{C}) \mid (\text{dom}(f) = X) \wedge (\text{cod}(f) = Y) \},$$

and works with these morphism sets as other category-theorists do. (But I will mention one notational consequence of the morphisms-only approach that can be confusing to the uninitiated: the use, by some categorists, of the name of an object as the name for its identity morphism as well.)

Changing the topic from technical details to attitudes, category theory has been seen by some as the new approach that would revolutionize, unify, and absorb all of mathematics; by others as a pointless abstraction whose content is trivial where it is not incomprehensible.

Neither of these characterizations is justified, but each has a grain of truth. The subject matter of essentially every branch of mathematics can be viewed as forming a category (or a family of categories); but this does not say how much value the category-theoretic viewpoint will have for workers in a given area. The actual role of category theory in mathematics is like that of group theory: Groups come up in all fields of mathematics, because for every sort of mathematical object we can look at its symmetries, and generally make use of them. In some situations the contribution of group theory is limited to a few trivial observations, and to providing a language consistent with that used for similar considerations in other fields. In others, deep group-theoretic results are applicable. Finally, group theory is a branch of algebra in its own right, with its own intrinsically interesting questions. All the corresponding observations are true of category theory.

As with the concept of “abstract group” for an earlier generation, many people are troubled by that of an “abstract category”, whose “objects” are structureless primitives, not mathematical objects with “underlying sets”, so that in particular, one cannot reason by “chasing elements” around diagrams. I think the difficulty is pedagogic. The problem comes from *expecting* to be able to “chase elements”. As one learns category theory (or a given branch thereof), one learns the techniques one *can* use, which is, after all, what one needs to do before one can feel at home in *any* area of mathematics. These include some reasonable approximations of element-chasing when one needs them.

And there is no objection to sometimes using a mental image in which objects are sets and morphisms are certain maps among them, since this is an important class of cases. One must merely bear in mind that, like all the mental images we use to understand mathematics, it is imperfect.

(Of course, strictly speaking, the objects of a category *are* sets, since in ZFC there are no “primitive objects”. But the morphisms of a category are not in general maps between these sets, and the set-theoretic structure of these objects is of no more relevance to the concept of category than the set-theoretic structure of “ $\frac{1}{2}$ ” is to the functional analyst.)

When one thinks of categories as *algebraic entities* themselves, one should note that the item in the definition of a category that corresponds to the *element* in the definition of a group, monoid

etc., is the *morphism*. It is on these that the composition operation, analogous to the multiplication in a group or monoid, is defined. The *object*-set of \mathbf{C} , which has no analog in groups or monoids, is essentially an index set, used to classify these elements.

While on the subject of terminology, I will mention one distinction among words (relevant to, but not limited to category theory) which many mathematicians are sloppy about, but which I try to maintain: the distinction between *composite* and *composition*. If f and g are maps of sets, or morphisms in a category, such that gf makes sense, it is their *composite*. The operation carrying the pair (f, g) to this element gf is *composition*. This is analogous to the distinction between the *sum* of two integers, $a + b$, and the operation of *addition*.

6.4. Universes. Let us now confront the problem we postponed, of how we can both handle category theory within set theory, and have category theory include concepts like “the category of sets”.

A first approach might be the following. Formulate the general definition of a category \mathbf{C} so that $\text{Ob}(\mathbf{C})$, and even the families $\mathbf{C}(X, Y)$, are *classes*. Do as much as we can in that context – the resulting animals are called *large* categories. Then go on to consider those categories in which at least the classes $\mathbf{C}(X, Y)$ are sets, and prove better results about these – they are called *legitimate* categories, and most of the examples of §6.2 are of this sort. Finally consider categories such that both the classes $\text{Ob}(\mathbf{C})$ and the classes $\mathbf{C}(X, Y)$ are sets. These are called *small* categories, and in studying them one can use the full power of set theory.

Unfortunately, in conventional set theory one has one’s hands tied behind one’s back when trying to work with large, or even legitimate categories, for there is no concept of a *collection of classes*. To get around this, one might try extending set theory. One could remove the assumption that every member of a class must be a set, so as to allow certain classes of proper classes, and extend the axioms to apply to such classes as well as sets – and one would find essentially no difficulty – except that what one had been calling “classes” are now looking more and more like sets!

So let us change their names, and call our old sets *small sets* and the classes, collections of classes, etc., *large sets*. (The word “class” itself we then restore to the function of referring to arbitrary collections of the sets, large and small, in our set theory. This includes the collection of all sets, and this again cannot itself be a member of that theory.) We would not distinguish between large and small sets in the *axioms*; we would assume the axioms of ZFC for arbitrary sets.

In formalizing the approach to which the above ideas lead, one makes the following definition, which summarizes the properties which we want the set of all “small sets” to have:

Definition 6.4.1. A universe is a set \mathcal{U} satisfying

- (i) $X \in Y \in \mathcal{U} \Rightarrow X \in \mathcal{U}$.
- (ii) $X, Y \in \mathcal{U} \Rightarrow \{X, Y\} \in \mathcal{U}$.
- (iii) $X, Y \in \mathcal{U} \Rightarrow X \times Y \in \mathcal{U}$.
- (iv) $X \in \mathcal{U} \Rightarrow \mathbf{P}(X) \in \mathcal{U}$.
- (v) $X \in \mathcal{U} \Rightarrow (\bigcup_{A \in X} A) \in \mathcal{U}$.
- (vi) $\omega \in \mathcal{U}$.
- (vii) If $X \in \mathcal{U}$ and $f: X \rightarrow \mathcal{U}$ is a function, then $\{f(x) \mid x \in X\} \in \mathcal{U}$.

The axioms of ZFC introduced in §4.4 do not guarantee the existence of a set with the above

properties (though if we replace “a set \mathcal{U} ” by “a class \mathcal{U} ” in the above definition, the class of *all* sets satisfies these conditions). Suppose, however, that a universe \mathcal{U} does exist, and suppose we look at the members of \mathcal{U} , and the relation \in among these. Then this subsystem of sets will itself satisfy the ZFC axioms (we have set up the definition of universe precisely to guarantee this) and the “operations” of power set, direct product, etc., will be the same within this “sub-set-theory” as in the total set theory. Hence we could call a member of \mathcal{U} a “small set” and an arbitrary set a “large set”, and in terms of these kinds of sets define “small category”, “legitimate category” and “large category” as above. We would define “group”, “ring”, “lattice”, “topological space”, etc., as we always did; we would further define one of these objects to be “small” if it is a member of \mathcal{U} . Then, although *all* groups still would not form a set, all *small* groups would (though not a small set). We would then define **Set**, **Group**, etc., to mean the categories of all small sets, small groups, etc., make the tacit assumption that small objects are all that “ordinary mathematics” cares about, and use large categories to study them! All that needs to be added to ZFC is an axiom saying that there exists a universe; and such an axiom is considered reasonable by set-theorists.

The above is the approach used by Mac Lane [17, pp.21-24]. However, we shall go a little further, and, following A. Grothendieck [66, §1.1], use ZFC plus an assumption that seems no less reasonable than the existence of a single universe, and more elegant. Namely,

Axiom of Universes: *Every set is a member of a universe.*

We shall assume ZFC with the Axiom of Universes from now on.

Given this set of axioms, we no longer have to think in terms of a 2-tiered set theory such that “ordinary” mathematicians work in the lower tier of small sets, and category theorists have access to the higher tier of large sets. Rather, categories, just like other mathematical objects, can exist “at any level”. But when we want to use categories to study a given sort of mathematical object, we study the category of these objects belonging to a fixed universe \mathcal{U} .

Let us now state things more formally.

Definition 6.4.2. *The concept of category will be defined as in the provisional Definition 6.1.2, but with the “system” of objects $\text{Ob}(\mathbf{C})$ explicitly meaning a set.*

Definition 6.4.3. *If \mathcal{U} is a universe, a set X will be called \mathcal{U} -small if $X \in \mathcal{U}$. A mathematical object (e.g., a group, a ring, a topological space, a category) will be called \mathcal{U} -small if it is so as a set. In addition, a category \mathbf{C} will be called \mathcal{U} -legitimate if $\text{Ob}(\mathbf{C}) \subseteq \mathcal{U}$ and for all $X, Y \in \text{Ob}(\mathbf{C})$, $\mathbf{C}(X, Y) \in \mathcal{U}$.*

The categories of \mathcal{U} -small sets, \mathcal{U} -small groups, etc., will be denoted $\mathbf{Set}_{(\mathcal{U})}$, $\mathbf{Group}_{(\mathcal{U})}$, etc..

Thus, $\mathbf{Set}_{(\mathcal{U})}$, $\mathbf{Group}_{(\mathcal{U})}$ etc., are \mathcal{U} -legitimate categories. Note that this implies that for every universe \mathcal{U}' having \mathcal{U} as a member, they are \mathcal{U}' -small categories.

But we don’t want to encumber our notation with these subscripts “ $_{(\mathcal{U})}$ ”, so we make a convention to suppress them:

Definition 6.4.4. *If we are not discussing universes, some chosen universe \mathcal{U} will be understood to be fixed, and the terms “small” and “legitimate” will mean “ \mathcal{U} -small” and “ \mathcal{U} -legitimate”. When we speak of mathematical objects (sets, groups, rings, topological spaces etc.), these will be assumed small if the contrary is not stated. As an exception, “category” will mean legitimate category if the contrary is not stated. In particular, the symbols **Set**, **Group**, **Top** etc., will*

denote the legitimate categories of all small sets, groups, topological spaces, etc..

Large will mean “not necessarily small or legitimate”.

Note that the term “large” does not specify any conditions on a set; it simply removes the assumption of smallness.

So things now look more or less as they did when we started, but we know what we are doing!

The distinctions between small and large objects will come into our considerations from time to time. For instance, when we generalize the construction of free groups and other universal objects as *subobjects of direct products*, we will see that the key condition we need is that we be able to choose an appropriate *small* set of objects over which to take the direct product.

Exercise 6.4:1. (i) Let $f: S \rightarrow \mathbb{U}$ be a function, where S is any set and \mathbb{U} is any universe. Show that $f \in \mathbb{U} \Leftrightarrow S \in \mathbb{U}$. (So in particular, any map from a member of \mathbb{U} to \mathbb{U} is a member of \mathbb{U} , but no map $\mathbb{U} \rightarrow \mathbb{U}$ is.)

(ii) Show that a large group G is small if and only if $|G|$ is small.

(iii) Show that a large category \mathbf{C} is small if and only if $\text{Ob}(\mathbf{C})$ is small, and for all $X, Y \in \text{Ob}(\mathbf{C})$, the set $\mathbf{C}(X, Y)$ is small.

Although, as we have seen, one uses *non-small* categories to study *small* mathematical objects of other sorts, the tables can be turned. For instance, we may consider closure operators on classes of small (or legitimate) categories, and the lattice of closed sets of such an operator will then be a *large* lattice.

The next couple of exercises show some properties of the class of universes. (The Axiom of Universes is, of course, to be assumed if the contrary is not stated.)

Exercise 6.4:2. (i) Show that the class of universes is not a set.

(ii) Will this same result hold if we weaken the Axiom of Universes to the statement that there is at least one universe (as in Mac Lane)? What if we use the intermediate statement that there is a universe, and that every universe is a member of a larger universe?

Exercise 6.4:3. Let us recursively define the *rank* of a set by the condition that $\text{rank}(X)$ is the least ordinal greater than all the ordinals $\text{rank}(Y)$ for $Y \in X$, and the *hereditary cardinality* of a set by the condition that $\text{her.card}(X)$ is the least cardinal that is both $\geq \text{card}(X)$ and $\geq \text{her.card}(Y)$ for all $Y \in X$.

(i) Explain why we can make these definitions. (Cf. Exercise 4.4:1.)

(ii) Show that for every universe \mathbb{U} there exists a cardinal α such that \mathbb{U} consists of all sets of hereditary cardinality $< \alpha$, and/or show that for every universe \mathbb{U} there exists a cardinal α such that \mathbb{U} consists of all sets of rank $< \alpha$.

(iii) Obtain bounds for the hereditary cardinality of a set in terms of its rank, and vice versa, and if you only did one part of the preceding point, deduce the other part.

(iv) Characterize the cardinals α which determine universes as in (ii).

Do the arguments you have used require the Axiom of Universes?

(Incidentally, the term “rank” is used as above by set theorists, who write V_α for the class of sets of rank $< \alpha$; but my use of “hereditary cardinality” above is a modification of their usage, which speaks of sets as being “hereditarily of cardinality $< \alpha$ ”. The class of sets with this property is denoted H_α .)

Exercise 6.4:4. Show that if $\mathbb{U} \neq \mathbb{U}'$ are universes, then either $\mathbb{U} \in \mathbb{U}'$ or $\mathbb{U}' \in \mathbb{U}$. Deduce that the relation “ \in or $=$ ” is a well-ordering on the class of universes. (You may wish to use some results from the preceding exercise.)

Exercise 6.4:5. Suppose that we drop from our axioms for set theory the Axiom of Infinity, and in our definition of “universe” replace the condition that every universe contain ω by the condition that every universe contain \emptyset . Show that under the new axiom-system, one can recover the Axiom of Infinity using the Axiom of Universes. Show that all but one of the sets which are “universes” under the new definition will be universes under our existing definition, and characterize the one exception.

Exercise 6.4:6. In Exercise 4.5:14 we found that “most” infinite cardinals were regular; namely, that all singular cardinals were limit cardinals; but that among limit cardinals, regular cardinals were rare; we found no example but \aleph_0 . Show now that the cardinality of any *universe* is a regular limit cardinal.

Remarks: Set-theorists call a regular limit cardinal a *weakly inaccessible cardinal*, because it cannot be “reached” from lower cardinals using either the cardinal successor operation or chains indexed by lower cardinals. The *inaccessible* cardinals, which are the cardinalities of universes, are the cardinals which cannot be reached from lower cardinals using *all* of the constructions of ZFC; i.e., the above two constructions together with the *power set* construction and the Axioms of the Empty Set and Infinity, which hand us 0 and \aleph_0 . Whether every weakly inaccessible cardinal is inaccessible depends on the assumptions one makes on one’s set theory. The student familiar with the Generalized Continuum Hypothesis will see that this assumption implies that these two concepts do coincide. Discussions of inaccessible cardinals can be found in basic texts on set theory. (For their relation to universes, cf. [4], [90]; for some alternative proposals for set-theoretic foundations of category theory, [95] and [65]; and for a proposal in the opposite direction, [15].)

Notice that introducing “large sets” has not eliminated the need for the concept of a “class” – in discussing set theory, one still needs to refer to the class of all sets; and one of the above exercises refers to the class of all universes. However, the need to refer to classes, and the difficulties arising from not being able to use set-theoretic techniques in such considerations, is greatly reduced, because for many purposes, references to large sets will now do.

We cannot be sure that the axiomatization we have adopted will be satisfactory for all the needs of category theory. It is based on the assumption that “ordinary mathematics” can be done within any universe \mathbb{U} , so that the set of all \mathbb{U} -small objects is a reasonable substitute for what was previously treated as the class of *all* objects. If some area of mathematics studied using category theory should itself require the full strength of the Axiom of Universes, then to get an adequate version of the category of “all” objects in that area, one might want to define a “second-order universe” to mean a universe \mathbb{U}' such that every set $X \in \mathbb{U}'$ is a member of a universe $\mathbb{U} \in \mathbb{U}'$, and introduce a Second Axiom of Universes, saying that every set belongs to a second-order universe! However, the fact that for pre-category-theoretic mathematics, ZFC seemed an adequate foundation suggests that the set theory we have adopted here should be good for a while.

Concerning the basic idea of what we have done, namely to assume a set theory that contains “sub-set-theories” which themselves look like traditional set theory, let us note that these are “sub-set-theories” in the strongest sense: They involve the same membership relation, the same power set operation, etc.. Set theorists often work with “sub-set-theories” in weaker senses; for example, allowing certain sets X to belong to the sub-set-theory without making all subsets of X members of the sub-set-theory. (E.g., they may allow only those that are “constructible” in some way.) The resulting model may still satisfy general axioms such as ZFC, but have other properties significantly different from those of the set theory one started with. This technique is used in proving results of the sort, “If a certain set of axioms is consistent, so is a modified set of axioms”. The distinction in question can be compared with the difference between considering a

sublattice of a lattice, which by assumption has the same meet and join operations, and considering a subset which also has least upper bounds and greatest lower bounds, and hence can again be regarded as a lattice, but where these least upper and greatest lower bounds are not the same as in the original lattice, so that the object is not a sublattice.

We will find the following concept useful at times.

Definition 6.4.5. *A mathematical object will be called quasi-small if it is isomorphic to a small object.*

Here the meaning of “isomorphic” will be determined by the context. A quasi-small set will be a set with the same cardinality as a small set; a quasi-small group is easily seen to be a group whose underlying set is a quasi-small set.

We shall now return to category theory proper. As we have indicated, our language will in general be, superficially, as before, but there is now a fixed arbitrary universe assumed in the background, and when the contrary is not stated, words such as “group” now mean “group that is small with respect to our fixed universe”, etc., while “category” means “category legitimate with respect to that universe”.

6.5. Functors. Since categories are themselves a sort of mathematical object, we should have a concept of “subcategory”, and some sort of concept of “homomorphism” between categories. The first of these concepts is described in

Definition 6.5.1. *If \mathbf{C} is a category, a subcategory of \mathbf{C} means a category \mathbf{S} such that (i) $\text{Ob}(\mathbf{S})$ is a subset of $\text{Ob}(\mathbf{C})$, (ii) for each $X, Y \in \text{Ob}(\mathbf{S})$, $\mathbf{S}(X, Y)$ is a subset of $\mathbf{C}(X, Y)$, and (iii) the composition and identity operations of \mathbf{S} are the restrictions of those of \mathbf{C} .*

Examples are clear: The category **Ab** of abelian groups is a subcategory of **Group**. Within **Monoid**, we can look at the subcategory whose objects are monoids all of whose elements are invertible (and whose morphisms are still all monoid-homomorphisms between these); this will be isomorphic to **Group**. **Lattice** is likewise isomorphic to a subcategory of **POSet**; here the lattice homomorphisms form a *proper* subset of the isotone maps. A subcategory of **POSet** with the same objects as the whole category, but a smaller set of morphisms, is the one we called **POSet**_{< . Similarly, **Set** is a subcategory of **RelSet** with the same set of objects, but a more restricted set of morphisms. The *empty category* (no objects, and hence no morphisms) is a subcategory of every category.}

The analog of homomorphism for categories is defined in

Definition 6.5.2. *If \mathbf{C} and \mathbf{D} are categories, then a functor $F: \mathbf{C} \rightarrow \mathbf{D}$ means a pair $(F_{\text{Ob}}, F_{\text{Ar}})$, where F_{Ob} is a map $\text{Ob}(\mathbf{C}) \rightarrow \text{Ob}(\mathbf{D})$, and F_{Ar} is a family $F_{\text{Ar}} = (F(X, Y))_{X, Y \in \text{Ob}(\mathbf{C})}$ of maps*

$$F(X, Y): \mathbf{C}(X, Y) \rightarrow \mathbf{D}(F_{\text{Ob}}(X), F_{\text{Ob}}(Y)) \quad (X, Y \in \text{Ob}(\mathbf{C})),$$

such that

(i) *for any two composable morphisms $X \xrightarrow{g} Y \xrightarrow{f} Z$ in \mathbf{C} , one has*

$$F(X, Z)(fg) = F(Y, Z)(f) F(X, Y)(g).$$

and

(ii) for every $X \in \text{Ob}(\mathbf{C})$,

$$F(X, X)(\text{id}_X) = \text{id}_{F_{\text{Ob}}(X)}.$$

When there is no danger of ambiguity, F_{Ob} , F_{Ar} , and $F(X, Y)$ are generally all abbreviated to F . Thus, in this notation, the last three displays become (more readably)

$$F: \mathbf{C}(X, Y) \rightarrow \mathbf{D}(F(X), F(Y)) \quad (X, Y \in \text{Ob}(\mathbf{C})),$$

$$F(fg) = F(f)F(g),$$

$$F(\text{id}_X) = \text{id}_{F(X)}.$$

How do functors arise in the prototypical situation where \mathbf{C} and \mathbf{D} consist of mathematical objects and homomorphisms among them? Since we must first specify the object of \mathbf{D} to which each object of \mathbf{C} is carried, such a functor must be based on a *construction* which gives us for each object of \mathbf{C} an object of \mathbf{D} . And in fact, most mathematical constructions, though often discussed as merely associating to each object of one sort an object of another, *also* have the property that to every morphism of objects of the first sort there corresponds naturally a morphism between the constructed objects, in a manner which satisfies just the conditions of the above definition.

Consider, for example the construction of the free group, with which we began this course. To every $X \in \text{Ob}(\mathbf{Set})$ this associates a group $F(X)$, together with a map $u_X: X \rightarrow |F(X)|$ having a certain universal property. Now if $f: X \rightarrow Y$ is a set map, it is easy to see how to get a homomorphism $F(f): F(X) \rightarrow F(Y)$. Intuitively, this homomorphism acts by “substituting $f(x)$ for x ” in elements of $F(X)$ and evaluating the results in $F(Y)$. Recall that in terms of the universal property of $F(X)$, “substituting values in a group G for the generators of $F(X)$ ” means determining a group homomorphism $F(X) \rightarrow G$ by specifying its composite with the set map $u_X: X \rightarrow |F(X)|$. In particular, for $f: X \rightarrow Y$, our above description of $F(f)$ translates to say that it is the unique group homomorphism $F(X) \rightarrow F(Y)$ such that $F(f) \circ u_X = u_Y \circ f$:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ \downarrow u_X & & \downarrow u_Y \\ |F(X)| & \xrightarrow{F(f)} & |F(Y)|. \end{array}$$

It is easy to check that when $F(f)$ is defined in this way for each set-map f , one has $F(fg) = F(f)F(g)$ and $F(\text{id}_X) = \text{id}_{F(X)}$. Hence the free group construction gives a functor $F: \mathbf{Set} \rightarrow \mathbf{Group}$.

Looking in the same way at the construction of *abelianization*, associating to each group G the abelian group $G^{\text{ab}} = G/[G, G]$, we see that every group homomorphism $f: G \rightarrow H$ yields a homomorphism of abelian groups $f^{\text{ab}}: G^{\text{ab}} \rightarrow H^{\text{ab}}$, describable either concretely in terms of cosets as in Exercise 3.4:3, or by a commutative diagram construction using the universal property of the canonical homomorphism $G \rightarrow G^{\text{ab}}$. The constructions of free semilattices, universal abelianizations of rings, etc., give similar examples.

Like most mathematical concepts, the concept of functor also has “trivial” examples, that by

themselves would not justify the general definition, yet which turn out to have important roles in the theory. The “construction” associating to every group G its underlying set $|G|$ is a functor $\mathbf{Group} \rightarrow \mathbf{Set}$, since homomorphisms of groups certainly give maps of underlying sets. One similarly has underlying-set functors from \mathbf{Ring}^1 , $\mathbf{Lattice}$, \mathbf{Top} , \mathbf{POSet} , etc., to \mathbf{Set} . These all belong to the class of constructions called “forgetful functors”. Those listed above “forget” all structure on the object, and so give functors to \mathbf{Set} ; other forgetful functors we have seen are the construction $G \mapsto G_{\text{md}}$ of §3.11, taking a group $(|G|, \cdot, {}^{-1}, e)$ to the monoid $(|G|, \cdot, e)$, which “forgets” the inverse operation, and the construction taking a ring to its underlying additive group, or to its underlying multiplicative monoid.

The term “forgetful functor” is not a technical one, so one cannot say precisely whether it should be applied to constructions like the one taking a lattice to its “underlying” partially ordered set (“underlying” in quotes because the partial ordering is not part of the 3-tuple formally defining the lattice); but in any case, this is another example of a functor. I likewise don’t know whether one would apply the term “forgetful” to the inclusion of the subcategory \mathbf{Ab} in the category \mathbf{Group} , which might be said to “forget” that the groups are abelian, but this too, and indeed, the inclusion of any subcategory in any category, is clearly a functor. In particular, every category \mathbf{C} has an identity functor, $\text{Id}_{\mathbf{C}}$, taking each object and each morphism to itself.

If, instead of looking at the whole underlying set of a group, we consider the set of its elements of exponent 2, we get another example of a functor $\mathbf{Group} \rightarrow \mathbf{Set}$; the reader should verify that every group homomorphism does indeed give a map between the corresponding sets.

If R is a ring, the *opposite* ring R^{op} is defined to have the same underlying set, and the same operations $+$, $-$, 0 , 1 as R , but reversed multiplication: $x*y = yx$. A ring homomorphism $f: R \rightarrow S$ will also be a homomorphism $R^{\text{op}} \rightarrow S^{\text{op}}$, and we see that this makes $(\)^{\text{op}}$ a functor $\mathbf{Ring}^1 \rightarrow \mathbf{Ring}^1$; one which, composed with itself, gives the identity functor. One has similar opposite-multiplication constructions for monoids and groups. The definitions of the opposite (or dual) of a partially ordered set or lattice give functors with similar properties.

Recall that \mathbf{HtpTop} is defined to have the same objects as \mathbf{Top} , but has for morphisms *equivalence classes* of continuous maps under homotopy. Thus we have a functor $\mathbf{Top} \rightarrow \mathbf{HtpTop}$ which preserves objects, and sends every morphism to its homotopy class.

We have mentioned *diagram categories*, such as the “commuting square diagram” $\begin{array}{ccc} & \rightarrow & \\ \downarrow & & \downarrow \\ & \rightarrow & \end{array}$ which is useful because “images” of it in any category \mathbf{C} correspond to commuting squares of objects and arrows in \mathbf{C} . We can now say this more precisely: Commuting squares in \mathbf{C} correspond to *functors* from this diagram-category into \mathbf{C} .

Let us note a few examples of mathematical constructions that are *not* functors. These tend to be of two sorts: those in which morphisms from one object to another can destroy some of the properties used by the construction, and those that involve arbitrary choices. We have noted that the construction associating to every group G the set of elements of exponent 2, $\{x \in |G| \mid x^2 = e\}$, is a functor $\mathbf{Group} \rightarrow \mathbf{Set}$. However, if we define $T(G)$ to be the set of elements of *order* 2, $\{x \in |G| \mid x^2 = e, x \neq e\}$, we find that a group homomorphism f may take some of these elements to the identity element, so there is no natural way to define “ $T(f)$ ”. Similarly, the important group-theoretic construction of the *center* $Z(G)$ of a group G (the subgroup of elements $a \in |G|$ that commute with all elements of G) is not functorial, because even if a is in the center of G , when we apply a homomorphism $f: G \rightarrow H$, some elements of H which are not in the image of G may fail to commute with $f(a)$. The construction Aut , taking a group G to its automorphism group, is also not a functor, roughly because when we map G into another

group H , there is no guarantee that H will have all the “symmetries” that G does.

Some constructions of these sorts can be “made into” functors by modifying the choice of domain category so as to restrict the morphisms thereof to maps that don’t “disturb” the structure involved. Thus, the construction associating to every group its set of elements of order 2 does give a functor $\mathbf{Group}_{\text{inj}} \rightarrow \mathbf{Set}$, if we define $\mathbf{Group}_{\text{inj}}$ to be the category whose objects are groups and whose morphisms are *injective* (one-to-one) group homomorphisms. The construction of the center likewise gives a functor $\mathbf{Group}_{\text{surj}} \rightarrow \mathbf{Group}$, where the morphisms of $\mathbf{Group}_{\text{surj}}$ are the *surjective* group homomorphisms. One may make \mathbf{Aut} a functor by restricting morphisms to *isomorphisms* of groups.

An example of the other sort, where a construction is not a functor because it involves choices that cannot be made in a canonical way, is that of finding a basis for a vector space. Even limiting ourselves to finite-dimensional vector spaces, so that bases may be constructed without the Axiom of Choice, the finite sequence of choices made is still arbitrary, so that if one chooses a basis B_V for a vector space V , and a basis B_W for a vector space W , there is no natural way to associate to every linear map $V \rightarrow W$ a set map $B_V \rightarrow B_W$.

In the above discussion we have merely indicated where straightforward attempts to make these constructions into functors went wrong. In several of the following exercises you are asked to prove more precise negative results.

Exercise 6.5:1. (i) Show that there can be no functor $F: \mathbf{Group} \rightarrow \mathbf{Set}$ taking each group to the set of its elements of order 2, no matter how F is made to act on morphisms.

On the other hand,

(ii) Show how to define a functor $\mathbf{Group} \rightarrow \mathbf{RelSet}$ taking every group to its set of elements of order 2. (Since \mathbf{RelSet} is an unfamiliar category, verify explicitly all parts of the definition of functor.)

Exercise 6.5:2. (i) Show that there can be no functor $F: \mathbf{Group} \rightarrow \mathbf{Group}$ taking each group to its center.

(ii) Can one construct a functor $\mathbf{Group} \rightarrow \mathbf{RelSet}$ taking every group to the set of its central elements?

Exercise 6.5:3. (i) Give an example of a group homomorphism $f: G \rightarrow H$ and an automorphism a of G such that there does not exist a unique automorphism a' of H such that $a'f = fa$. In fact, find such examples with f one-to-one but not onto, and with f onto but not one-to-one, and in each of these situations, if possible, an example where such a' does not exist, and an example where such a' exists but is not unique. (If you cannot get an example of one of the above combinations, can you show that it does not occur?)

(ii) Find similar examples involving partially ordered sets in place of groups.

(iii) Prove that there is no functor from \mathbf{Group} (alternatively, from \mathbf{POSet}) to \mathbf{Set} (or even to \mathbf{RelSet}) taking each object to its set of automorphisms.

Exercise 6.5:4. If K is a field, let \bar{K} denote the algebraic closure of K . We recall that any field homomorphism $f: K \rightarrow L$ can be extended to a homomorphism of algebraic closures, $\bar{f}: \bar{K} \rightarrow \bar{L}$.

(i) Show, however, that there is no way to choose an extension \bar{f} of each field homomorphism f so as to make the algebraic closure construction a functor.

(ii) If we remove the restriction that \bar{f} be an extension of f , can we make algebraic closure a functor?

The next exercise is instructive and entertaining. A full solution to the second part is difficult, but one can get many interesting partial results.

Exercise 6.5:5. Let **FSet** denote the subcategory of **Set** having for objects the finite sets, and for morphisms all set maps among these.

- (i) Show that every functor F from **FSet** to **FSet** determines a unique function f from the nonnegative integers to the nonnegative integers, such that for every finite set X , $\text{card}(F(X)) = f(\text{card}(X))$.
- (ii) Investigate *which* integer-valued functions f can occur as the functions associated to such functors. If possible, determine necessary and sufficient conditions on f for such an F to exist.

Note that given functors $\mathbf{C} \xrightarrow{G} \mathbf{D} \xrightarrow{F} \mathbf{E}$ between any three categories, we can form the *composite* functor $\mathbf{C} \xrightarrow{FG} \mathbf{E}$ taking each object X to $F(G(X))$ and each morphism f to $F(G(f))$. Composition of functors is clearly associative, and identity functors satisfy the identity laws, so we have a “category of categories”! This is named in

Definition 6.5.3. *Cat* will denote the (legitimate) category whose objects are all small categories, and where for two small categories \mathbf{C} and \mathbf{D} , $\mathbf{Cat}(\mathbf{C}, \mathbf{D})$ is the set of all functors $\mathbf{C} \rightarrow \mathbf{D}$, with composition of functors defined as above.

You might be disappointed with this definition, since only a few of the categories we have mentioned have been small (the diagram-categories, and the categories $S_{\mathbf{cat}}$ and $P_{\mathbf{cat}}$ constructed from monoids S and partially ordered sets P). Thus, **Cat** would appear to be of limited importance. But here the Axiom of Universes comes to our aid. The universe \mathbb{U} relative to which we have defined “small category” is arbitrary. If we want to study the categories of all groups, rings, etc., belonging to a universe \mathbb{U} , and functors among these categories, we may choose a universe \mathbb{U}' having \mathbb{U} as a member, and note that the abovementioned categories, and indeed all \mathbb{U} -legitimate categories, are \mathbb{U}' -small, hence are objects of $\mathbf{Cat}_{(\mathbb{U}')}$. Thus we can apply general results about the construction **Cat** to this situation.

For some purposes, it might also be useful to give a name to the category of all \mathbb{U} -legitimate categories, which lies strictly between $\mathbf{Cat}_{(\mathbb{U})}$ and $\mathbf{Cat}_{(\mathbb{U}')}$, but we shall not do so here.

Considering functors as “homomorphisms” among categories, we should like to define properties of functors analogous to “one-to-one-ness” and “onto-ness”. The complication is that a functor acts both on objects and on morphisms. We have observed that it is the *morphisms* in a category that are like the *elements* of a group or monoid; this leads to the pair of concepts named below. They are not the only analogs of one-one-ness and onto-ness that one ever uses, but they are the most important:

Definition 6.5.4. Let $F: \mathbf{C} \rightarrow \mathbf{D}$ be a functor.

F is called *faithful* if for all $X, Y \in \text{Ob}(\mathbf{C})$, the map $F(X, Y): \mathbf{C}(X, Y) \rightarrow \mathbf{D}(F(X), F(Y))$ is one-to-one.

F is called *full* if for all $X, Y \in \text{Ob}(\mathbf{C})$, the map $F(X, Y): \mathbf{C}(X, Y) \rightarrow \mathbf{D}(F(X), F(Y))$ is onto.

A subcategory of \mathbf{C} is said to be *full* if the corresponding inclusion functor is full.

Thus, a full subcategory of \mathbf{C} is determined by specifying a subset of the object-set; the morphisms of the subcategory are *all* the morphisms among these objects. The subcategory **Ab** of **Group** is an example. Some examples of nonfull subcategories are $\mathbf{Set} \subseteq \mathbf{RelSet}$ and $\mathbf{POSet}_{<} \subseteq \mathbf{POSet}$. The inclusion of a full subcategory in a category is a full and faithful functor, while the inclusion of a nonfull subcategory is a faithful functor, but is not full. The reader should

verify that most of our examples of forgetful functors are faithful but not full, as is, also, the free-group functor $\mathbf{Set} \rightarrow \mathbf{Group}$. The functor $\mathbf{Top} \rightarrow \mathbf{HtpTop}$ which takes every object (topological space) to itself, and each morphism to its *homotopy class*, is an example of a functor that is full but not faithful. The functor associating to every group the set of its elements of exponent 2 is neither full nor faithful.

Exercise 6.5:6. Show that the abelianization construction, $\mathbf{Group} \rightarrow \mathbf{Ab}$ is neither full nor faithful.

Exercise 6.5:7. Is the functor $\mathbf{Monoid} \rightarrow \mathbf{Group}$ associating to every monoid its group of invertible elements full? Faithful?

Exercise 6.5:8. (i) Show that the construction associating to each partially ordered set P the category $P_{\mathbf{cat}}$ can be made in a natural way into a functor $F: \mathbf{POSet} \rightarrow \mathbf{Cat}$, and that as such it is full and faithful. Essentially, this says that the concept of functor, when restricted to the class of categories that correspond to partially ordered sets, just gives the concept of isotone map between these sets!

(ii) Which isotone maps between partially ordered sets correspond under F to full functors? To faithful functors?

(iii) Show similarly that the construction associating to each monoid S the category $S_{\mathbf{cat}}$ is a full and faithful functor $E: \mathbf{Monoid} \rightarrow \mathbf{Cat}$. Which monoid homomorphisms are sent by E to full, respectively faithful functors?

Exercise 6.5:9. Show that for $F: \mathbf{C} \rightarrow \mathbf{D}$ a functor, neither of the following conditions implies the other: (a) F is full, (b) for all $X, Y \in \mathbf{Ob}(\mathbf{D})$ and $f \in \mathbf{D}(X, Y)$ there exist $X_0, Y_0 \in \mathbf{Ob}(\mathbf{C})$ and $f_0 \in \mathbf{C}(X_0, Y_0)$ such that $F(X_0) = X$, $F(Y_0) = Y$, and $F(f_0) = f$.

In §6.1 we sketched a way of “concretizing” any small category \mathbf{C} (Exercise 6.1:1 and preceding discussion). Let us make the details precise now.

Definition 6.5.5. A concrete category *means* a category \mathbf{C} given with a faithful functor $U: \mathbf{C} \rightarrow \mathbf{Set}$ (a “concretization functor”). (More formally, one would say that the concrete category is the ordered pair (\mathbf{C}, U) .)

So the result sketched earlier was that given any small category \mathbf{C} , there exists a faithful functor $U: \mathbf{C} \rightarrow \mathbf{Set}$. The idea was to let the family of representing sets – in our present language, the system of sets $U(X)$ ($X \in \mathbf{Ob}(\mathbf{C})$) – be “generated” by a family of elements $z_Y \in U(Y)$, one for each $Y \in \mathbf{Ob}(\mathbf{C})$, so that the general element of $U(X)$ would look like $U(a)(z_Y)$ ($Y \in \mathbf{Ob}(\mathbf{C})$, $a \in \mathbf{C}(Y, X)$); and to impose no additional relations on these elements, so that they are all distinct.

Let us use the ordered pair (Y, a) for the element that is to become $U(a)(z_Y)$. Then we should define U to take $X \in \mathbf{Ob}(\mathbf{C})$ to $\{(Y, a) \mid Y \in \mathbf{Ob}(\mathbf{C}), a \in \mathbf{C}(Y, X)\}$. Given $b \in \mathbf{C}(X, W)$, we see that $U(b)$ should take $(Y, a) \in U(X)$ to $(Y, ba) \in U(W)$. It is easy to verify that this defines a faithful functor $U: \mathbf{C} \rightarrow \mathbf{Set}$, proving

Theorem 6.5.6 (Cayley’s Theorem for small categories). *Every small category admits a concretization, i.e., a faithful functor to the category of small sets.* \square

Exercise 6.5:10. Verify that the above construction U is a functor, and is faithful. Which element of each set $U(Y)$ corresponds to the z_Y of our motivating discussion?

Incidentally, if we had required that categories have disjoint morphism-sets, we could have

dropped the Y 's from the pairs (Y, a) , since each a would determine its domain. Then we could simply have taken $U(X) = \cup_{Y \in \text{Ob}(\mathbf{C})} \mathbf{C}(Y, X)$.

It is natural to hope for stronger results, so you can try

Exercise 6.5:11. (i) Does every legitimate category admit a concretization – a faithful functor to the (legitimate) category of small sets? (Obviously, most of those we are familiar with do.)

Since this question involves “big” cardinalities, you might prefer to examine a mini-version of the same problem:

(ii) Suppose \mathbf{C} is a category with countably many objects, and such that for all $X, Y \in \text{Ob}(\mathbf{C})$, the set $\mathbf{C}(X, Y)$ is finite. Must \mathbf{C} admit a faithful functor into the category of finite sets?

(iii) If the answer to either question is negative, can you find necessary and sufficient conditions on \mathbf{C} for such concretizations to exist?

Of course, a given concretizable category will admit many concretizations, just as a given group has many faithful representations by permutations.

Recall that the proof of Theorem 6.5.6 sketched above came out of our proof of the corresponding result for “bimonoids”, and that in trying to prove that result, we first wondered whether it would suffice to adjoin a generator in just one of the two representing sets X and Y , but realized that the resulting representation of the bimonoid would not necessarily be faithful. Given a category \mathbf{C} and an object Y of \mathbf{C} , we can similarly construct a functor $U: \mathbf{C} \rightarrow \mathbf{Set}$ by introducing only one generator $z_Y \in U(Y)$, again with no relations imposed among the elements $U(a)(z_Y)$. Though these functors also generally fail to be faithful, they will play an important role in our subsequent work. Note that each such functor is the “part” of the construction we used in Theorem 6.5.6 consisting of the elements $U(a)(z_Y)$ for one fixed Y . With Y fixed, each such element is determined by $a \in \mathbf{C}(Y, X)$, so U may be described as taking each object X to the set $\mathbf{C}(Y, X)$; hence its name, coming from the term “hom-sets” for the sets $\mathbf{C}(Y, X)$:

Definition 6.5.7. For $Y \in \text{Ob}(\mathbf{C})$, the hom functor induced by Y , $h_Y: \mathbf{C} \rightarrow \mathbf{Set}$, is defined on objects by

$$h_Y(X) = \mathbf{C}(Y, X) \quad (X \in \text{Ob}(\mathbf{C})),$$

while for a morphism $b \in \mathbf{C}(X, W)$, $h_Y(b)$ is defined to carry $a \in \mathbf{C}(Y, X)$ to $ba \in \mathbf{C}(Y, W)$.

Examples: On the category **Group**, the functor $h_{\mathbb{Z}}$ takes each group G to **Group** (\mathbb{Z}, G) . But a homomorphism from \mathbb{Z} to G is determined by what it does on the generator $1 \in |\mathbb{Z}|$, so the elements of $h_{\mathbb{Z}}(G)$ correspond to the elements of the underlying set of G ; i.e., $h_{\mathbb{Z}}$ is essentially the underlying set functor. You should verify that its behavior on morphisms also agrees with that functor. Similarly, $h_{\mathbb{Z}_2}$ may be identified with the functor taking each group to the set of its elements of exponent 2.

Recalling that $2 \in \text{Ob}(\mathbf{Set})$ is a 2-element set, we see that $h_2: \mathbf{Set} \rightarrow \mathbf{Set}$ is essentially the construction $X \mapsto X^2$.

For a topological example, consider the category of topological spaces with basepoint, and homotopy classes of basepoint-preserving maps, and let $(S^1, 0)$ denote the circle with a basepoint chosen. Then $h_{(S^1, 0)}(X, x_0) = |\pi_1(X, x_0)|$. (Of course, the most interesting thing about $\pi_1(X, x_0)$ is its group structure. How *this* can be described category-theoretically we shall discover in Chapter 9!)

In the last few paragraphs, we have said a couple of times that a certain functor is “essentially”

a certain construction. What we mean should be intuitively clear; we will see how to make these statements precise in §6.9.

6.6. Contravariant functors, and functors of several variables. Consider the construction associating to every set X the additive group \mathbb{Z}^X of integer-valued functions on X , with pointwise operations. This takes objects of **Set** to objects of **Ab**, but given a set map $f: X \rightarrow Y$, there is not a natural map $\mathbb{Z}^X \rightarrow \mathbb{Z}^Y$ – rather, there is a homomorphism $\mathbb{Z}^Y \rightarrow \mathbb{Z}^X$ carrying each integer-valued function a on Y to the function af on X .

There are many similar examples – the construction associating to any set X the Boolean algebra $(\mathbf{P}(X), \cup, \cap, {}^c, \emptyset, X)$ of its subsets, the construction associating to a set X the lower semilattice $(\mathbf{E}(X), \cap)$ of equivalence relations on X , the construction associating to a vector space V its dual V^* , the construction associating to a commutative ring the partially ordered set of its prime ideals. All have the property that from a map going one way among the given objects, one gets a map going the other way among constructed objects. It is clear that these constructions take identity maps to identity maps and composite maps to composite maps (though the order of composition must be reversed because of the reversal of the direction of the maps). These properties look like the definition of a functor turned backwards. Let us set up a definition to cover this:

Definition 6.6.1. *If \mathbf{C} and \mathbf{D} are categories, then a contravariant functor $F: \mathbf{C} \rightarrow \mathbf{D}$ means a pair $(F_{\text{Ob}}, F_{\text{Ar}})$, where F_{Ob} (written F when there is no danger of ambiguity) is a map $\text{Ob}(\mathbf{C}) \rightarrow \text{Ob}(\mathbf{D})$, and F_{Ar} is a family of maps*

$$F(X, Y): \mathbf{C}(X, Y) \rightarrow \mathbf{D}(F(Y), F(X)) \quad (X, Y \in \text{Ob}(\mathbf{C})),$$

such that (also abbreviating these maps $F(X, Y)$ to F),

(i) *for any two composable morphisms $X \xrightarrow{g} Y \xrightarrow{f} Z$ in \mathbf{C} , one has*

$$F(fg) = F(g)F(f) \text{ in } \mathbf{D},$$

and

(ii) *for every $X \in \text{Ob}(\mathbf{C})$, one has*

$$F(\text{id}_X) = \text{id}_{F(X)}.$$

Functors of the sort defined in the preceding section are called covariant functors when one wants to contrast them with contravariant functors. When the contrary is not indicated “functor” will still mean covariant functor.

It is easy to see that a composite of two contravariant functors is a covariant functor, while a composite of a covariant and a contravariant functor, in either order, is a contravariant functor.

Contravariant functors can in fact be expressed in terms of covariant functors, thus eliminating the need to prove results separately for them. We shall do this with the help of

Definition 6.6.2. *If \mathbf{C} is a category, then \mathbf{C}^{op} will denote the category defined by*

$$\begin{aligned} \text{Ob}(\mathbf{C}^{\text{op}}) &= \text{Ob}(\mathbf{C}), & \mathbf{C}^{\text{op}}(X, Y) &= \mathbf{C}(Y, X), \\ \mu(\mathbf{C}^{\text{op}})(f, g) &= \mu(\mathbf{C})(g, f), & \text{id}(\mathbf{C}^{\text{op}})_X &= \text{id}(\mathbf{C})_X. \end{aligned}$$

Thus, a *contravariant* functor $\mathbf{C} \rightarrow \mathbf{D}$ is equivalent to a covariant functor $\mathbf{C}^{\text{op}} \rightarrow \mathbf{D}$. Of course, one could also describe it as equivalent to a covariant functor $\mathbf{C} \rightarrow \mathbf{D}^{\text{op}}$, and at this point we have no way of deciding which reduction is preferable. However, we shall see soon that putting the “op” on the domain category is more convenient.

As in the theory of partially ordered sets, the “opposite” construction introduced above allows us to dualize results. Whenever we have proved a result about a general category \mathbf{C} , the statement obtained by reversing the directions of all morphisms and the orders of all compositions is also a theorem, which may be proved by applying the original theorem to \mathbf{C}^{op} .

There is a slight notational difficulty in dealing with a category \mathbf{C}^{op} , while referring also to the original category \mathbf{C} , for though in the formal definition given above we could distinguish the two composition operations as $\mu(\mathbf{C})$ and $\mu(\mathbf{C}^{\text{op}})$, the usual notation for composition, $f \circ g$ or fg , does not allow such a distinction. There are various ways of getting around this problem. One can use a modified symbol, such as \circ^{op} or $*$, for the composition of \mathbf{C}^{op} . Or one can continue to denote composition by juxtaposition, but use different symbols for the same objects and morphisms when considered as elements of \mathbf{C} and of \mathbf{C}^{op} ; e.g., let the morphism written $f \in \mathbf{C}(X, Y)$ also be written $\tilde{f} \in \mathbf{C}^{\text{op}}(\tilde{Y}, \tilde{X})$, so that one can write $\widetilde{fg} = \tilde{g}\tilde{f}$, relying on the convention that the meaning of juxtaposition is determined by context – specifically, by the structure to which the elements being juxtaposed belong. Still other solutions are possible. E.g., one could be daring, and denote the same composite by fg in both \mathbf{C} and \mathbf{C}^{op} , using different conventions, $fg = \mu(f, g)$ in \mathbf{C} and $fg = \mu(g, f)$ in \mathbf{C}^{op} ; i.e., writing morphisms with domains “on the right” in one category and “on the left” in the other.

Most often, one avoids the problem by not writing equations in \mathbf{C}^{op} . One uses this category as an auxiliary construct in discussing contravariant functors and in dualizing results, but avoids dealing explicitly with objects and morphisms inside it.

In these notes, we shall regularly write a contravariant functor from \mathbf{C} to \mathbf{D} as $F: \mathbf{C}^{\text{op}} \rightarrow \mathbf{D}$, where F is a covariant functor on \mathbf{C}^{op} , and shall take advantage of the principle of duality mentioned. These are the main uses we shall make of the $^{\text{op}}$ construction; in the rare cases where we have to work explicitly inside \mathbf{C}^{op} , we will generally use modified symbols such as \tilde{X} , \tilde{f} (or X^{op} , f^{op}) for objects and morphisms in \mathbf{C}^{op} .

Note that in the category of categories, \mathbf{Cat} , the morphisms are the *covariant* functors.

Exercise 6.6:1. (i) Show how to make $^{\text{op}}$ a functor R from \mathbf{Cat} to \mathbf{Cat} . Is R a covariant or a contravariant functor?

(ii) Let $R: \mathbf{Cat} \rightarrow \mathbf{Cat}$ be as in part (i), let $R': \mathbf{POSet} \rightarrow \mathbf{POSet}$ be the functor taking every partially ordered set P to the opposite partially ordered set P^{op} , and let $C: \mathbf{POSet} \rightarrow \mathbf{Cat}$ denote the functor taking each partially ordered set P to the category $P_{\mathbf{cat}}$ (§6.2). Show that $RC = CR'$. This means that the construction of the opposite of a partially ordered set is essentially a case of the construction of the opposite of a category!

(iii) State the analogous result with *monoids* in place of partially ordered sets.

We noted in earlier chapters that given a set map $X \rightarrow Y$, there are ways of getting both a map $\mathbf{P}(X) \rightarrow \mathbf{P}(Y)$ and a map $\mathbf{P}(Y) \rightarrow \mathbf{P}(X)$ (where \mathbf{P} denotes the power-set construction). The next few exercises look at these and some similar situations.

Exercise 6.6:2. (i) Write down explicitly how to get from a set map $f: X \rightarrow Y$ a set map $P_1(f): \mathbf{P}(X) \rightarrow \mathbf{P}(Y)$ and a set map $P_2(f): \mathbf{P}(Y) \rightarrow \mathbf{P}(X)$. Show that these constructions make the power set construction a functor $P_1: \mathbf{Set} \rightarrow \mathbf{Set}$ and a functor $P_2: \mathbf{Set}^{\text{op}} \rightarrow \mathbf{Set}$ respectively. (These are called the *covariant* and *contravariant power set functors*.)

(ii) Examine what structure on $\mathbf{P}(X)$ is *respected* by maps of the form $P_1(f)$ and $P_2(f)$ defined as above. In particular, determine whether each sort of map always respects the operations of finite meets, finite joins, empty meet, empty join, unions of chains, intersections of chains, complements, and the relations “ \subseteq ” and “ \subset ” in power-sets $\mathbf{P}(X)$. (Cf. Exercise 5.1:11. If you are familiar with the standard topologization of $\mathbf{P}(X)$, you can also investigate whether maps of the form $P_1(f)$ and $P_2(f)$ are continuous.) Accordingly, determine whether the constructions P_1 and P_2 which we referred to above as functors from \mathbf{Set} , respectively \mathbf{Set}^{op} , to \mathbf{Set} , can in fact be made into functors from \mathbf{Set} and/or \mathbf{Set}^{op} to $\vee\text{-Semilat}$, to \mathbf{Bool}^1 , etc..

Exercise 6.6:3. Investigate similarly the construction associating to every set X the set $\mathbf{E}(X)$ of *equivalence relations* on X . I.e., for a set map $f: X \rightarrow Y$, look for functorial ways of inducing maps in one or both directions between the sets $\mathbf{E}(X)$, $\mathbf{E}(Y)$, and determine what structure on these sets is respected by each such construction.

Exercise 6.6:4. (i) Do the same for the construction associating to every group G the set of subgroups of G .

(ii) Do the same for the construction associating to every group G the set of *normal* subgroups of G .

As with covariant functors, there is an important class of contravariant functors which one can define on every category:

Definition 6.6.3. For any category \mathbf{C} and any object $Y \in \text{Ob}(\mathbf{C})$, the contravariant hom functor induced by Y , $h^Y: \mathbf{C}^{\text{op}} \rightarrow \mathbf{Set}$, is defined on objects by

$$h^Y(X) = \mathbf{C}(X, Y) \quad (X \in \text{Ob}(\mathbf{C})),$$

while for a morphism $b \in \mathbf{C}(W, X)$ the morphism $h^Y(b): \mathbf{C}(X, Y) \rightarrow \mathbf{C}(W, Y)$ is defined to carry $a \in \mathbf{C}(X, Y)$ to $ab \in \mathbf{C}(W, Y)$. (The functor h_Y which we previously named “the hom functor induced by Y ” will henceforth be called “the covariant hom functor induced by Y ”.)

Examples: Let $\mathbf{C} = \mathbf{Set}$, and let Y be the set $2 = \{0, 1\}$. Recall that every map from a set X into 2 is the characteristic function of a unique subset $S \subseteq X$. Hence $\mathbf{Set}(X, 2)$ can be identified with $\mathbf{P}(X)$. The reader should verify that the behavior of $h^2: \mathbf{Set} \rightarrow \mathbf{Set}$ on morphisms is exactly that of the contravariant power-set functor.

Let k be a field, and in the category $k\text{-Mod}$ of k -vector spaces, let k denote this field considered as a one-dimensional vector space. Then for any space V , $h^k(V)$ is the underlying set of the *dual* vector space, and for any linear map $b: W \rightarrow V$, $h^k(b)$ is the induced map from the dual of the space W to the dual of the space V .

Let $\mathbb{R} \in \text{Ob}(\mathbf{Top})$ denote the real line. Then $h^{\mathbb{R}}$ is the construction associating to every topological space X the set of continuous real-valued functions on X . One can vary this example using categories of differentiable manifolds and differentiable maps, etc., in place of \mathbf{Top} .

Here are three examples for students familiar with more specialized topics:

In the category of commutative algebras over the rational numbers, if \mathbb{C} denotes the algebra of complex numbers, then $h^{\mathbb{C}}$ is the functor associating to every algebra the set of its “complex-valued points”, its *classical spectrum*. In particular, if R is an algebra presented by generators x_0, \dots, x_{n-1} and relations $p_0 = 0, \dots, p_{m-1} = 0$, then $h^{\mathbb{C}}(R)$ can be identified with the solution-set of the system of polynomial equations $p_0 = 0, \dots, p_{m-1} = 0$ in complex n -space.

If $\mathbf{LocCpAb}$ is the category of locally compact abelian groups, and $S = \mathbb{R}/\mathbb{Z}$ is the circle group, then $h^S(A)$ is the underlying set of the *Pontryagin dual* of the group A [111, §1.7]. (In

the study of nontopological abelian groups, the object $S = \mathbb{Q}/\mathbb{Z}$ plays a somewhat similar role [31, p.145, Remark 2].)

Finally, in the category **HtpTop**, $h^{S^n}(X)$ (where S^n denotes the n -sphere) gives the underlying set of the n th *cohomotopy* group, $\pi^n(X)$.

Exercise 6.6:5. Let $2 \in \text{Ob}(\mathbf{POSet})$ denote the set $2 = \{0, 1\}$, ordered in the usual way.

- (i) Show that $h^2: \mathbf{POSet}^{\text{op}} \rightarrow \mathbf{Set}$ is faithful.
- (ii) Show that for $P \in \text{Ob}(\mathbf{POSet})$, the set $h^2(P)$ can be made a *lattice* with a greatest and a least element, under pointwise operations. Show that in this way h^2 induces a functor $A: \mathbf{POSet}^{\text{op}} \rightarrow \mathbf{Lattice}^{0,1}$, where $\mathbf{Lattice}^{0,1}$ denotes the category of lattices with greatest and least elements, and lattice homomorphisms respecting these elements.
- (iii) Let us also write $2 \in \text{Ob}(\mathbf{Lattice}^{0,1})$ for the 2-element lattice! Thus we get a functor $h^2: (\mathbf{Lattice}^{0,1})^{\text{op}} \rightarrow \mathbf{Set}$. Show that this functor is *not* faithful.
- (iv) Show that for $L \in \text{Ob}(\mathbf{Lattice}^{0,1})$, the set $h^2(L)$ is not in general closed under pointwise meet or join, and may not contain a greatest or least element, but that if we partially order lattice homomorphisms by pointwise comparison, h^2 yields a functor $B: (\mathbf{Lattice}^{0,1})^{\text{op}} \rightarrow \mathbf{POSet}$.
- (v) Show that for P a *finite* partially ordered set, $B(A(P)) \cong P$.

The above is just a teaser regarding this pair of functors. The student can discover more for him or herself now, or wait till we resume this investigation in §9.12 with more general concepts at our disposal.

Exercise 6.6:6. Following up on the idea of Exercise 6.5:5, observe that every *contravariant* functor from the category **FSet** of finite sets into itself also determines a nonnegative integer-valued function on the nonnegative integers. Investigate which functions on the nonnegative integers arise as functions associated with contravariant functors.

Exercise 6.6:7. Let **RelFSet** denote the full subcategory of **RelSet** whose objects are finite sets. Investigate similarly the integer-valued functions associated with functors **RelFSet** \rightarrow **FSet**, **FSet** \rightarrow **RelFSet**, and **RelFSet** \rightarrow **RelFSet**. In these cases, it does not matter whether we look at covariant or contravariant functors – why not?

Exercise 6.6:8. We have noted that a composite of two contravariant functors is a covariant functor, etc.. But in terms of the description of contravariant functors as covariant functors $\mathbf{C}^{\text{op}} \rightarrow \mathbf{D}$, it is not clear how to formally describe the composite of two contravariant functors (or a composite of the form (contravariant functor) $^{\circ}$ (covariant functor)). Show how to reduce these cases to composition of covariant functors, with the help of Exercise 6.6:1(i).

There are still some types of well-behaved mathematical constructions which we have not yet fitted into our functorial scheme: (a) Given a *pair* of sets (A, B) , we can form the *product* set $A \times B$. We likewise have product constructions for groups, rings, topological spaces, etc., *coproducts* for most of the same types of objects, and the *tensor product* construction on abelian groups. (b) From two objects A and B of any category \mathbf{C} , one gets $\mathbf{C}(A, B) \in \text{Ob}(\mathbf{Set})$. (c) There are also constructions that combine objects of different categories. For instance, from a commutative ring R and a set X , one can form the *polynomial ring* over R in an X -tuple of indeterminates, $R[X]$.

In each of these cases, maps on the given objects yield maps on the constructed objects. In cases (a) and (c), the maps of constructed objects go the same way as the maps of the given objects, while in case (b) the direction depends on which argument one varies: A morphism $Y \rightarrow Y'$ yields a map $\mathbf{C}(X, Y) \rightarrow \mathbf{C}(X, Y')$, but a morphism $X \rightarrow X'$ yields a map $\mathbf{C}(X', Y) \rightarrow \mathbf{C}(X, Y)$.

It is natural to call such constructions *functors of two variables*. Like the concept of contravariant functor, that of a functor of more than one variable can be reduced to our original definition of functor via an appropriate construction on categories.

Definition 6.6.4. Let $(\mathbf{C}_i)_{i \in I}$ be a family of categories. Then the product category $\prod_{i \in I} \mathbf{C}_i$ will mean the category \mathbf{C} defined by

$$\begin{aligned} \text{Ob}(\mathbf{C}) &= \prod_I \text{Ob}(\mathbf{C}_i) & \mathbf{C}((X_i)_I, (Y_i)_I) &= \prod_I \mathbf{C}_i(X_i, Y_i), \\ \mu((f_i)_I, (g_i)_I) &= (\mu(f_i, g_i))_I, & \text{id}_{(X_i)_I} &= (\text{id}_{X_i})_I. \end{aligned}$$

The product of a finite family of categories is often written $\mathbf{C} \times \dots \times \mathbf{E}$.

A functor F on a product category is called a functor of several variables; a functor of two variables is often called a bifunctor.

Thus, a functor on a category of the form $\mathbf{C} \times \mathbf{D}^{\text{op}}$ may be described as a ‘‘bifunctor covariant in a \mathbf{C} -valued variable and contravariant in a \mathbf{D} -valued variable’’. Note that if we tried to express contravariance by putting ‘‘ op ’’ onto the *codomains* instead of the domains of functors, we would not be able to express this mixed type of functor; hence the preference for putting op on domains.

A product category $\prod_{i \in I} \mathbf{C}_i$ has a *projection functor* onto each of the categories \mathbf{C}_i ($i \in I$), taking each object and each morphism to its i th component, and as we might expect from our experience with products of other sorts of mathematical objects, this is characterizable by a universal property:

Theorem 6.6.5. Let $(\mathbf{C}_i)_{i \in I}$ be a family of categories, $\mathbf{C} = \prod \mathbf{C}_i$ their product, and $P_i: \mathbf{C} \rightarrow \mathbf{C}_i$ the projection functors. Then for every category \mathbf{D} and family of functors $F_i: \mathbf{D} \rightarrow \mathbf{C}_i$, there exists a unique functor $F: \mathbf{D} \rightarrow \mathbf{C}$ such that for each $i \in I$, $F_i = P_i F$. \square

Exercise 6.6:9. Prove the above theorem.

Exercise 6.6:10. Show that a family of categories also has a *coproduct*. (First state the universal property desired.)

Let us note that the two sorts of hom-functors, h_X and h^Y , are in fact pieces of a single bifunctor. In the definition of this functor below, we use ‘‘ \tilde{X} ’’-notation for objects and morphisms in opposite categories, though in presentations elsewhere, you are likely to see no distinctions made.

Definition 6.6.6. The bivariant hom-functor of a category \mathbf{C} means the functor

$$h: \mathbf{C}^{\text{op}} \times \mathbf{C} \rightarrow \mathbf{Set}$$

which is defined on objects by

$$h(\tilde{X}, Y) = \mathbf{C}(X, Y) \quad (X, Y \in \text{Ob}(\mathbf{C})),$$

while for a morphism $(\tilde{p}, q) \in \mathbf{C}^{\text{op}}(\tilde{X}, \tilde{W}) \times \mathbf{C}(Y, Z)$ (formed from morphisms $p \in \mathbf{C}(W, X)$, $q \in \mathbf{C}(Y, Z)$) we define $h(\tilde{p}, q)$ to carry $a \in \mathbf{C}(X, Y)$ to $qap \in \mathbf{C}(W, Z)$.

Thus, each covariant hom-functor $h_X: \mathbf{C} \rightarrow \mathbf{Set}$ can be described as taking objects Y to the objects $h(\tilde{X}, Y)$, and morphisms q to the morphisms $h(\text{id}_{\tilde{X}}, q)$, and the contravariant hom-functors $h^Y: \mathbf{C}^{\text{op}} \rightarrow \mathbf{Set}$ are similarly obtained by putting Y and id_Y in the right-hand slot of

the bifunctor h .

Exercise 6.6:11. Extend further the ideas of Exercises 6.5:5, 6.6:6 and 6.6:7, by investigating functions in two nonnegative-integer-valued variables induced by bifunctors $\mathbf{FSet} \times \mathbf{FSet} \rightarrow \mathbf{FSet}$, $\mathbf{FSet}^{\text{op}} \times \mathbf{FSet} \rightarrow \mathbf{FSet}$, etc..

6.7. Category-theoretic versions of some common mathematical notions: properties of morphisms. We have mentioned that in an abstract category, one cannot speak of “elements” of an object, hence one cannot meaningfully ask whether a given morphism is one-to-one or onto. However, we have occasionally spoken of two objects of a category \mathbf{C} being “isomorphic”. What we meant was, I hope, clear: An *isomorphism* between X and Y means an element $f \in \mathbf{C}(X, Y)$ for which there exists a 2-sided inverse, that is, a morphism $g \in \mathbf{C}(Y, X)$ such that $fg = \text{id}_Y$, $gf = \text{id}_X$. It is clear that in virtually any naturally occurring category, the invertible morphisms are the things one wants to think of as the isomorphisms. (However, for some mathematical objects other words are traditionally used: In set theory the term is *bijection*, an invertible morphism in \mathbf{Top} is called a *homeomorphism*, and differential geometers call their invertible maps *diffeomorphisms*.) If X and Y are isomorphic, we will as usual write $X \cong Y$. An isomorphism of an object X with itself is called an *automorphism* of X ; these together comprise the *automorphism group* of X .

Exercise 6.7:1. Let \mathbf{C} be a category.

- (i) Show that if a morphism $f \in \mathbf{C}(X, Y)$ has both a right inverse g and a left inverse g' , then these are equal. (Hence if h and h' are both two-sided inverses of f , then $h = h'$.)
- (ii) Show that the relation $X \cong Y$ is an equivalence relation on $\text{Ob}(\mathbf{C})$.
- (iii) Show that isomorphic objects in a category have isomorphic automorphism groups.

Our aim in this and the next section will be to look at various other concepts occurring in “concrete mathematics” and ask, in each case, whether we can define a concept for abstract categories which will yield the given concept in *many* concrete cases. We cannot expect that there will always be as perfect a fit as there was for the concept of isomorphism! But lack of perfect fit with existing concepts will not necessarily detract from the usefulness of the concepts we find.

Let us start with the concepts of “one-to-one map” and “onto map”. The next exercise shows that no condition can give a perfect fit in these cases.

Exercise 6.7:2. Show that a category \mathbf{C} can have concretizations $T, U, V, W: \mathbf{C} \rightarrow \mathbf{Set}$ such that for a particular morphism f in \mathbf{C} , $T(f)$ is one-to-one and onto, $U(f)$ is one-to-one but not onto, $V(f)$ is onto but not one-to-one, and $W(f)$ is neither one-to-one nor onto. (Suggestion: Take $\mathbf{C} = S_{\text{cat}}$, where S is the free monoid on one generator, or $\mathbf{C} = 2_{\text{cat}}$, where 2 is the 2-element totally ordered set.)

Nevertheless, there is a category-theoretic property which in the vast majority of naturally occurring concrete categories does correspond to one-one-ness.

Definition 6.7.1. A morphism $f: X \rightarrow Y$ in a category \mathbf{C} is called a *monomorphism* if for all $W \in \text{Ob}(\mathbf{C})$ and all pairs of morphisms $g, h \in \mathbf{C}(W, X)$, one has $fg = fh \Rightarrow g = h$; equivalently, if every covariant hom-functor $h_W: \mathbf{C} \rightarrow \mathbf{Set}$ ($W \in \text{Ob}(\mathbf{C})$) carries f to a one-to-one set map.

Exercise 6.7:3. (i) Show that if (\mathbf{C}, U) is a concrete category (i.e., \mathbf{C} is a category and $U: \mathbf{C} \rightarrow \mathbf{Set}$ a faithful functor) and f is a morphism in \mathbf{C} such that $U(f)$ is one-to-one, then f is a monomorphism in \mathbf{C} .

(ii) Show conversely that if \mathbf{C} is a *small* category and f a monomorphism in \mathbf{C} , then there exists a faithful functor $U: \mathbf{C} \rightarrow \mathbf{Set}$ such that $U(f)$ is one-to-one.

Exercise 6.7:4. Show that in the categories **Set**, **Group**, **Monoid**, **Ring**¹, **POSet** and **Lattice**, a morphism is one-to-one if and only if it is a monomorphism. (Suggestion: look for one method that works in all six cases.) If you are familiar with the basic definitions of general topology, also verify this for **Top**.

Naturally occurring concrete categories where monomorphisms are not the one-to-one maps are rare, but here is an example:

Exercise 6.7:5. A group G is called *divisible* if for every $x \in |G|$ and every positive integer n , there exists $y \in |G|$ such that $x = y^n$.

(i) Show that in the category of divisible groups (a full subcategory of **Group**), the quotient map $\mathbb{Q} \rightarrow \mathbb{Q}/\mathbb{Z}$ (where \mathbb{Q} is the additive group of rational numbers and \mathbb{Z} the subgroup of integers) is a monomorphism, though it is not a one-to-one map.

(ii) Can you characterize group-theoretically the homomorphisms that are monomorphisms in the category of divisible abelian groups? Of all divisible groups?

(iii) Can you find a category-theoretic property equivalent in either of these categories to being one-to-one?

If you are familiar with topological group theory, you may in the above questions consider the category of connected Lie groups and the quotient map $\mathbb{R} \rightarrow \mathbb{R}/\mathbb{Z}$, instead of or in addition to divisible groups and $\mathbb{Q} \rightarrow \mathbb{Q}/\mathbb{Z}$.

It is natural to dualize the concept of monomorphism.

Definition 6.7.2. A morphism $f: X \rightarrow Y$ in a category \mathbf{C} is called an *epimorphism* if for all $Z \in \text{Ob}(\mathbf{C})$ and all pairs of morphisms $g, h \in \mathbf{C}(Y, Z)$ one has $gf = hf \Rightarrow g = h$; equivalently, if all the contravariant hom-functors $h^Z: \mathbf{C} \rightarrow \mathbf{Set}$ ($Z \in \text{Ob}(\mathbf{C})$) carry f to one-to-one set maps; equivalently, if in \mathbf{C}^{op} the morphism \tilde{f} is a monomorphism.

This concept coincides with that of a surjective map in many naturally occurring concrete categories, but in about equally many it does not:

Exercise 6.7:6. (i) Show that if (\mathbf{C}, U) is a concrete category, and f a morphism in \mathbf{C} such that $U(f)$ is surjective, then f is an epimorphism in \mathbf{C} .

(ii) Show that in the categories **Set** and **Ab**, the epimorphisms are precisely the surjective morphisms.

(iii) Show that in the category **Monoid**, the inclusion of the free monoid on one generator in the free group on one generator is an epimorphism, though not surjective with respect to the underlying-set concretization. (Hint: uniqueness of inverses.) Show similarly that in **Ring**¹, the inclusion of any integral domain in its field of fractions is an epimorphism.

(iv) If you are familiar with elementary point-set topology, show that in the category **HausTop** of Hausdorff topological spaces, the epimorphisms are precisely the continuous maps with *dense* image.

Exercise 6.7:7. (i) Determine the epimorphisms in **Group**.

(ii) Show the relation between this problem and Exercise 3.10:9.

(iii) Does the method you used in (i) also yield a description of the epimorphisms in the category of *finite* groups? If not, can you nevertheless determine these?

Exercise 6.7:8. (i) Show that for an object A of \mathbf{Ring}^1 (or if you prefer, $\mathbf{CommRing}^1$), the following conditions are equivalent: (a) The unique morphism $\mathbb{Z} \rightarrow A$ is an epimorphism. (b) For each object R , there is at most one morphism $A \rightarrow R$ in \mathbf{C} .

(ii) Investigate the class of rings A with the above property. (Cf. Exercise 3.12:7, and last sentence of Exercise 6.7:6(iii).)

Exercise 6.7:9. (i) Show that if R is a commutative ring, and $f: R \rightarrow S$ is an epimorphism in \mathbf{Ring}^1 , then S is also commutative.

(Hint: Given a ring A , construct a ring A' of formal sums $a + b\varepsilon$ ($a, b \in A$) with multiplication given by $(a + b\varepsilon)(c + d\varepsilon) = ac + (ad + bc)\varepsilon$. For fixed $r \in A$, on what elements of A do the two homomorphisms $A \rightarrow A'$ given by $x \mapsto x$ and $x \mapsto (1+r\varepsilon)^{-1}x(1+r\varepsilon)$ agree?)

(ii) Show that if $f: R \rightarrow S$ is an epimorphism in $\mathbf{CommRing}^1$, then it is also an epimorphism in \mathbf{Ring}^1 .

(Hint: Given homomorphisms $g, h: S \rightarrow T$ agreeing on R , reduce to the situation where the image of R in T is in the center. Then look at the ring of endomorphisms of the additive group of T generated by left multiplications by elements of $g(S)$ and right multiplications by elements of $h(S)$.)

(iii) Prove the converse of (ii), i.e., that if a homomorphism of commutative rings is an epimorphism in \mathbf{Ring}^1 , then it is also an epimorphism in $\mathbf{CommRing}^1$. In fact, show that this is an instance of a general property of epimorphisms in a category and a subcategory.

Unlike the result of (iii), the results of (i) and (ii) are rather exceptional, as indicated by

(iv) Show that for a commutative ring k , the inclusion of the ring of upper triangular 2×2 matrices over k (matrices (a_{ij}) such that $a_{21} = 0$) in the ring of all 2×2 matrices over k is an epimorphism in \mathbf{Ring}^1 . Show, however, that the identity $(xy - yx)^2 = 0$ holds in the former ring but not the latter.

Thus, although the result of (i) can be formulated as saying “If $f: R \rightarrow S$ is an epimorphism in \mathbf{Ring}^1 , and R satisfies the identity $xy - yx = 0$, then so does S ”, the corresponding statement with $xy - yx$ replaced by $(xy - yx)^2$ is false.

(v) Similarly, give an example showing that the analog of (ii) is not true for the general case of a category and a full subcategory.

(vi) Does the analog of (i) and/or (ii) hold for the category \mathbf{Monoid} and its subcategory $\mathbf{AbMonoid}$?

Though as some of the above exercises show, the property of being an epimorphism is not a reliable equivalent of surjectivity, we see that it is an interesting concept in its own right. In concrete categories, the statement that $f: A \rightarrow B$ is an epimorphism means intuitively that the image $f(A)$ “controls” all of B , in terms of behavior under morphisms.

There is an unfortunate tendency for some categorical enthusiasts to consider “epimorphism” to be the “category-theoretically correct” translation of “surjective map”, even in cases when the concepts do not agree. For instance, a standard definition in module theory calls a module P *projective* if for every surjective module homomorphism $f: M \rightarrow N$, every homomorphism $P \rightarrow N$ factors through f . (If you haven’t seen this concept, draw a diagram, and verify that every *free* module is projective.) I have heard it claimed that one should therefore define an object P of a general category \mathbf{C} to be projective if and only if for every *epimorphism* $f: M \rightarrow N$ of \mathbf{C} , every morphism $P \rightarrow N$ factors through f . This property is certainly of interest, but there is no reason to consider it to the exclusion of others. In particular, if \mathbf{C} is a category having some natural concretization functor $U: \mathbf{C} \rightarrow \mathbf{Set}$, there is no reason to reject the concept of projective object defined in terms of factorization through “surjective” maps, i.e., maps $f: M \rightarrow N$ such that $U(f)$ is surjective. The fact that a property can be defined purely category-theoretically does not

make it automatically superior to another property.

(The right context for developing a theory of “projective objects” is probably that of a category \mathbf{C} given with a subfamily of morphisms S , which we wish to put in the role of surjections. To make things behave nicely, one will presumably want to put certain restrictions on S ; for instance that it be *contained* in the class of epimorphisms, as the surjective maps in concrete categories always are by Exercise 6.7:6(i); probably also that it contain all invertible morphisms, and be closed under composition. We would then say that an object P is “projective with respect to the class S ” if for every morphism $f: M \rightarrow N$ belonging to S , every morphism $P \rightarrow N$ factors through f . Such an approach is taken in [87], where a large number of properties are defined relative to a *pair* of classes of morphisms, one in the role of the surjections and the other in the role of the injections.)

The use of the words “monomorphism” and “epimorphism” is itself unsettled. In the days before category theory, the words were introduced by Bourbaki with the meanings “injective (i.e., one-to-one) homomorphism” and “surjective (i.e., onto) homomorphism”. The early category-theorists brazenly gave these words the abstract category-theoretic meanings we have been discussing. This, of course, made the terms ambiguous in situations where the category-theoretic definition did not agree with the old meaning. Mac Lane [17] tries to remedy the situation by restoring “monomorphism” and “epimorphism” to their old meanings (applicable in concrete categories) and calling the general category-theoretic concepts that we have been discussing “monic” and “epic” morphisms, or “monos” and “epis” for short. However, the category-theoretic meanings are already well-established in many areas; e.g., there have been many published papers dealing with epimorphisms in categories of rings. (A concept which includes the construction of the field of fractions of a commutative domain is bound to be of interest!) My feeling is that the words “epimorphism” and “epic morphism” sound too similar to usefully carry Mac Lane’s distinction; and that we should now stick with the category-theoretic meanings of “epimorphism” and “monomorphism”. The phrases “surjective (or onto) homomorphism” and “injective (or one-to-one) homomorphism” give us more than enough ways of referring to the concrete concepts.

In any case, when you see these words used by other authors, you should make sure which meaning they are giving them.

Exercise 6.7:10. Suppose $f \in \mathbf{C}(Y, Z)$, $g \in \mathbf{C}(X, Y)$. Investigate implications holding among the conditions “ f is a monomorphism”, “ g is a monomorphism”, “ fg is a monomorphism” “ f is an epimorphism”, “ g is an epimorphism” and “ fg is an epimorphism”.

A full answer would be an exact determination of which among the 64 possible combinations of truth-values for these 6 statements can hold for a pair of morphisms, and which cannot! As a partial answer, you might determine which of the 8 possible combinations of truth-values of the first 3 conditions can hold. Then see whether duality allows you to deduce which combinations of the last 3 can hold, and whether, by examining when morphisms in a *product* of categories are monomorphisms or epimorphisms, you can use the results you have found to get a complete or nearly complete answer to the full 64-case question.

Exercise 6.7:11. Although in most natural categories of mathematical objects the two obvious questions about a morphism are whether it is one-to-one and whether it is onto, in the category **RelSet** we can ask additional questions such as whether a given relation is a *function*.

- (i) Can you find a general condition on a morphism in an arbitrary category, which for a morphisms $f: X \rightarrow Y$ in **RelSet** is equivalent to being a set-theoretic function $X \rightarrow Y$?
- (ii) Examine other properties of relations, and whether they can be characterized by category-theoretic properties in **RelSet**. For instance, which members of **RelSet**(X, X) represent partial

orderings on X ? Given $f, g \in \mathbf{RelSet}(X, Y)$, how can one determine whether $f \subseteq g$ as relations? Can one construct from the category-structure of \mathbf{RelSet} the contravariant functor $R: \mathbf{RelSet}^{\text{op}} \rightarrow \mathbf{RelSet}$ taking each relation $f \in \mathbf{RelSet}(X, Y)$ to the opposite relation, $R(f) \in \mathbf{RelSet}(Y, X)$?

Because of the way we used duality in getting from the concept of monomorphism to that of epimorphism, both of them refer to *one-one-ness* of the images of a morphism under certain hom-functors. Let us look at the conditions that these same images be *onto*:

Exercise 6.7:12. (i) Given $f \in \mathbf{C}(X, Y)$, show that the following conditions are equivalent:

- (a) For all $Z \in \text{Ob}(\mathbf{C})$, $h_Z(f)$ is surjective.
- (b) f is right invertible; i.e., there exists $g \in \mathbf{C}(Y, X)$ such that $fg = \text{id}_Y$.
- (c) For every covariant functor $F: \mathbf{C} \rightarrow \mathbf{Set}$, $F(f)$ is surjective.
- (d) For every contravariant functor $F: \mathbf{C} \rightarrow \mathbf{Set}$, $F(f)$ is injective.
- (e) For every category \mathbf{D} and covariant functor $F: \mathbf{C} \rightarrow \mathbf{D}$, $F(f)$ is an epimorphism.
- (f) For every category \mathbf{D} and contravariant functor $F: \mathbf{C} \rightarrow \mathbf{D}$, $F(f)$ is a monomorphism.

(For partial credit, simply establish the equivalence of (a) and (b). Hint: $\text{id}_Y \in h_Y(Y)$.)

(ii) State the result which follows from the result of (i) by duality, indicating briefly how one deduces this dual result from that of (i).

Let us look at what condition (b) of the above exercise means in familiar categories; in other words, what it means to have two morphisms satisfying a one-sided inverse relation,

$$(6.7.3) \quad fg = \text{id}_Y \quad (f \in \mathbf{C}(X, Y), g \in \mathbf{C}(Y, X)).$$

Let us first take $\mathbf{C} = \mathbf{Set}$. Then we see that if (6.7.3) holds, g must be one-to-one (if two elements of Y fell together under g , there would be no way for f to “separate” them); so let us think of g as embedding a copy of Y in X . The map f sends X to Y so as to take each element $g(y)$ back to y , while acting in an unspecified way on elements of X that are not in the image of g . Thus the composite $gf \in \mathbf{C}(X, X)$ leaves elements of the image of g fixed, and carries all elements not in that image into that image; i.e., it “retracts” X onto the embedded copy of Y . Hence in an arbitrary category, a pair of morphisms satisfying (6.7.3) is called a *retraction* of the object X onto the object Y . In this situation Y is said to be a *retract* of X (via the morphisms f and g).

Exercise 6.7:13. (i) Show that a morphism in \mathbf{Set} is left invertible if and only if it is one-to-one, with the exception of certain cases involving \emptyset (which you should show are indeed exceptions) and right invertible if and only if it is onto (without exceptions).

(ii) Show that X is a retract of Y in the category \mathbf{Ab} of abelian groups (or more generally, the category $R\text{-Mod}$ of left R -modules) if and only if X is isomorphic to a direct summand in Y .

(iii) Give examples of a morphism in \mathbf{Ab} that is surjective, but not right invertible, and a morphism that is one-to-one, but not left invertible.

(iv) Characterize retractions in \mathbf{Group} in terms of familiar group-theoretic constructions. Do they all arise from direct-product decompositions, as in \mathbf{Ab} ?

Combining part (i) of the above exercise with Exercises 6.7:3(i) and 6.7:6(i), we see that in a concrete category, one has

left invertible \Rightarrow one-to-one \Rightarrow monomorphism,
 right invertible \Rightarrow onto \Rightarrow epimorphism.

On the other hand, part (iii) of the above exercise and similar examples given in earlier exercises show that none of these implications are reversible.

Exercise 6.7:14. Give an example of a morphism in some category which is both an epimorphism and a monomorphism, but not an isomorphism. Investigate what combinations of the properties “epimorphism”, “monomorphism”, “left invertible” and “right invertible” force a morphism to be an isomorphism.

Warning in connection with the above discussion and exercises: The meanings of the terms “left” and “right” invertible become reversed when category-theorists – or other mathematicians – compose their maps in the opposite sense to the one we are using!

We have noted that in the situation of (6.7.3) the composite $e = gf$ is an idempotent endomorphism of the object X , whose image, in concrete situations, is a copy of the retract Y . The next exercise establishes two category-theoretic versions of the idea that this idempotent morphism “determines” the structure of the retract Y of X .

Exercise 6.7:15. (i) Let $X, Y, Y' \in \text{Ob}(\mathbf{C})$, and suppose that $f \in \mathbf{C}(X, Y)$, $f' \in \mathbf{C}(X, Y')$ have right inverses g, g' respectively. Show that $gf = g'f' \Rightarrow Y \cong Y'$.

(ii) Let \mathbf{C} be a category, and $e \in \mathbf{C}(X, X)$ be an idempotent morphism: $e^2 = e$. Show that \mathbf{C} may be embedded as a full subcategory in a category \mathbf{D} , *unique up to isomorphism*, with one additional object Y (i.e., with $\text{Ob}(\mathbf{D}) = \text{Ob}(\mathbf{C}) \cup \{Y\}$) and such that there exist morphisms $f \in \mathbf{D}(X, Y)$, $g \in \mathbf{D}(Y, X)$ satisfying

$$fg = \text{id}_Y \text{ (in } \mathbf{D}(Y, Y)), \quad gf = e \text{ (in } \mathbf{D}(X, X) = \mathbf{C}(X, X)).$$

Returning to our search for conditions which correspond to familiar mathematical concepts in many cases, let us ask whether we can define a concept of a *subobject* of an object X in a category \mathbf{C} .

If by this we mean a criterion telling *which* objects of a category such as **Set** or **Group** are actually *contained* in which other objects, the answer is “certainly not”: There can be no way to distinguish an object that *is* a subobject of another from one that is simply *isomorphic* to such a subobject. However, in particular categories of mathematical objects, we may well be able to say when a given morphism is an *embedding*, i.e., corresponds to an isomorphism of its domain object with a subobject of its codomain. For instance, in the categories **Set**, **Group**, **Monoid**, **Ring**¹, **Lattice** and similar categories, the embeddings are the monomorphisms. In these cases, and more generally, whenever we know which morphisms we want to regard as embeddings, we can recover the partially ordered set of subobjects of X as equivalence classes of these morphisms:

Exercise 6.7:16. Let \mathbf{C} be a category, and suppose we are given a subcategory \mathbf{C}_{emb} of \mathbf{C} which includes all the objects of \mathbf{C} , and whose set of morphisms is contained in the set of *monomorphisms* of \mathbf{C} . The morphisms of \mathbf{C}_{emb} are the morphisms of \mathbf{C} that we intend to *think of* as embeddings. (But you may not assume anything about \mathbf{C}_{emb} except the conditions stated above.) For any object X of \mathbf{C} , let \mathbf{Emb}_X denote the category whose objects are pairs (Y, f) , where $Y \in \text{Ob}(\mathbf{C})$ and $f \in \mathbf{C}_{\text{emb}}(Y, X)$, and where a morphism from (Y, f) to (Z, g) means a morphism $a: Y \rightarrow Z$ of \mathbf{C} such that $f = ga$.

- (i) Show that each hom-set $\mathbf{Emb}_X(U, V)$ has at most one element. Deduce that \mathbf{Emb}_X is of the form $\text{Emb}(X)_{\text{cat}}$ for some (possibly large) preorder $\text{Emb}(X)$.
- (ii) Let us call the partially ordered set constructed from the preorder $\text{Emb}(X)$ as in

Proposition 4.2.2 “Sub(X)”. Show that if \mathbf{C} is one of **Set**, **Group**, **Ring** or **Lattice**, and we take \mathbf{C}_{emb} to have for its morphisms all the monomorphisms of \mathbf{C} , then $\text{Sub}(X)$ is isomorphic to the partially ordered set of subsets, subgroups, etc., of X .

(iii) Let X be a set, in general infinite, and S the monoid of set maps of X into itself. Form the category S_{cat} , and take $(S_{\text{cat}})_{\text{emb}}$ to have the monomorphisms of S_{cat} for its morphisms. Calling the one object of S_{cat} “0”, describe the partially ordered set $\text{Sub}(0)$.

The categories of algebraic objects mentioned so far in discussing one-one-ness have had the property that every one-to-one morphism gives an isomorphism of its domain with a subobject of its codomain. An example of a category for which this is not true is **POSet**. For instance if P and Q are finite partially ordered sets having the same underlying set, but the order-relation on Q is stronger than that of P , then the identity map of the underlying set is a one-to-one isotone map from P to Q , but some elements of Q will satisfy order-relations that they don’t satisfy in P , so we cannot regard P as a subobject of Q with the induced ordering. This leads to the questions

Exercise 6.7:17. (i) Suppose the construction of the preceding exercise is applied with \mathbf{C} the category **POSet** and \mathbf{C}_{emb} taken to consist of all the monomorphisms of \mathbf{C} . For $X \in \text{Ob}(\mathbf{C})$, describe the partially ordered set $\text{Sub}(X)$.

(ii) Can you find a category-theoretic property characterizing those morphisms of **POSet** which are “genuine” embeddings, i.e., correspond to isomorphisms of their domain with subsets of their codomain, partially ordered under the induced ordering?

6.8. More categorical versions of common mathematical notions: special objects. I shall start this section with some “trivialities”.

In many of the classes of structures we have dealt with, there were one, or sometimes two objects that one would call the “trivial” objects: the one-element group; the one-element set and also the empty set; the one-element lattice and likewise the empty lattice. The following definition abstracts the common properties of these objects.

Definition 6.8.1. An initial object in a category \mathbf{C} means an object I such that for every $X \in \text{Ob}(\mathbf{C})$, $\mathbf{C}(I, X)$ has exactly one element.

A terminal object in a category \mathbf{C} means an object T such that for every $X \in \text{Ob}(\mathbf{C})$, $\mathbf{C}(X, T)$ has exactly one element.

An object that is both initial and terminal is often called a zero object.

Thus, in **Set**, the empty set is the unique initial object, while any one-element set is a terminal object. In **Group**, a one-element group is both initial and terminal, hence is a zero object. The categories **Lattice**, **POSet**, **Top** and **Semigroup** are like **Set** in this respect, while **Top**^{Pt} and **Monoid** are like **Group**. In **Ring**¹, the initial object is \mathbb{Z} , which we would generally not consider “trivial”; the terminal object is the one-element ring with $1 = 0$ (which some people do not call a ring).

A category need not have an initial or terminal object: The category of nonempty sets, or nonempty partially ordered sets, or nonempty lattices, or finite rings, has no initial object; **POSet**_< has no terminal object, nor does the category of nonzero rings (rings in which $1 \neq 0$). If P is the partially ordered set of the integers, then P_{cat} has neither an initial nor a terminal object. Terminal objects are also called “final” objects, and I may sometimes slip and use that word in class.

Lemma 6.8.2. *If I, I' are two initial objects in a category \mathbf{C} , then they are isomorphic, via a unique isomorphism. Similarly, any two terminal objects are isomorphic via a unique isomorphism. \square*

Exercise 6.8:1. Prove Lemma 6.8.2.

Exercise 6.8:2. Consider the following conditions on a category \mathbf{C} :

- (a) \mathbf{C} has a zero object (an object that is both initial and terminal).
 - (b) It is possible to choose in each hom-set $\mathbf{C}(X, Y)$ a morphism $0_{X, Y}$ in such a way that for all $X, Y, Z \in \text{Ob}(\mathbf{C})$ and $f \in \mathbf{C}(X, Y), g \in \mathbf{C}(Y, Z)$ one has $0_{Y, Z}f = 0_{X, Z} = g0_{X, Y}$.
 - (c) It is possible to choose in each hom-set $\mathbf{C}(X, Y)$ a morphism $0_{X, Y}$ such that for all $X, Y, Z \in \text{Ob}(\mathbf{C})$ one has $0_{Y, Z}0_{X, Y} = 0_{X, Z}$.
 - (d) For all $X, Y \in \text{Ob}(\mathbf{C}), \mathbf{C}(X, Y) \neq \emptyset$.
- (i) Show that (a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (d), but that none of these implications is reversible.
 - (ii) Show that if \mathbf{C} has either an initial or a terminal object, then the first and third implications are reversible, but not, in general, the second.
 - (iii) Show that if \mathbf{C} has an initial object *and* a terminal object (as the majority of naturally occurring categories do), then (d) \Rightarrow (a), so that all four conditions are equivalent.

Exercise 6.8:3. If \mathbf{C} is a category with a terminal object T , let \mathbf{C}^{pt} denote the category whose objects are pairs (X, p) , where $X \in \text{Ob}(\mathbf{C}), p \in \mathbf{C}(T, X)$, and where $\mathbf{C}^{\text{pt}}((X, p), (Y, q)) = \{f \in \mathbf{C}(X, Y) \mid fp = q\}$.

- (i) Verify that this defines a category, and that \mathbf{C}^{pt} will have a zero object.
- (ii) Show that if $\mathbf{C} = \mathbf{Top}$, this gives the category we earlier named \mathbf{Top}^{pt} .
- (iii) Show that if \mathbf{C} already had a zero object, then \mathbf{C}^{pt} will be isomorphic to \mathbf{C} .

Exercise 6.8:4. If \mathbf{C} is a category, call an object A of \mathbf{C} *quasi-initial* if it satisfies the condition of Exercise 6.7:8(i)(b). Generalize the result “(a) \Leftrightarrow (b)” of that exercise to a characterization of quasi-initial objects in categories with initial objects.

What about the concept of *free* object? The definition of a free group F on a set X refers to *elements* of groups, hence the generalization should apply to a *concrete* category (\mathbf{C}, U) . You should verify that when $\mathbf{C} = \mathbf{Group}$ and U is the underlying set functor, the following definition reduces to the usual definition of free group.

Definition 6.8.3. *If \mathbf{C} is a category, $U: \mathbf{C} \rightarrow \mathbf{Set}$ a faithful functor, and X a set, then a free object of \mathbf{C} on X with respect to the concretization U will mean a pair (F_X, u) , where $F_X \in \text{Ob}(\mathbf{C}), u \in \mathbf{Set}(X, U(F_X))$, and this pair has the universal property that for any pair (G, v) with $G \in \text{Ob}(\mathbf{C}), v \in \mathbf{Set}(X, U(G))$, there is a unique morphism $h \in \mathbf{C}(F_X, G)$ such that $v = U(h)u$.*

Loosely, we often call the object F_X the free object, and u the associated universal map.

Exercise 6.8:5. Let V denote the functor associating to every group G the set $|G|^2$ of ordered pairs (x, y) of elements of G , and W the functor associating to G the set of “unordered pairs” $\{x, y\}$ of elements of G (where $x = y$ is allowed).

- (i) State how these functors should be defined on morphisms. (I don’t ask you to verify the fairly obvious fact that these descriptions do give morphisms.) Show that they are both faithful.
- (ii) Show that for any set X , there exists a free group with respect to the functor V , and describe this group.
- (iii) Show that there do not in general exist free groups with respect to W .

Exercise 6.8:6. Let $U: \mathbf{Ring}^1 \rightarrow \mathbf{Set}$ be the functor associating to every ring R the set of 2×2 invertible matrices over R . Show that U is faithful. Does there exist for every set X a free ring R_X on X with respect to U ?

The next exercise shows why the property of being a monomorphism characterizes the one-to-one maps in most of the concrete categories we know – or at least, shows that this characterization follows from another property we have noted in these categories.

Exercise 6.8:7. Let (\mathbf{C}, U) be a concrete category. Show that if there exists a free object on a one-element set with respect to U , then a morphism f of \mathbf{C} is a monomorphism if and only if $U(f)$ is one-to-one.

We could go further into the study of free objects, proving, for instance, that they are unique up to isomorphism when they exist, and that when \mathbf{C} has free objects on all sets, the free-object construction gives a functor $\mathbf{Set} \rightarrow \mathbf{C}$. Some of this will be done in Exercise 6.9:8 later in this chapter, but for the most part we shall get such results in the next chapter, as part of a theory embracing wide classes of universal constructions.

Let us turn to another pair of constructions that we have seen in many categories (including \mathbf{Cat} itself), those of *product* and *coproduct*. No concretization or other additional structure is needed to translate these concepts into category-theoretic terms.

Definition 6.8.4. Let \mathbf{C} be a category, I a set, and $(X_i)_{i \in I}$ a family of objects of \mathbf{C} .

A product of this family in \mathbf{C} means a pair $(P, (p_i)_{i \in I})$, where $P \in \text{Ob}(\mathbf{C})$ and for each $i \in I$, $p_i \in \mathbf{C}(P, X_i)$, having the universal property that for any pair $(Y, (y_i)_{i \in I})$ ($Y \in \text{Ob}(\mathbf{C})$, $y_i \in \mathbf{C}(Y, X_i)$) there exists a unique morphism $r \in \mathbf{C}(Y, P)$ such that $y_i = p_i r$ ($i \in I$).

Likewise, a coproduct of the family $(X_i)_{i \in I}$ means a pair $(Q, (q_i)_{i \in I})$, where $Q \in \text{Ob}(\mathbf{C})$ and for each $i \in I$, $q_i \in \mathbf{C}(X_i, Q)$, having the universal property that for any pair $(Y, (y_i)_{i \in I})$ ($Y \in \text{Ob}(\mathbf{C})$, $y_i \in \mathbf{C}(X_i, Y)$) there exists a unique morphism $r \in \mathbf{C}(Q, Y)$ such that $y_i = r q_i$ ($i \in I$).

Loosely, we call P and Q the product and coproduct of the objects X_i , the $p_i: P \rightarrow X_i$ the projection maps, and the $q_i: X_i \rightarrow Q$ the coprojection maps. (The term injection is used by some authors instead of coprojection.)

The category \mathbf{C} is said to have finite products if every finite family of objects of \mathbf{C} has a product in \mathbf{C} , and to have small products (often simply “to have products”) if every family of objects of \mathbf{C} indexed by a small set has a product; and similarly for finite and small coproducts.

Standard notations for product and coproduct objects are $P = \prod_{i \in I} X_i$ and $Q = \coprod_{i \in I} X_i$. For a product of finitely many objects one also writes $X_0 \times \dots \times X_{n-1}$. There is no analogous standard notation for coproducts of finitely many objects; we used “*” as the operation-symbol in Chapter 3, following group-theorists’ notation for “free products”; one sometimes sees $+$ or \oplus , based on module-theoretic notation. In these notes we shall from now on write $X_0 \amalg \dots \amalg X_{n-1}$, which also occurs in the literature.

Observe that a product of the empty family is equivalent to a terminal object, while a coproduct of the empty family is equivalent to an initial object.

Exercise 6.8:8. If P is a partially ordered set, what does it mean for a family of objects of $P_{\mathbf{cat}}$ to have a product? A coproduct?

Exercise 6.8:9. (i) Suppose we are given a family of families of objects in a category \mathbf{C} , $((X_{ij})_{i \in I_j})_{j \in J}$, such that for each j , $\prod_{I_j} X_{ij}$ exists, and such that we can also find a product of these product objects, $P = \prod_J (\prod_{I_j} X_{ij})$. Show that P will be a product of the family $(X_{ij})_{i \in I_j, j \in J}$.

(ii) Deduce that if a category has products of pairs of objects, it has products of all finite nonempty families of objects.

Exercise 6.8:10. (i) Let X be a set (in general infinite) and S the monoid of maps of X into itself. When, if ever, does the category $S_{\mathbf{cat}}$ have products of pairs of objects? (Of course, there is only one ordered pair of objects, and only one object to serve as their product, so the question comes down to whether two morphisms p_1 and p_2 can be found having appropriate properties.)

(ii) Is there, in some sense, a “universal” example of a monoid S such that $S_{\mathbf{cat}}$ has products of pairs of objects?

Exercise 6.8:11. Let k be a field. Show that one can define a category \mathbf{C} whose objects are the k -vector-spaces, and such that for vector spaces U and V , $\mathbf{C}(U, V)$ is the set of equivalence classes of linear maps $U \rightarrow V$ under the equivalence relation that makes $f \sim g$ if and only if the linear map $f - g$ has finite rank. Show that in this category, finite families of objects have products and coproducts, but infinite families in general have neither.

We saw in Exercise 6.7:13(ii) that in \mathbf{Ab} and $R\text{-Mod}$, any retraction of an object arises from a decomposition as a direct sum, which in those categories is both a product and coproduct. The next exercise examines the relation between retractions, products and coproducts in general.

Exercise 6.8:12. (i) Show that if \mathbf{C} is a category with a zero object, then for any objects A and B of \mathbf{C} , if the product $A \times B$ exists, then A can be identified with a retract of this product, and if the coproduct $A \amalg B$ exists, then A can be identified with a retract of this coproduct.

(ii) Can you find a condition more general than the existence of a zero object under which these conclusions hold?

One does not have the converse to either part of (i). Indeed, let A and B be nontrivial objects of \mathbf{Group} , so that by (i) above, A can be identified with a retract of $A \times B$ and also with a retract of $A \amalg B$. Now

(iii) Show that the subgroup $A \subseteq A \amalg B$, though a retract, is not a factor in any *product* decomposition of that group, and that $A \subseteq A \times B$, though a retract, is not a factor in any *coproduct* decomposition of that group.

Some related facts are noted in the next exercise. (Part (iii) thereof requires some group-theoretic expertise, or some ingenuity.)

Exercise 6.8:13. (i) Show that if A is the free group or free abelian group on a generating set X , and Y is a subset of X , then the subgroup of A generated by Y is a retract of A .

(ii) Conversely, show that if A is a free abelian group and B a retract of A , then A has a basis X such that B is the subgroup generated by a subset of X .

(iii) On the other hand, show that if A is the free group on two generators x and y , then A has cyclic subgroups which are retracts, but are not generated by any subset of any free generating set for A . (Suggestion: try the cyclic subgroup generated by x^2y^3 , or by $x^2yx^{-1}y^{-1}$.)

The next exercise shows that when one requires products of *large* families of objects, one’s categories tend to become degenerate.

Exercise 6.8:14. Let \mathbf{C} be a category and α a cardinal such that $\text{Ob}(\mathbf{C})$ and all morphism sets $\mathbf{C}(X, Y)$ have cardinality $\leq \alpha$ (e.g., the cardinality of a universe with respect to which \mathbf{C} is legitimate).

(i) Show that if every family of objects of \mathbf{C} indexed by a set of cardinality $\leq \alpha$ has a product in \mathbf{C} , then \mathbf{C} has the form $P_{\mathbf{cat}}$, where P is a preorder whose associated partially ordered set P/\approx is a complete lattice.

(ii) Deduce that in this case every family of objects of \mathbf{C} (indexed by any set whatsoever) has a product and a coproduct.

It is an easy fallacy to say “since product is a category-theoretic notion, functors must respect products”. Rather

Exercise 6.8:15. Find an example of categories \mathbf{C} and \mathbf{D} having finite products, and a functor $\mathbf{C} \rightarrow \mathbf{D}$ which does not respect such products.

On the other hand:

Exercise 6.8:16. Show that if (\mathbf{C}, U) is a concrete category, and there exists a free object on one generator with respect to U , then U respects all products which exist in \mathbf{C} . (Cf. Exercise 6.8:7.)

Thus, in most of the concrete categories we have been interested in, the underlying set of a product object is the direct product of the underlying sets of the given objects. However, there is a well-known example for which this fails:

Exercise 6.8:17. A *torsion* group (also called a “periodic group”) is a group all of whose elements are of finite order. Let \mathbf{TorAb} be the category of all torsion abelian groups.

(i) Show that a product in \mathbf{Ab} of an infinite family of torsion abelian groups is not in general a torsion group.

(ii) Show, however, that the category \mathbf{TorAb} has small products.

(iii) Deduce that the underlying set functor $\mathbf{TorAb} \rightarrow \mathbf{Set}$ does not respect products.

Exercise 6.8:18. Does the category $\mathbf{TorGroup}$ of all torsion groups have small products?

For future reference, let us make

Definition 6.8.5. Let I be a set (for instance, a natural number or other cardinal), and \mathbf{C} a category having I -fold products. If X is an object of \mathbf{C} , then when the contrary is not stated, X^I will denote the I -fold product of X with itself, which we may call the “ I th power of X ”. Likewise, if F is a functor from another category \mathbf{D} to \mathbf{C} , then when the contrary is not stated, F^I will denote the functor taking each object Y of \mathbf{D} to the object $F(Y)^I$ of \mathbf{C} , and behaving in the obvious way on morphisms.

(Note that if $F: \mathbf{C} \rightarrow \mathbf{C}$ is an endofunctor of some category \mathbf{C} , we might want to write F^n for the n -fold composite of F with itself. In such a case we would have to make an explicit exception to the above convention.)

What about category-theoretic versions of the constructions of *kernel* and *cokernel*?

We saw that these constructions were specific to fairly limited kinds of mathematical objects, such as groups and rings, but that a pair of concepts which embrace them but are much more versatile are those of *equalizer* and *coequalizer*. The latter concepts are abstracted in

Definition 6.8.6. Let \mathbf{C} be a category, $X, Y \in \text{Ob}(\mathbf{C})$, and $f, g \in \mathbf{C}(X, Y)$.

Then an equalizer of f and g means a pair (K, k) , where K is an object, and $k: K \rightarrow X$ a morphism which satisfies $fk = gk$, and is universal for this property, in the sense that for any pair (W, w) with W an object and $w: W \rightarrow X$ a morphism such that $fw = gw$, there exists a unique morphism $h: W \rightarrow K$ such that $w = kh$.

Likewise, a coequalizer of f and g means a pair (C, c) where C is an object, and $c: Y \rightarrow C$ a morphism which satisfies $cf = cg$, and is universal for this property, in the sense that for any pair (Z, z) with Z an object and $z: Y \rightarrow Z$ a morphism such that $zf = zg$, there exists a unique morphism $h: C \rightarrow Z$ such that $z = hc$.

Loosely, K and C are called the equalizer and coequalizer objects, and k, c the equalizer and coequalizer morphisms, or the canonical morphisms associated with the (co)equalizer construction. We say that \mathbf{C} has equalizers (respectively coequalizers) if every pair of morphisms between every pair of objects of \mathbf{C} has an equalizer (coequalizer).

It turns out that in familiar categories, the concept of coequalizer yields a better approximation to that of *surjective map* than does the concept of epimorphism:

Exercise 6.8:19. (i) Show that in each of the categories **Group**, **Ring**¹, **Set**, **Monoid**, a morphism out of an object Y is surjective on underlying sets if and only if it is a coequalizer morphism of some pair of morphisms from an object X into Y .

(ii) Is the same true in **POSet**? In the category of *finite* groups?

(iii) In the categories considered in (i) (and optionally, those considered in (ii)) investigate whether, likewise, the condition of being an *equalizer* is equivalent to one-one-ness.

(iv) Investigate what implications hold in a general category between the conditions of being an epimorphism, being right invertible, and being a coequalizer map.

Exercise 6.8:20. Let $f, g \in \mathbf{Set}(X, Y)$ be morphisms, and (C, c) their coequalizer.

(i) Show that $\text{card}(X) + \text{card}(C) \geq \text{card}(Y)$. If you wish, assume X and Y are finite.

(ii) Can one establish some similar relation between the cardinalities of X , of Y , and of the *equalizer* of f and g in **Set**?

(iii) What can be said of the corresponding questions in **Ab**? In **Group**?

In categories such as **Group**, **Ab** and **Monoid** which have a zero object, concepts of *kernel* and *cokernel* of a morphism $f: X \rightarrow Y$ may also be defined, namely as the equalizer and coequalizer of f with the zero morphism $X \rightarrow Y$ (see Exercise 6.8:2).

We turn next to a pair of constructions which we have not discussed before, but which are related both to products and coproducts and to equalizers and coequalizers.

Definition 6.8.7. Given objects X_1, X_2, X_3 of a category \mathbf{C} , and morphisms $f_1: X_1 \rightarrow X_3$, $f_2: X_2 \rightarrow X_3$ (diagram below), a pullback of the pair of morphisms f_1, f_2 means a 3-tuple (P, p_1, p_2) , where P is an object, and $p_1: P \rightarrow X_1$, $p_2: P \rightarrow X_2$ are morphisms satisfying $f_1 p_1 = f_2 p_2$, and which is universal for this property, in the sense that any 3-tuple (Y, y_1, y_2) , with $y_1: Y \rightarrow X_1$, $y_2: Y \rightarrow X_2$ satisfying $f_1 y_1 = f_2 y_2$, is induced by a unique morphism $h: Y \rightarrow P$.

(6.8.8)

$$\begin{array}{ccccc}
 & & Y & & \\
 & & \swarrow & & \searrow \\
 & & P & \xrightarrow{p_1} & X_1 \\
 & & \downarrow p_2 & & \downarrow f_1 \\
 & & X_2 & \xrightarrow{f_2} & X_3
 \end{array}$$

Dually, for objects X_0, X_1, X_2 and morphisms $g_1: X_0 \rightarrow X_1, g_2: X_0 \rightarrow X_2$, a pushout of g_1 and g_2 means a 3-tuple (Q, q_1, q_2) , where $q_1: X_1 \rightarrow Q, q_2: X_2 \rightarrow Q$ satisfy $q_1 g_1 = q_2 g_2$, and which is universal for this property in the sense shown below:

(6.8.9)

$$\begin{array}{ccc}
 X_0 & \xrightarrow{g_1} & X_1 \\
 \downarrow g_2 & & \downarrow q_1 \\
 X_2 & \xrightarrow{q_2} & Q
 \end{array}$$

\searrow
 \searrow
 \searrow
 Y

As in the case of products and coproducts, the universal morphisms p_1, p_2 from a pullback object P are called its projection morphisms to the X_i , and the universal morphisms q_1, q_2 to a pushout object Q are called its coprojection morphisms.

A commuting square in \mathbf{C} is called a pullback diagram (respectively, a pushout diagram) if the upper left-hand (lower right-hand) object is a pullback (pushout) of the remainder of the diagram. We say that a category \mathbf{C} has pullbacks if every diagram of objects and morphisms X_1, X_2, X_3, f_1, f_2 as in (6.8.8) has a pullback P , and that \mathbf{C} has pushouts if every diagram of objects and morphisms X_0, X_1, X_2, g_1, g_2 as in (6.8.9) has a pushout Q .

The next exercise shows how to construct these creatures.

Exercise 6.8:21. (i) Show that if a category \mathbf{C} has finite products and has equalizers, then it has pullbacks. Namely, for every system of objects and morphisms, X_1, X_2, X_3, f_1, f_2 as in the first part of the above definition, construct a pullback as the equalizer of a certain pair of morphisms $X_1 \times X_2 \rightarrow X_3$.

(ii) State the dual result for pushouts.

To get a picture of pullbacks in **Set**, note that any set map $f: X \rightarrow Y$ can be regarded as a decomposition of the set X into subsets $f^{-1}(y)$, indexed by the elements $y \in Y$. When looking at f this way, one calls X a set *fibred* by Y , and calls $f^{-1}(y)$ the *fiber* of X at $y \in Y$. Now in a pullback situation (6.8.8) in **Set**, we see that from two sets X_1 and X_2 , each fibred by X_3 , we obtain a third set P fibred by X_3 , with maps into the first two. From the preceding exercise one can verify that the fiber of P at each $y \in X_3$ is the direct product of the fibers of X_1 and of X_2 at y . Consequently, pullbacks are sometimes called *fibred products*, whether or not one is working in a concrete category. The next exercise shows that “fibred products” can be regarded as products in an appropriate category of “fibred objects”.

Exercise 6.8:22. Given a category \mathbf{C} and any $Z \in \text{Ob}(\mathbf{C})$, let \mathbf{C}_Z denote the category of “objects of \mathbf{C} fibered by Z ”, that is, the category having for objects all pairs (X, f) where $X \in \text{Ob}(\mathbf{C})$ and $f \in \mathbf{C}(X, Z)$, and having for morphisms $(X, f) \rightarrow (Y, g)$ all members of $\mathbf{C}(X, Y)$ making commuting triangles with the morphisms f and g into Z .

Show that a *pullback* (6.8.8) in \mathbf{C} is equivalent to a *product* of the objects $(X_1, f_1), (X_2, f_2)$ in \mathbf{C}_{X_3} .

The *pushout* Q of a diagram (6.8.9) is also often called suggestively the “coproduct of X_1 and X_2 with *amalgamation* of X_0 ”, especially in concrete situations where the morphisms f_1 and f_2 are embeddings. It also has names specific to particular fields: In topology, the Q of (6.8.9) is the space gotten by “gluing together” the spaces X_1 and X_2 along a common image of X_0 . In commutative ring theory, where the X_0, X_1 and X_2 of (6.8.9) might be denoted K, R and S , the pushout Q is written $R \otimes_K S$, and called the tensor product of R and S over K as K -algebras.

In the spirit of Chapter 3, you might do

Exercise 6.8:23. (i) Show by a generators-and-relations argument that the category **Group** has pushouts.

(ii) Obtain a normal form or other explicit description for the pushout, in the category of groups, of one-to-one group homomorphisms $f_1: G_0 \rightarrow G_1$ and $f_2: G_0 \rightarrow G_2$. Assume for notational convenience that these maps are inclusions, and that the underlying sets of G_1 and G_2 are disjoint except for the common subgroup G_0 .

This is a classical construction, called by group theorists “the free product of G_1 and G_2 with amalgamation of the common subgroup G_0 ”. (If you are already familiar with this construction, and the proof of its normal form by van der Waerden’s trick, skip to the next part.)

(iii) Describe how to reduce the construction of an arbitrary pushout of groups to the case where the given maps f_1 and f_2 are one-to-one, as above.

Exercise 6.8:24. Let \mathbf{C} be a category having pullbacks and pushouts, and let X_1, X_2, X_3, f_1, f_2 be as in (6.8.8). Suppose we form their pullback P , then form the pushout of the system P, X_2, X_3, p_1, p_2 , and so on, going back and forth between pullbacks and pushouts. Will this process ever “stabilize”?

(Suggestion: Given the two objects X_1 and X_2 , consider the set A of all objects W given with morphisms into X_1 and X_2 , and the set B of all objects Y given with morphisms into them from X_1 and X_2 , and let $R \subseteq A \times B$ denote the relation “the four morphisms form a commuting square”. Examine the resulting Galois connection between A and B .)

We note

Lemma 6.8.10. A morphism $f: X \rightarrow Y$ of a category \mathbf{C} is a monomorphism if and only if the diagram

$$\begin{array}{ccc}
 X & \xrightarrow{\text{id}_X} & X \\
 \downarrow \text{id}_X & & \downarrow f \\
 X & \xrightarrow{f} & Y
 \end{array}$$

is a pullback diagram. Similarly f is an epimorphism if and only if

$$\begin{array}{ccc}
 X & \xrightarrow{f} & Y \\
 \downarrow f & & \downarrow \text{id}_Y \\
 Y & \xrightarrow{\text{id}_Y} & Y
 \end{array}$$

is a pushout diagram. \square

Exercise 6.8:25. Prove Lemma 6.8.10.

The category of “objects of \mathbf{C} fibered over Z ” used in Exercise 6.8:22 has a far-reaching generalization:

Definition 6.8.11. Given three categories and two functors, $\mathbf{D} \xrightarrow{S} \mathbf{C} \xleftarrow{T} \mathbf{E}$, we shall denote by $(S \downarrow T)$ the category having for objects all 3-tuples (D, f, E) , where $D \in \text{Ob}(\mathbf{D})$, $E \in \text{Ob}(\mathbf{E})$, and $f \in \mathbf{C}(S(D), T(E))$, and where a morphism $(D, f, E) \rightarrow (D', f', E')$ means a pair consisting of morphisms $d: D \rightarrow D'$, $e: E \rightarrow E'$, such that $S(d)$ and $T(e)$ make a commuting square with f and f' .

This construction is sometimes written (S, T) . We follow Mac Lane [17] in writing it $(S \downarrow T)$, because, as he observes, “the comma is already overworked”. However, the older notation is the source of its name, the *comma category* construction. The most frequently used cases of this construction are those noted in (ii) and (iii) of the next exercise.

Exercise 6.8:26. (i) Verify that Definition 6.8.11 makes sense. Namely, write out the indicated commutativity condition, say how composition should be defined in $(S \downarrow T)$, and verify that the result is a category.

(ii) Given a category \mathbf{C} , suppose we let $\mathbf{D} = \mathbf{C}$, with $S: \mathbf{C} \rightarrow \mathbf{C}$ the identity functor, while we take for \mathbf{E} the trivial category, with only one object, denoted 0 , and its identity morphism. Let $T: \mathbf{E} \rightarrow \mathbf{C}$ be any functor; thus T will be determined by the choice of one object, $T(0)$, which we shall call Z .

Show that the category $(S \downarrow T)$ can then be identified with the category we called \mathbf{C}_Z in Exercise 6.8:22. This is often denoted $(\mathbf{C} \downarrow Z)$.

(iii) For S and T as in (ii) above, also describe the category $(T \downarrow S)$ (often denoted $(Z \downarrow \mathbf{C})$).

(In the symbols for particular comma categories mentioned at the ends of (ii) and (iii) above, note that the object-name “ Z ” is used as an abbreviation for the functor on the trivial category taking its one object to Z , while \mathbf{C} is used as an abbreviation for the identity functor of \mathbf{C} . The latter is an instance of the use, mentioned in §6.3, of the symbol for an object to denote that object’s identity morphism. Though we are not adopting that usage in general, it is convenient in this case, where we know the two slots in the comma category symbol must be filled with names of functors, so there is no danger of confusion.)

If \mathbf{C} is a category with a terminal object T , the construction \mathbf{C}^{Pt} of Exercise 6.8:3 can clearly be described as $(T \downarrow \mathbf{C})$. However, there is a different comma category construction that is also sometimes called the category of “pointed objects” of \mathbf{C} :

Exercise 6.8:27. Suppose (\mathbf{C}, U) is a concrete category having a free object $F(1)$ on the one-element set $1 = \{0\}$. Show that the following categories are isomorphic:

- (i) $(F(1) \downarrow \mathbf{C})$.
- (ii) The category of objects X of \mathbf{C} given with a distinguished element of $U(X)$, and having for morphisms the morphisms of \mathbf{C} that respect distinguished elements.
- (iii) $(1 \downarrow U)$, where “1” denotes the functor from the one-object one-morphism category to **Set** taking the unique object to the one-element set 1.

Since the one-point topological space is both the terminal object T of **Top** and the free object $F(1)$ on one generator in that category (under the concretization by underlying sets), the constructions $(T \downarrow \mathbf{C})$ and $(F(1) \downarrow \mathbf{C})$ agree in this case, leading to the above situation of one concept with two natural but inequivalent generalizations. As mentioned above, each of these constructions is sometimes called the category of “pointed objects of \mathbf{C} ”, though they may be quite different from one another. The term “pointed” in both cases presumably comes from the term “pointed topological space” for an object of **Top**^{Pt}, regarded as a space with a distinguished basepoint.

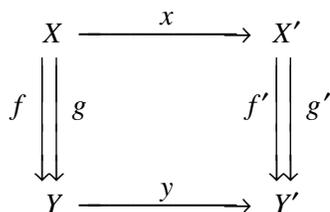
Note that since the terminal object of **Group** is also initial (i.e., is a zero object, as defined in Definition 6.8.1), every object of **Group** admits a unique homomorphism of this terminal object into it. Hence such a homomorphism contains no new information, and **Group**^{Pt} is isomorphic to **Group**. Therefore when an author speaks of the category of “pointed groups” one can guess that he or she does not mean **Group**^{Pt}, but $(F(1) \downarrow \mathbf{Group})$, the category of groups with a distinguished element.

We end this section with a particularly simple example of a category-theoretic translation of a familiar concept. Let G be a group, and recall that a G -set is a set with an *action* of G on it by permutations. More generally, one can consider an action of G by automorphisms on any object X of a category \mathbf{C} ; one defines such an action as a homomorphism f of G into the monoid $\mathbf{C}(X, X)$. Now observe that the pair consisting of such an object X and such a homomorphism $f: G \rightarrow \mathbf{C}(X, X)$ is equivalent to a functor $G_{\mathbf{cat}} \rightarrow \mathbf{C}$; the object X gives the image of the one object of $G_{\mathbf{cat}}$, and f determines the images of the morphisms. Thus, group actions are examples of functors!

6.9. Morphisms of functors (or “natural transformations”). We have seen that various sorts of mathematical structures can be regarded as functors from “diagram” categories to categories of simpler objects: As just noted, G -sets are equivalent to functors from $G_{\mathbf{cat}}$ to **Set**; another example is the type of structure which is the input of the equalizer and coequalizer constructions, consisting of two objects of a category \mathbf{C} and a pair of morphisms from the first object to the second, (X, Y, f, g) . If we call such a 4-tuple a “parallel pair” of morphisms in \mathbf{C} , then as observed in §6.2, parallel pairs correspond to functors from the 2-object diagram category $\cdot \rightrightarrows \cdot$ to \mathbf{C} .

Now if we regard such functors as new sorts of mathematical “objects”, it is natural to ask whether we can define *morphisms* among these objects.

There is a standard concept of a morphism of G -sets – a set map which “respects” the action of G . Is there a similar concept of “morphism of parallel pairs”? Given two parallel pairs $S = (X, Y, f, g)$ and $S' = (X', Y', f', g')$, it seems reasonable to define a morphism $S \rightarrow S'$ to be a pair of morphisms $x \in \mathbf{C}(X, X')$, $y \in \mathbf{C}(Y, Y')$ which respects the structure of parallel pairs, in the sense that $yf = f'x$ and $yg = g'x$:



It is clear how to compose such morphisms, and immediate to verify that this composition makes the class of parallel pairs in \mathbf{C} into a category.

We find that with this definition, equalizers and coequalizers join the ranks of constructions which, though originally thought of only as defined on objects, can also be applied to morphisms. Indeed, if the two parallel pairs of the above diagram each have an equalizer, then it is not hard to check that the morphism (x, y) induces a morphism z of the equalizer objects, and if every parallel pair in \mathbf{C} has an equalizer, then this way of associating to every morphism of parallel pairs a morphism of their equalizer objects makes the equalizer construction a functor. Likewise, if each parallel pair has a coequalizer, the coequalizer construction becomes a functor.

Exercise 6.9:1. Prove the assertions about equalizers in the above paragraph.

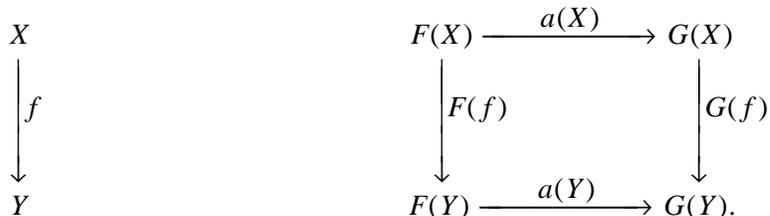
Exactly similar considerations apply to the configurations in a category \mathbf{C} for which we defined the concepts of *pullbacks* and *pushouts*. Such configurations can be regarded as functors from diagram categories $\cdot \rightarrow \downarrow$, respectively $\downarrow \rightarrow \cdot$ into \mathbf{C} , and the set of all configurations of one or the other of these kinds can be made into a category, by letting a morphism from one such configuration to another mean a system of maps between corresponding objects, which respect the given morphisms among these. One can verify that this makes the pullback and pushout constructions, when they exist, into functors on these categories of configurations.

In each of these cases, we have had a diagram category \mathbf{D} and a general category \mathbf{C} , and we have discovered a concept of “morphism” between functors from \mathbf{D} to \mathbf{C} . So, although we have so far regarded functors as the morphisms of \mathbf{Cat} , it seems that there is also a concept of morphisms among functors! We formalize this as

Definition 6.9.1. Let \mathbf{C} and \mathbf{D} be categories and $F, G: \mathbf{D} \rightarrow \mathbf{C}$ functors. Then a morphism of functors $a: F \rightarrow G$ means a family $(a(X))_{X \in \text{Ob}(\mathbf{D})}$ such that for each $X \in \text{Ob}(\mathbf{D})$, $a(X) \in \mathbf{C}(F(X), G(X))$, and for each morphism $f: X \rightarrow Y$ in \mathbf{D} , one has

$$(6.9.2) \quad a(Y)F(f) = G(f)a(X) \text{ in } \mathbf{C}.$$

Pictorially, this means that for each arrow f of \mathbf{D} as at left below, we have commutativity of the square at right:



Given functors $F, G, H: \mathbf{D} \rightarrow \mathbf{C}$ and morphisms $F \xrightarrow{a} G \xrightarrow{b} H$, the composite morphism $ba: F \rightarrow H$ is defined by

$$ba(X) = b(X)a(X) \quad (X \in \text{Ob}(\mathbf{D})).$$

Likewise, the identity morphism id_F of a functor $F: \mathbf{D} \rightarrow \mathbf{C}$ is defined by

$$\text{id}_F(X) = \text{id}_{F(X)} \quad (X \in \text{Ob}(\mathbf{D})).$$

The category whose objects are all the functors from \mathbf{D} to \mathbf{C} , with morphisms, composition, and identity defined as above, will be denoted $\mathbf{C}^{\mathbf{D}}$.

Note that if \mathbf{D} is small, then $\mathbf{C}^{\mathbf{D}}$ will be small or legitimate according as \mathbf{C} is, but that if \mathbf{D} is legitimate, then $\mathbf{C}^{\mathbf{D}}$ will in general be large! (Cf. Exercise 6.4:1(i).) But again, the Axiom of Universes shows us that we may consider these large functor categories as small categories with respect to a larger universe.

We see that if G is a group, the above definition of a morphism between functors $G_{\text{cat}} \rightarrow \mathbf{Set}$ indeed agrees with the concept of a morphism between G -sets, hence the category $G\text{-Set}$ can be identified with $\mathbf{Set}^{(G_{\text{cat}})}$. Since G_{cat} is a small category, $\mathbf{Set}^{(G_{\text{cat}})}$ is a legitimate category.

Let us note some examples, where \mathbf{D} is not a small category, of morphisms between functors we have seen before. Let $F, A: \mathbf{Set} \rightarrow \mathbf{Group}$ be the functors taking a set X to the free group and the free abelian group on X respectively. For every set X there is a homomorphism $a(X): F(X) \rightarrow A(X)$ taking each generator of $F(X)$ to the corresponding generator of $A(X)$. It is easy to see that these form commuting squares with group homomorphisms induced by set maps, hence they constitute a morphism of functors $a: F \rightarrow A$.

Let F again be the free group construction, and let $U: \mathbf{Group} \rightarrow \mathbf{Set}$ be the underlying set functor. Recall that for each $X \in \text{Ob}(\mathbf{Set})$, the universal property of $F(X)$ involves a set map $u(X): X \rightarrow U(F(X))$. It is easy to check that these maps $u(X)$, taken together, give a morphism $u: \text{Id}_{\mathbf{Set}} \rightarrow UF$ of functors $\mathbf{Set} \rightarrow \mathbf{Set}$, where $\text{Id}_{\mathbf{Set}}$ denotes the identity functor of the category \mathbf{Set} .

Exercise 6.9:2. Verify the above claim that u is a morphism of functors.

Statements that two different constructions are “essentially the same” can usually be formulated precisely as saying that they are isomorphic as functors. For instance

Exercise 6.9:3. (i) Let $F: \mathbf{Set} \rightarrow \mathbf{Group}$ denote the free group construction, $A: \mathbf{Set} \rightarrow \mathbf{Group}$ the free abelian group construction, and $C: \mathbf{Group} \rightarrow \mathbf{Group}$ the abelianization construction. Show that $CF \cong A$. (In what functor category?)

(ii) When we gave examples of covariant hom-functors $h_X: \mathbf{C} \rightarrow \mathbf{Set}$ at the end of §6.5, we observed that for $\mathbf{C} = \mathbf{Group}$, the functor $h_{\mathbb{Z}}$ was “essentially” the underlying set functor, and that for $\mathbf{C} = \mathbf{Set}$ and $2 = \{0, 1\} \in \text{Ob}(\mathbf{Set})$, h_2 was “essentially” the construction $X \mapsto X \times X$. Similarly, in §6.6 we noted that, the contravariant hom-functor h^2 on \mathbf{Set} “could be identified with” the contravariant power-set functor. Verify that in each of these cases, we have an *isomorphism* of functors.

(iii) Let $T: \mathbf{Ab} \times \mathbf{Ab} \rightarrow \mathbf{Ab}$ be the tensor product construction, and $R: \mathbf{Ab} \times \mathbf{Ab} \rightarrow \mathbf{Ab} \times \mathbf{Ab}$ the construction taking each pair of abelian groups (A, B) to the pair (B, A) , and acting similarly on morphisms. Show that $T \cong TR$.

(iv) Show that the isomorphisms of Exercise 6.6:5(v) give an isomorphism of functors $\text{Id}_{\mathbf{POSet}} \cong BA$.

Exercise 6.9:4. Give an example of two functors $F, G: \mathbf{D} \rightarrow \mathbf{C}$ such that for every object X of \mathbf{D} , $F(X) \cong G(X)$ in \mathbf{C} , but such that F and G are not isomorphic as functors. In fact, if possible give one example in which \mathbf{C} and \mathbf{D} are both one-object categories, and another in which they are naturally occurring categories of mathematical objects.

A venerable example of an isomorphism of functors arises in considering duality of finite-dimensional vector spaces. We know that a finite-dimensional vector space V , its dual V^* , and its double dual V^{**} are all isomorphic. Now the isomorphism $V \cong V^*$ is not “natural” – these spaces are isomorphic simply because they have the same dimension. But there *is* a natural way to construct an isomorphism $V \cong V^{**}$, by taking each vector v to the operator \bar{v} defined by $\bar{v}(f) = f(v)$ ($f \in V^*$). What this natural construction shows is that for \mathbf{C} the category of finite-dimensional k -vector spaces, the functors $\text{Id}_{\mathbf{C}}$ and $**$ are isomorphic. (One cannot even attempt to construct an isomorphism between $\text{Id}_{\mathbf{C}}$ and $*$, since one functor is covariant and the other contravariant.)

Examples such as this had long been referred to as “natural isomorphisms”, and people had gradually noticed that these and other sorts of “natural” constructions respected maps among objects. When Eilenberg and Mac Lane introduced category theory in [7], they therefore gave the name *natural transformation* to what we are calling a *morphism of functors*. The former term is still widely used, though we shall not use it here. One can also call such an entity a *functorial map*, to emphasize that it is not merely a system of maps between individual objects $F(X)$ and $G(X)$, but that these respect the morphisms $F(f)$ and $G(f)$ that make the constructions F and G functors.

In fact, we used this term “functorial” – deferring explanation – in Exercises 2.3:6 and 2.3:7. What we called there a “functorial group-theoretic operation in n variables” is in our new language a morphism $U^n \rightarrow U$, where U is the underlying-set functor $\mathbf{Group} \rightarrow \mathbf{Set}$, and U^n is (as indicated in Definition 6.8.5) the functor associating to every group G the direct product of n copies of $U(G)$, i.e., the set of n -tuples of elements of $U(G)$. Some cases of those exercises reappear, along with other problems, in the following exercises, which should give you practice thinking about morphisms of functors.

Exercise 6.9:5. In each part below, attempt to describe *all* morphisms among the functors listed, including morphisms from functors to themselves. (I describe functors below in terms of their behavior on objects. The definitions of their behavior on morphisms should be clear. If you are at all in doubt, begin your answer by saying how you think these functors should act on morphisms.)

- (i) The functors Id , A and $B: \mathbf{Set} \rightarrow \mathbf{Set}$ given by $\text{Id}(S) = S$, $A(S) = S \times S$, $B(S) = \{\{x, y\} \mid x, y \in S\}$. (Note that a member of $B(S)$ may have either one or two elements.)
- (ii) The functors U , V and $W: \mathbf{Group} \rightarrow \mathbf{Set}$ given by $U(G) = |G|$, $V(G) = |G| \times |G|$, $W(G) = \{x \in |G| \mid x^2 = e\}$.
- (iii) The underlying set functor $U: \mathbf{FGroup} \rightarrow \mathbf{Set}$, where \mathbf{FGroup} is the category of finite groups.

Exercise 6.9:6. (i) Show that for any category \mathbf{C} , the monoid $\mathbf{C}^{\mathbf{C}}(\text{Id}_{\mathbf{C}}, \text{Id}_{\mathbf{C}})$ of endomorphisms of the identity functor of \mathbf{C} is commutative.

- (ii) Attempt to determine this monoid for the following categories $\mathbf{C}: \mathbf{Set}, \mathbf{Group}, \mathbf{Ab}, \mathbf{FAb}$, the last being the category of finite abelian groups.
- (iii) Do the same for $\mathbf{C} = S_{\mathbf{cat}}$ where S is an arbitrary monoid.
- (iv) Is the endomorphism monoid of a full and faithful functor $F: \mathbf{C} \rightarrow \mathbf{D}$ in general isomorphic to the endomorphism monoid of the full subcategory of \mathbf{D} that is its image? If not, is it at least abelian? If you get such a result, can either “full” or “faithful” be deleted from the hypothesis?

Exercise 6.9:7. (i) Let $F: \mathbf{Set} \rightarrow \mathbf{Set}$ be the functor associating to every set S the set S^ω of all sequences (s_0, s_1, \dots) of elements of S . Determine all morphisms from F to the identity functor of \mathbf{Set} .

(ii) Let $G: \mathbf{FSet} \rightarrow \mathbf{Set}$ be the restriction of the above functor to the category of finite sets; i.e., the functor taking every finite set S to the (generally infinite) set of all sequences of members of S . Determine all morphisms from G to the inclusion functor $\mathbf{FSet} \rightarrow \mathbf{Set}$.

We have mentioned that constructions such as that of free groups, product objects, etc., could be made into functors by using the universal properties to get the required morphisms between the constructed objects. Since then, we have talked about *the* free group functor, *the* product functor on a category, etc.. Part (ii) of the next exercise justifies this use of the definite article.

Exercise 6.9:8. (i) Let (\mathbf{C}, U) be a concrete category having free objects, and let Φ be a function associating to every $X \in \text{Ob}(\mathbf{Set})$ a free object on X in \mathbf{C} , $\Phi(X) = (F(X), u(X))$. Show that there is a unique way of extending F (the first component of Φ) to a functor (i.e., defining $F(f)$ for each morphism f of \mathbf{Set} in a functorial manner) so that u becomes a morphism of functors $\text{Id}_{\mathbf{C}} \rightarrow UF$.

(ii) Suppose $\Phi: X \mapsto (F(X), u(X))$ and $\Psi: X \mapsto (G(X), v(X))$ are two constructions *each* assigning to every set X a free object in \mathbf{C} with respect to U . Show that the functors F and G obtained from Φ and Ψ as in part (i) above (so that the second components u and v become morphisms of functors) are isomorphic; in fact, that there is a *unique* isomorphism making an appropriate diagram commute.

(iii) Write up the analogs of (i) and (ii) for one other functor associated with a universal construction, e.g., products, equalizers, tensor products of abelian groups, etc.. You may abbreviate steps that parallel the free-object case closely.

Exercise 6.9:9. Consider a category \mathbf{C} having finite products. When we spoke of making the product construction into a functor (in motivating the concept of a functor of two variables), the domain category was to be the set of *pairs* of objects of \mathbf{C} . Clearly we can do the same using I -tuples for any *fixed* finite set I . But what if we look at the product construction as simultaneously applying to I -tuples of objects as I ranges over *all* finite index sets?

To make this question precise, let $\text{Ob}(\mathbf{C})^+$ denote the class of all families $(X_i)_{i \in I}$ such that I is a finite set (varying from family to family) and the X_i are objects of \mathbf{C} . Can you make this the object-set of a category \mathbf{C}^+ in a natural way, so that the product construction becomes a functor $\mathbf{C}^+ \rightarrow \mathbf{C}$? If so, will the same category \mathbf{C}^+ serve as domain for the *coproduct* construction, assuming \mathbf{C} has finite coproducts?

Exercise 6.9:10. (i) Suppose $F, G: \mathbf{C} \rightarrow \mathbf{D}$ are functors, and $a: F \rightarrow G$ a morphism of functors. What is the relation between the conditions: (a) for all $X \in \text{Ob}(\mathbf{C})$, $a(X) \in \mathbf{D}(F(X), G(X))$ is a monomorphism in \mathbf{D} , and (b) $a \in \mathbf{D}^{\mathbf{C}}(F, G)$ is a monomorphism in $\mathbf{D}^{\mathbf{C}}$?

(ii) Suppose $F_1, F_2, P: \mathbf{C} \rightarrow \mathbf{D}$ are functors, and $p_1: P \rightarrow F_1, p_2: P \rightarrow F_2$ are morphisms. What is the relation between the conditions (a) for all $X \in \text{Ob}(\mathbf{C})$, $P(X)$ is a product of $F_1(X)$ and $F_2(X)$ in \mathbf{D} , with projection morphisms $p_1(X)$ and $p_2(X)$, and (b) P is a product of F_1 and F_2 in $\mathbf{D}^{\mathbf{C}}$, with projection morphisms p_1 and p_2 ?

To motivate what comes next, let us consider the following three pairs of constructions: (a) To every group G , we may associate the set of its elements of exponent 2, and also its set of elements of exponent 4; this gives two functors V_2 and V_4 from \mathbf{Group} to \mathbf{Set} such that for every G , $V_2(G) \subseteq V_4(G)$. (b) To every set X we can associate the set $\mathbf{P}(X)$ of its subsets, and also the set $\mathbf{P}_f(X)$ of its finite subsets. If we regard the power-set construction as a covariant functor $P: \mathbf{Set} \rightarrow \mathbf{Set}$, this gives a second covariant functor $P_f: \mathbf{Set} \rightarrow \mathbf{Set}$ such that for all X , $P_f(X) \subseteq P(X)$. (We used the covariant power-set functor here because the inverse image of a finite set under a set map may not be finite, so there is no natural way to make a *contravariant* functor out of \mathbf{P}_f .) (c) If $\text{Inv}: \mathbf{Monoid} \rightarrow \mathbf{Monoid}$ denotes the functor associating to every

monoid its submonoid of invertible elements, then for each monoid S , $\text{Inv}(S)$ is a submonoid of $S = \text{Id}_{\text{Monoid}}(S)$.

These examples suggest that we want a concept of “subfunctor” of a functor. Of course, the examples were based on having the concept of a “subobject” of an object, and as we have observed, there is no unique way to define this in a category. However, if we assume a concept of subobject *given*, we can define the concept of subfunctor relative to it:

Lemma 6.9.3. *Let \mathbf{C} be a category, and \mathbf{C}_{incl} be a subcategory having for objects all the objects of \mathbf{C} , and having for morphisms a subclass of the monomorphisms of \mathbf{C} , called the inclusions, such that there is at most one inclusion morphism between any unordered pair of objects (i.e., such that \mathbf{C}_{incl} is a (possibly large) partially ordered set). For $X_0, X \in \text{Ob}(\mathbf{C})$, let us call X_0 a subobject of X (or when there is a possibility of ambiguity, a “subobject with respect to the distinguished subcategory \mathbf{C}_{incl} ”) if there exists an inclusion morphism $X_0 \rightarrow X$. If X_0 and Y_0 are subobjects of X and Y respectively, and $f \in \mathbf{C}(X, Y)$, let us say f carries X_0 into Y_0 if there exists a (necessarily unique!) morphism $f_0 \in \mathbf{C}(X_0, Y_0)$ making a commuting square with f and the inclusions of X_0 and Y_0 in X and Y .*

Then for \mathbf{C} and \mathbf{C}_{incl} as above, and F any functor from another category \mathbf{D} into \mathbf{C} , the following data are equivalent:

- (a) *A choice for each $X \in \text{Ob}(\mathbf{D})$ of a subobject $F_0(X)$ of $F(X)$ such that for each $f \in \mathbf{D}(X, Y)$, $F(f)$ carries $F_0(X)$ into $F_0(Y)$.*
- (b) *A functor $F_0: \mathbf{D} \rightarrow \mathbf{C}$ such that each $F_0(X)$ is a subobject of $F(X)$, and such that the inclusion maps give a morphism of functors $F_0 \rightarrow F$.*
- (c) *A subobject F_0 of F as objects of $\mathbf{C}^{\mathbf{D}}$ with respect to the subcategory of thereof having for objects all the objects of that category (all functors $\mathbf{D} \rightarrow \mathbf{C}$), and for morphisms those morphisms of functors whose values at all objects of \mathbf{D} are inclusion morphisms (relative to \mathbf{C}_{incl}). We may call such an F_0 a subfunctor of F . \square*

Exercise 6.9:11. Prove the above lemma, including the assertion of unicity noted parenthetically near the end of the first paragraph, and the implicit assertion that the subcategory referred to in (c) has the same properties assumed for \mathbf{C}_{incl} . (Can that subcategory of $\mathbf{C}^{\mathbf{D}}$ be described as $(\mathbf{C}_{\text{incl}})^{\mathbf{D}}$?)

In considering categories \mathbf{C} of familiar algebraic objects, when we speak of subobjects and subfunctors, the distinguished subcategory \mathbf{C}_{incl} will be understood to have for morphisms the “ordinary” inclusions, unless the contrary is stated.

Exercise 6.9:12. Let G be a group.

- (i) Show that if S is a subfunctor of the identity functor of **Group**, then $S(G)$ will be a subgroup of G which is carried into itself by every endomorphism of G . (Group theorists call such a subgroup *fully invariant*.)
- (ii) Is it true, conversely, that if H is any fully invariant subgroup of G , then there exists a subfunctor S of Id_{Group} such that $H = S(G)$?
- (iii) Given a subgroup H of G such that *some* subfunctor S of Id_{Group} exists for which $H = S(G)$, will there exist a *least* S with this property? A *greatest*?
- (iv) Generalize your answers to (i)-(iii), in one way or another.

Exercise 6.9:13. Let k be a field of characteristic 0, and $k\text{-Mod}$ the category of k -vector-spaces. For each positive integer n let $\otimes^n: k\text{-Mod} \rightarrow k\text{-Mod}$ denote the n -fold tensor product functor, $V \mapsto V^{\otimes n} =_{\text{def}} V \otimes \dots \otimes V$ (n factors).

- (i) Determine all subfunctors of the functors \otimes^1 and \otimes^2 .
- (ii) Investigate subfunctors of higher \otimes^n 's.
- (iii) Are the results you obtained in (i) and/or (ii) valid over fields k of arbitrary characteristic?

We have observed that the idea that two constructions of some sort of mathematical object are “equivalent” can often be made precise as a statement that two functors are isomorphic. A different type of statement is that two *sorts* of mathematical object are “equivalent”. In some cases, this can be formalized by giving an *isomorphism* (invertible functor) between the categories of the two sorts of objects. E.g., the category of Boolean rings is isomorphic to the category of Boolean algebras, and **Group** is isomorphic to the category of those monoids all of whose elements are invertible. But there are times when this does not work, because the two sorts of objects differ in certain “irrelevant” structure which makes it impossible, or unnatural, to set up such an isomorphism. For instance, groups with underlying set contained in ω are “essentially” the same as arbitrary countable groups, although there cannot be an isomorphism between these two categories of groups, because one is small while the object-set of the other has the cardinality of the universe in which we are working. Monoids are “essentially the same” as categories with just one object, but the natural construction taking one-object categories to monoids is not one-to-one, because it forgets what element was the one object; and the way we found to go in the other direction (inserting “1” as the object) is likewise not onto. For these purposes, a concept weaker than isomorphism is useful.

Definition 6.9.4. A functor $F: \mathbf{C} \rightarrow \mathbf{D}$ is called an equivalence between the categories \mathbf{C} and \mathbf{D} if there exists a functor $G: \mathbf{D} \rightarrow \mathbf{C}$ such that $GF \cong \text{Id}_{\mathbf{C}}$ and $FG \cong \text{Id}_{\mathbf{D}}$ (isomorphisms of functors). If such an equivalence exists, one says “ \mathbf{C} is equivalent to \mathbf{D} ”, often written $\mathbf{C} \approx \mathbf{D}$.

Lemma 6.9.5. A functor $F: \mathbf{C} \rightarrow \mathbf{D}$ is an equivalence if and only if it is full and faithful, and every object of \mathbf{D} is isomorphic to $F(X)$ for some $X \in \text{Ob}(\mathbf{C})$.

Idea of Proof. “ \Rightarrow ” is straightforward. To show “ \Leftarrow ”, choose for each object Y of \mathbf{D} an object $G(Y)$ of \mathbf{C} and an isomorphism $i(Y): Y \rightarrow FG(Y)$. One finds that there is a unique way to make G a functor so that i becomes an isomorphism $\text{Id}_{\mathbf{D}} \cong FG$, and a straightforward way to construct an isomorphism $\text{Id}_{\mathbf{C}} \cong GF$. \square

Note that it is clear from Definition 6.9.4 that the relation \approx is symmetric and reflexive, but it is not entirely clear whether a composite of equivalences is an equivalence, hence whether \approx is transitive. That condition, however, is easily seen from Lemma 6.9.5. So the relation \approx of equivalence between categories is, as one would hope, an equivalence relation.

Exercise 6.9:14. Give the details of the proof of the above lemma.

If one merely assumes a functor $F: \mathbf{C} \rightarrow \mathbf{D}$ is full and faithful, but not the final condition of the above lemma, then it is not hard to deduce that this is equivalent to saying that it is an equivalence of \mathbf{C} with a full subcategory of \mathbf{D} .

Exercise 6.9:15. Let k be a field and $k\text{-FMod}$ the category of finite-dimensional vector spaces over k . Let \mathbf{Mat}_k denote the category whose objects are the nonnegative integers, and such that a morphism from m to n is an $n \times m$ matrix over k , with composition of morphisms given by matrix multiplication. Show that $\mathbf{Mat}_k \approx k\text{-FMod}$.

Exercise 6.9:16. Let k and $k\text{-FMod}$ be as in the preceding exercise. Show that duality of vector spaces gives a *contravariant equivalence* of $k\text{-FMod}$ with itself, i.e., an equivalence between $k\text{-FMod}^{\text{op}}$ and $k\text{-FMod}$.

Exercise 6.9:17. Let \mathbf{FBool}^1 denote the category of finite Boolean rings, and \mathbf{FSet} the category of finite sets. In \mathbf{FBool}^1 , 2 will denote the 2-element Boolean ring with underlying set $\{0, 1\}$, while in \mathbf{FSet} , 2 will as usual denote the set $\{0, 1\}$.

For each $B \in \text{Ob}(\mathbf{FBool}^1)$, if we define $B^* = \mathbf{FBool}^1(B, 2)$, we get a natural homomorphism $m_B: B \rightarrow \prod_{B^*} 2$ which takes $x \in B$ to $(h(x))_{h \in B^*}$.

(i) Show that m_B is always an isomorphism. In particular, this says that every finite Boolean ring is a finite product of copies of the ring 2 .

(ii) Show with the help of the preceding result that the category \mathbf{FBool}^1 is equivalent to $\mathbf{FSet}^{\text{op}}$.

Exercise 6.9:18. Let R be a ring, n a positive integer, and $M_n(R)$ the ring of $n \times n$ matrices over R . For any left R -module M , let $\text{Col}_n(M)$ denote the set of column vectors of height n of elements of M , and let this be made a left $M_n(R)$ -module in the obvious way. This gives a functor $\text{Col}_n: R\text{-Mod} \rightarrow M_n(R)\text{-Mod}$.

Show that Col_n is an equivalence of categories.

(Pairs of rings such as R and $M_n(R)$ which have equivalent module categories are said to be *Morita equivalent*. Morita equivalence was mentioned in Exercise 6.2:3, in terms of isomorphisms in a peculiar category having rings as objects and bimodules as morphisms. I hope in the future to add to Chapter 9 an introduction to Morita theory, from which we will be able to see why the “invertible bimodule” property and the above condition are equivalent.)

The following definition and lemma reduce the question of whether two categories are *equivalent* to the question of whether two other categories are *isomorphic*.

Definition 6.9.6. If \mathbf{C} is a category, then a *skeleton* of \mathbf{C} means a full subcategory having exactly one representative of each isomorphism class of objects of \mathbf{C} ; i.e., by Lemma 6.9.5, a minimal full subcategory \mathbf{C}_0 such that the inclusion of \mathbf{C}_0 in \mathbf{C} is an equivalence.

The Axiom of Choice allows us to construct a skeleton for every category.

Lemma 6.9.7. Let \mathbf{C} and \mathbf{D} be categories, with skeleta \mathbf{C}_0 and \mathbf{D}_0 . Then \mathbf{C} and \mathbf{D} are equivalent if and only if \mathbf{C}_0 and \mathbf{D}_0 are isomorphic. \square

Exercise 6.9:19. Write out the proof of Lemma 6.9.7.

Lemma 6.9.7 shows that equivalent categories agree in all properties that respect isomorphism of categories and “don’t depend on how many isomorphic copies each object has”; that is, intuitively speaking, in all “genuinely category-theoretic” properties.

Exercise 6.9:20. Show that \mathbf{Set} is not equivalent to \mathbf{Set}^{op} by finding a category-theoretic property possessed by one of these categories but not the other, and proving that equivalent categories must agree with respect to whether this property holds. For additional credit, demonstrate the non-equivalence of a few other pairs of familiar categories, e.g., show that \mathbf{Set} is not equivalent to \mathbf{Group} .

Exercise 6.9:21. Let X be a pathwise connected topological space. Recall that one can define a category $\pi_1(X)$ whose objects are the points of X , and in which a morphism from x to y means a homotopy class of paths from x to y . What does a skeleton of this category look like?

Exercise 6.9:22. Suppose \mathbf{C} and \mathbf{D} are equivalent categories, and \mathbf{C}_0 is a subcategory of \mathbf{C} . Must \mathbf{D} have a subcategory \mathbf{D}_0 equivalent to \mathbf{C}_0 ?

6.10. Properties of functor categories. In the preceding section we defined morphisms of functors, and saw some applications of the resulting category structure of $\mathbf{C}^{\mathbf{D}}$. Let us now set down a few basic properties of such categories.

First, consider any bifunctor

$$F: \mathbf{D} \times \mathbf{E} \rightarrow \mathbf{C},$$

in other words, any object of $\mathbf{C}^{\mathbf{D} \times \mathbf{E}}$. If we fix an object $Y \in \text{Ob}(\mathbf{E})$, it is easy to verify that F induces a functor $F(-, Y): \mathbf{D} \rightarrow \mathbf{C}$, i.e., an object of $\mathbf{C}^{\mathbf{D}}$, sending each object X of \mathbf{D} to $F(X, Y)$ and each morphism f of \mathbf{D} to $F(f, \text{id}_Y)$.

Having made this observation for each *object* of \mathbf{E} , let us now note that for each *morphism* between such objects, $g \in \mathbf{E}(Y, Y')$, the morphisms $F(\text{id}_X, g)$ ($X \in \text{Ob}(\mathbf{D})$) yield a morphism of functors $F(-, g): F(-, Y) \rightarrow F(-, Y')$. Thus our system of objects $F(-, Y)$ ($Y \in \text{Ob}(\mathbf{E})$) of $\mathbf{C}^{\mathbf{D}}$ has become a functor $F': \mathbf{E} \rightarrow \mathbf{C}^{\mathbf{D}}$. That is, from our object F of $\mathbf{C}^{\mathbf{D} \times \mathbf{E}}$ we have gotten an object F' of $(\mathbf{C}^{\mathbf{D}})^{\mathbf{E}}$.

In constructing F' , we have not used the values of F at all the morphisms of $\mathbf{D} \times \mathbf{E}$, but only at morphisms of the forms (id_X, g) and (f, id_Y) ; so we might wonder whether F' embodies all the information contained in F . But in fact, an arbitrary morphism of $\mathbf{D} \times \mathbf{E}$, $(f, g): (X, Y) \rightarrow (X', Y')$, can be written $(f, \text{id}_{Y'}) (\text{id}_X, g)$, so the images of morphisms of those two sorts do indeed determine the images of all morphisms of $\mathbf{D} \times \mathbf{E}$. In fact, we have

Lemma 6.10.1 (Law of exponents for categories). *For any categories \mathbf{C} , \mathbf{D} , \mathbf{E} one has $\mathbf{C}^{\mathbf{D} \times \mathbf{E}} \cong (\mathbf{C}^{\mathbf{D}})^{\mathbf{E}}$, via the construction sketched above. \square*

Exercise 6.10:1. Prove the above lemma. In particular, describe how to map morphisms of $\mathbf{C}^{\mathbf{D} \times \mathbf{E}}$ to morphisms of $(\mathbf{C}^{\mathbf{D}})^{\mathbf{E}}$.

Exercise 6.10:2. Does one have other laws of exponents for functor categories? In particular, is $(\mathbf{C} \times \mathbf{D})^{\mathbf{E}} \cong (\mathbf{C}^{\mathbf{E}}) \times (\mathbf{D}^{\mathbf{E}})$, and is $\mathbf{C}^{\mathbf{D} \amalg \mathbf{E}} \cong (\mathbf{C}^{\mathbf{D}}) \times (\mathbf{C}^{\mathbf{E}})$? (For the meaning of $\mathbf{D} \amalg \mathbf{E}$, cf. Exercise 6.6:10.)

Next, suppose that $G_1, G_2: \mathbf{D} \rightarrow \mathbf{C}$ are functors, and $a: G_1 \rightarrow G_2$ is a morphism between them. If H is a functor from any other category into \mathbf{D} , we can form the composite functors $G_1 H$ and $G_2 H$, and we find, not surprisingly, that the morphism $a: G_1 \rightarrow G_2$ induces a morphism $G_1 H \rightarrow G_2 H$. Likewise, given a functor F out of \mathbf{C} , a induces a morphism $F G_1 \rightarrow F G_2$. These induced morphisms of functors are written $a \circ H: G_1 H \rightarrow G_2 H$ and $F \circ a: F G_1 \rightarrow F G_2$ respectively.

Example: Let a be the canonical morphism from the free group functor F to the free abelian group functor A . If we compose on the right with, say, the functor U taking every lattice to its underlying set,

$$\mathbf{Lattice} \xrightarrow{U} \mathbf{Set} \begin{array}{c} \xrightarrow{F} \\ \downarrow a \\ \xrightarrow{A} \end{array} \mathbf{Group},$$

we get a morphism of functors $a \circ U$ mapping free groups on the underlying sets of lattices L homomorphically to the free abelian groups on the same underlying sets. If instead we compose on the left with the underlying-set functor V out of the category of groups,

$$\mathbf{Set} \begin{array}{c} \xrightarrow{F} \\ \downarrow a \\ \xrightarrow{A} \end{array} \mathbf{Group} \xrightarrow{V} \mathbf{Set},$$

we get a morphism of functors $V \circ a$ mapping the underlying set of the free group on each set X to the underlying set of the free abelian group on X .

We record below the above constructions and note the basic laws that they satisfy. The reader is advised to draw (or visualize) pictures like those above for the various situations described.

Lemma 6.10.2. *Let \mathbf{C} , \mathbf{D} and \mathbf{E} be categories.*

(i) *Given a morphism $a: F_1 \rightarrow F_2$ of functors $\mathbf{D} \rightarrow \mathbf{C}$, and any functor $G: \mathbf{E} \rightarrow \mathbf{D}$, a morphism $a \circ G: F_1 G \rightarrow F_2 G$ is defined by setting $(a \circ G)(X) = a(G(X))$ ($X \in \text{Ob}(\mathbf{E})$).*

(ii) *Given any functor $F: \mathbf{D} \rightarrow \mathbf{C}$, and a morphism $b: G_1 \rightarrow G_2$ of functors $\mathbf{E} \rightarrow \mathbf{D}$, a morphism $F \circ b: F G_1 \rightarrow F G_2$ is defined by setting $(F \circ b)(X) = F(b(X))$ ($X \in \text{Ob}(\mathbf{E})$).*

(iii) *Given morphisms $F_1 \xrightarrow{a_1} F_2 \xrightarrow{a_2} F_3$ of functors $\mathbf{D} \rightarrow \mathbf{C}$, and any functor $G: \mathbf{E} \rightarrow \mathbf{D}$, one has*

$$(a_2 a_1 \circ G) = (a_2 \circ G)(a_1 \circ G).$$

(iv) *Given any functor $F: \mathbf{D} \rightarrow \mathbf{C}$, and morphisms $G_1 \xrightarrow{b_1} G_2 \xrightarrow{b_2} G_3$ of functors $\mathbf{E} \rightarrow \mathbf{D}$, one has*

$$(F \circ b_2 b_1) = (F \circ b_2)(F \circ b_1).$$

(v) *Given both a morphism $a: F_1 \rightarrow F_2$ of functors $\mathbf{D} \rightarrow \mathbf{C}$, and a morphism $b: G_1 \rightarrow G_2$ of functors $\mathbf{E} \rightarrow \mathbf{D}$, one has*

$$(6.10.3) \quad (a \circ G_2)(F_1 \circ b) = (F_2 \circ b)(a \circ G_1)$$

as morphisms $F_1 G_1 \rightarrow F_2 G_2$.

(vi) *Given functors $F: \mathbf{D} \rightarrow \mathbf{C}$ and $G: \mathbf{E} \rightarrow \mathbf{D}$, one has*

$$\text{id}_F \circ G = \text{id}_{FG} = F \circ \text{id}_G.$$

(vii) *Hence, composition of functors $\mathbf{E} \rightarrow \mathbf{D} \rightarrow \mathbf{C}$ may be made a functor $\mathbf{C}^{\mathbf{D}} \times \mathbf{D}^{\mathbf{E}} \rightarrow \mathbf{C}^{\mathbf{E}}$ by taking each morphism $(a, b): (F_1, G_1) \rightarrow (F_2, G_2)$ to the common value of the two sides of (6.10.3). \square*

Exercise 6.10.3. (i) Prove statements (i)-(vi) of the above lemma.

(ii) Show that statement (vii) summarizes all of statements (i)-(vi), except for the descriptions of how $F \circ a$ and $b \circ G$ are defined.

The above lemma shows that the operation of composing functors, which, to begin with, was defined as a set map

$$(6.10.4) \quad \mathbf{Cat}(\mathbf{D}, \mathbf{C}) \times \mathbf{Cat}(\mathbf{E}, \mathbf{D}) \rightarrow \mathbf{Cat}(\mathbf{E}, \mathbf{C}),$$

actually gives a functor

$$(6.10.5) \quad \mathbf{C}^{\mathbf{D}} \times \mathbf{D}^{\mathbf{E}} \rightarrow \mathbf{C}^{\mathbf{E}}.$$

In making \mathbf{Cat} a category, we had to verify that the set map (6.10.4) satisfied the associativity and identity laws; we now ought to check that these laws hold, not merely as equalities of set maps, but as equalities of functors! The case of the identity laws is easy, but as part (ii) of the next exercise shows, is still useful:

Exercise 6.10:4. (i) Given a morphism $a: G_1 \rightarrow G_2$ of functors $G_1, G_2: \mathbf{D} \rightarrow \mathbf{C}$, show that

$$a \circ \text{Id}_{\mathbf{D}} = a = \text{Id}_{\mathbf{C}} \circ a.$$

(ii) Show that the above result, together with Lemma 6.10.2(v), immediately gives the result of Exercise 6.9:6(i).

It is more work to write out the details of

Exercise 6.10:5. For categories \mathbf{B} , \mathbf{C} , \mathbf{D} and \mathbf{E} establish identities (like those of Lemma 6.10.2) showing that the two iterated-composition functors $\mathbf{B}^{\mathbf{C}} \times \mathbf{C}^{\mathbf{D}} \times \mathbf{D}^{\mathbf{E}} \rightarrow \mathbf{B}^{\mathbf{E}}$ are equal as functors.

In doing the above exercises, you may wish to use the notation which represents the common value of the two sides of (6.10.3) as $a \circ b$. Note, however, point (i) of

Exercise 6.10:6. (i) Show that if the above notation is adopted, there are situations where $a \circ b$ and ab are both defined, but are unequal.

(ii) Can you find any important class of cases where they must be equal?

Exercise 6.10:7. Suppose we have an equivalence of categories, given by functors $F: \mathbf{C} \rightarrow \mathbf{D}$, $G: \mathbf{D} \rightarrow \mathbf{C}$ with $\text{Id}_{\mathbf{C}} \cong GF$, $\text{Id}_{\mathbf{D}} \cong FG$. Given a particular isomorphism of functors $i: \text{Id}_{\mathbf{C}} \rightarrow GF$, can one in general choose an isomorphism $j: \text{Id}_{\mathbf{D}} \rightarrow FG$ such that the two isomorphisms of functors, $i \circ G$, $G \circ j: G \rightarrow GFG$ are equal, and likewise the two isomorphisms $j \circ F$, $F \circ i: F \rightarrow FGF$?

How are we to look at a functor category $\mathbf{C}^{\mathbf{D}}$? Should we think of the functors which are its objects as “maps” or as “things”? As a category, is it “like” \mathbf{C} , “like” \mathbf{D} , or like neither?

My general advice is to think of its objects as “things” and its morphisms as “maps”; more precisely, its objects are “things” composed of *systems* of objects of \mathbf{C} linked together by morphisms of \mathbf{C} in a way *parametrized* by \mathbf{D} ; its morphisms are further systems of morphisms of \mathbf{C} uniting parallel structures of this sort. With respect to basic properties, such a functor category usually behaves more like \mathbf{C} than like \mathbf{D} . For example, if \mathbf{C} has finite products, so does $\mathbf{C}^{\mathbf{D}}$: One can construct the product $F \times G$ of two functors $F, G \in \mathbf{C}^{\mathbf{D}}$ “objectwise”, by taking $(F \times G)(X)$ to be the product $F(X) \times G(X)$ for each $X \in \text{Ob}(\mathbf{D})$ (cf. Exercise 6.9:10(ii)). On the other hand, so far as I know, existence of products in \mathbf{D} tells us nothing about $\mathbf{C}^{\mathbf{D}}$.

6.11. Enriched categories (a sketch). A recurring trick in category theory is to characterize some type of mathematical entity as a certain kind of structure in a particular category, such as \mathbf{Set} , analyze what properties of \mathbf{Set} are needed for the concept to make sense, and then create a generalized definition, which is like the original one except that \mathbf{Set} is replaced by a general category having the required properties.

There is in fact an important application of this idea to the concept of *category* itself! We shall sketch this briefly below. We will begin with a few examples to motivate the idea, and then discuss what is involved in the general case, though we shall not give formal definitions.

Recall that a category, as we have defined it, is given by a *set* of objects, and a *set* of morphisms between any two objects, with composition operations given by *set maps*, $\mu: \mathbf{C}(Y, Z) \times \mathbf{C}(X, Y) \rightarrow \mathbf{C}(X, Z)$. But now consider the category **Cat**. Though we still have a *set* of objects, we have seen that for each pair of objects **C**, **D** we can speak of a *category* of morphisms $\mathbf{C}^{\mathbf{D}}$, and composition in **Cat** is given by *bifunctors* $\mu_{\mathbf{CDE}}: \mathbf{E}^{\mathbf{D}} \times \mathbf{D}^{\mathbf{C}} \rightarrow \mathbf{E}^{\mathbf{C}}$.

Likewise, for any ring R , it is well known that the homomorphisms from one R -module to another form an additive group, so the category $R\text{-Mod}$ can be described as having, for each pair of objects, an *abelian group* of morphisms $X \rightarrow Y$. Here composition is given by *bilinear maps* $(R\text{-Mod}(Y, Z), R\text{-Mod}(X, Y)) \rightarrow R\text{-Mod}(X, Z)$ among these abelian groups.

One expresses these facts by saying that **Cat** can be regarded as a **Cat**-based category, or a **Cat**-category for short, and $R\text{-Mod}$ as an **Ab**-category. Similarly, in various situations where one has a natural topological structure on sets of morphisms and the composition maps are bicontinuous, one can say one has a **Top**-based category.

These generalized categories are called *enriched* categories.

Note that when we referred to $R\text{-Mod}$ as being an **Ab**-category, this included the observation that the composition maps μ_{XYZ} are *bilinear*. Thus, they correspond to abelian group homomorphisms (which by abuse of notation we shall denote by the same symbols):

$$\mu_{XYZ}: \mathbf{C}(Y, Z) \otimes \mathbf{C}(X, Y) \rightarrow \mathbf{C}(X, Z).$$

The general definition of enriched category requires that the base category (the category in which the hom-objects are taken to lie; i.e., **Set**, **Cat**, **Ab**, **Top**, etc.) be given with a bifunctor into itself having certain properties, which is used, as above, in describing the composition maps. In the case where the base category is **Ab**, this is the *tensor-product* bifunctor, while in the cases of **Set**, **Cat**, and **Top**, the corresponding role is filled by the *product* bifunctor. See [17, §VII.7] for a few more details, and [85] for a thorough development of the subject.

One should, strictly, distinguish between $R\text{-Mod}$ as an ordinary (i.e., **Set**-based) category and as an **Ab**-category, writing these two entities as, say, $R\text{-Mod}$ and $R\text{-Mod}_{(\mathbf{Ab})}$, and similarly distinguish **Cat** and $\mathbf{Cat}_{(\mathbf{Cat})}$ – just as one ought to distinguish between the set of integers, the additive group of integers, the lattice of integers, the ring of integers, etc.. This notational problem will not concern us, however, since we will not formalize the concept of enriched category in these notes. Outside this section, when we occasionally refer to the special properties of categories such as $R\text{-Mod}$ or **Cat**, we shall not assume any general theory of enriched categories, but simply use in an ad hoc fashion what we know about the extra structure.

We remark that **Ab**-based categories, and more generally, $k\text{-Mod}$ -based categories for k a commutative ring (these are called “ k -linear categories”), are probably more widely used than all the other sorts of enriched categories together. See [9] for a lively development of the subject.

The **Cat**-based category **Cat** contains a vast number of interesting sub-**Cat**-categories. Here is one:

Exercise 6.11:1. Consider the full subcategory of **Cat** whose objects are the categories $G_{\mathbf{cat}}$, for groups G . Characterize in group-theoretic terms the morphisms, and morphisms of morphisms, in this **Cat**-category.

Translate your answer into a description of a **Cat**-category structure one can put on the category **Group**.

The student interested in ring theory might note that the category of Exercise 6.2:3 (with rings as objects, and bimodules ${}_R B_S$ as morphisms) can be made a **Cat**-category, by using bimodule homomorphisms as the morphisms-among-morphisms; moreover, each morphism-category $\mathbf{C}(R, S)$ (for R and S rings) is in fact an **Ab**-category! What this says is that this category is an **AbCat**-category, where **AbCat** is the category of **Ab**-categories. There is an explanation: This category is equivalent to the subcategory of **Cat** whose objects are the **Ab**-categories $R\text{-Mod}$ for rings R , and whose morphisms are the functors ${}_R B_S -: S\text{-Mod} \rightarrow R\text{-Mod}$ induced by bimodules ${}_R B_S$. So this observation is really a special case of the fact that **AbCat** is an **AbCat**-category, just as **Ab** is an **Ab**-category and **Cat** is a **Cat**-category.

We have mentioned (in Exercise 6.3:1 and the preceding discussion) that there is a version of the definition of category which eliminates reference to objects, and thus involves only one kind of element, the morphism (the objects being hidden under the guise of their identity morphisms). If we apply this idea twice to the concept of a **Cat**-category, we likewise get a structure with only one type of element – what we have been calling the morphisms of morphisms – but with two *partial* composition operations on these elements, ab and $a \circ b$ (Exercise 6.10:6). Described in this way, **Cat**-categories have been called “2-categories” [17, p.44]. (The relation between the two types of composition is slightly asymmetric. If one drops the asymmetric condition – that every identity element with respect to the *first* composition is also an identity element with respect to the second – one gets a slightly more general concept, also defined in [17], and called a “double category”.)

Having begun by considering **Cat** as an ordinary, i.e., **Set**-based category, with objects and morphisms (i.e., functors), and then having found that there was an important concept of *morphisms between morphisms* (morphisms of functors), we may ask whether one can define, further, *morphisms between morphisms between morphisms* in this category. The answer is “yes and no”. On the yes side, observe that one can set up a concept of “morphisms between morphisms” in any category **C**! For a morphism in **C** is the same as an object of \mathbf{C}^2 , where **2** denotes the diagram category $\rightarrow\rightarrow$, and we know how to make \mathbf{C}^2 a category. So in particular, given categories **C** and **D** we can define a “morphism of morphisms in $\mathbf{C}^{\mathbf{D}}$ ”, which is thus a “morphism of morphisms of morphisms” in **Cat**.

However, this construction does not constitute a nontrivial enrichment of structure, since the concept of morphism of morphisms we have just described in an arbitrary category **C** is defined in terms of its existing category structure. (Indeed, when applied to **Cat**, it does not give the concept of “morphism of functors”, but that of “commuting square of functors”.) So we come to the “no” side of the answer – so far as I know, the category **Cat** has no enriched structure beyond that of a **Cat**-category.

However, if one turns from the category **Cat** of all **Set**-based categories, to the category **CatCat** of all **Cat**-based categories, one finds that here one has a natural and nontrivial concept of morphisms between morphisms between morphisms – in other words, **CatCat** is a **CatCat**-based category. And this process can be iterated ad infinitum.

But it is time to return from these vertiginous heights to the main stream of our subject.