Category Theory

Notes by Peter Johnstone

Michaelmas 2020

1 Definitions and Examples

Category theory begins with the observation, made by Samuel Eilenberg and Saunders Mac Lane in the early 1940s, that the collection of all mathematical structures of a given type, together with all the appropriate mappings betwen them, is itself an instance of a structure which can be studied mathematically. (This came about through a chance meeting between the two men, when they realized that the calculations Eilenberg was performing in his research in topology were almost identical to those which Mac Lane was doing in algebra, and set out to find the underlying reason for this.)

They arrived at the following formal definition:

Definition 1.1 A *category* $\mathcal{C}$ consists of the following data:

(a) a collection ob $\mathcal{C}$ of *objects* $A, B, C, \ldots$;

(b) a collection mor $\mathcal{C}$ of *morphisms* $f, g, h, \ldots$;

(c) two operations dom, cod assigning to each morphism f its *domain* dom f and its *codmain* cod f, which are objects. We use the arrow notation '$f : A \to B$' as shorthand for 'f is a morphism and dom $f = A$ and cod $f = B$'.

(d) an operation assigning to each object A a morphism $1_A : A \to A$;

(e) and a partial binary *composition* operation $(f, g) \mapsto fg$ on morphisms, subject to the condition that fg is defined iff cod g = dom f, and then we have dom fg = dom g and cod fg = cod f.

These data satisfy the two axioms

(f) $f1_A = f$ and $1_A g = g$ whenever the composites are defined, and

(g) $f(gh) = (fg)h$ whenever fg and gh are defined. (Note that the 'book-keeping' conditions in (e) then ensure that the two triple composites are defined.)

Remarks 1.2 Before giving a list of examples,we need to make a few comments about this definition.

(a) The use of the vague word 'collection' rather than something precise like 'class' or 'set' in (a) and (b) was deliberate; we want the definition to be an elementary one interpretable in any first-order context, since we shall wish to consider categories such as the category of all sets and functions, whose collection of objects is definitely not a set. (So when we say '$\mathcal{C}$ is a category', all we mean is that we have given well-defined meanings to the statements 'A is an object of $\mathcal{C}$' and 'f is a morphism of $\mathcal{C}$', and that these are compatible with the conditions above.) On the other hand, some of the categories we consider will be formalizable within a given model of set theory, in that ob $\mathcal{C}$ and mor $\mathcal{C}$ will be sets; we distinguish such categories by calling them *small*. (But it should be realized that the term 'small category' only has meaning once we have specified which model of set theory we are working in.)

(*b*) The axioms ensure that in any category there is a bijection between the objects and the *identity morphisms*, i.e. those f such that $fg = g$ and $hf = h$ whenever the composites are defined; so we could if we wished formulate the notion of category as a single-sorted theory, in which the only entities are morphisms equipped with a partial binary composition operation and axioms which say that there are 'enough identities'. (This approach was in particular followed by the French school of which the late Charles Ehresmann was a leading exponent.) But in most of the categories we consider the objects are in some sense 'logically prior' to the morphisms (for example, how can one define a group homomorphism without having first defined a group?), so we take the view that objects as well as morphisms deserve to be treated as primitive concepts.

(*c*) There is a convention implicit in 1.1(*e*), namely that fg means 'first do g, then do f'; i.e., if we think of morphisms as functions (which they very often are), we follow the usual convention of writing them on the left of their arguments (i.e. one writes $\sin x$ rather than $x \sin$). Clearly, it is possible to adopt the opposite convention, and some writers (particularly those in computer science) do so. But we shall stick with the 'left-handed' convention in these notes.

Examples 1.3 (*a*) As already hinted, one of our principal examples will be the category **Set** of all sets and all functions between them. (Actually, there is a slight glitch here, in that the usual set-theoretic definition of a function as a set of ordered pairs does not satisfy axiom (*c*): a set-theoretic function f has a well-defined domain (the set of all first members of ordered pairs in f), but its codomain could be any set containing all the second members of such pairs. Formally, therefore, a morphism of **Set** has to be taken to be a pair of the form (f, B) where f is a set-theoretic function and B is the set which we are considering to be its codomain. But this is not something to worry about; it is really a defect of set theory rather than of category theory — if A is a subset of B, a set-theorist thinks of the identity function on A and the inclusion $A \to B$ as identically the same thing, whereas most mathematicians would wish to differentiate between them.)

(*b*) For any type of algebraic structure, there is a corresponding notion of homomorphism, which is 'stable' under composition; thus we have the category **Gp** of all groups and group homomorphisms, the category **Rng** of all rings and ring homomorphisms, the category $\mathbf{Vect}_k$ of vector spaces over a given field k, and so on (in each case with the proviso about specifying codomains of morphisms — and indeed we have to specify domains too, since the set-theoretic view of a homomorphism specifies the underlying set of its domain but not the algebraic structure on it).

(*c*) Similarly, we have the category **Top** of topological spaces and continuous maps, the category **Met** of metric spaces and non-expansive maps (i.e. functions f satisfying $d(f(x), f(y)) \leq d(x, y)$ for all x and y), the category **Mfd** of smooth manifolds and C^∞ maps, and so on.

(*d*) Even if the objects of a category are sets (with structure), the morphisms do not have to be individual functions. For example, many of you will be familiar with the equivalence relation of *homotopy* between continuous maps of spaces, and we can form a new category **Htpy** whose objects are those of **Top**, but whose morphisms are homotopy classes of continuous maps. The conditions required for this, given an equivalence relation $\simeq$ on $\mathrm{mor}\,\mathcal{C}$, are that it should respect domains and codomains (i.e. $f \simeq g$ implies $\mathrm{dom}\, f = \mathrm{dom}\, g$ and $\mathrm{cod}\, f = \mathrm{cod}\, g$) and be compatible with composition (i.e. $f \simeq g$ implies $fh \simeq gh$ and $kf \simeq kg$ when the composites are defined); given such a relation, which we call a *congruence* on $\mathcal{C}$, we may form a quotient category $\mathcal{C}/\simeq$ with the same objects as $\mathcal{C}$, but equivalence classes as morphisms.

(*e*) In the opposite direction, we may allow things more general than functions to be morphisms. The category **Rel** has the same objects as **Set**, but its morphisms $A \to B$ are arbitrary relations, i.e. subsets of $A \times B$; the composite of $R\colon A \to B$ and $S\colon B \to C$ is

$$S \circ R = \{(a, c) \mid (\exists b \in B)((a, b) \in R) \text{ and } ((b, c) \in S)\} .$$

It will be seen that if we regard functions, set-theoretically, as special cases of relations, this agrees with the usual composition of functions; thus we may think of **Set** as a subcategory of **Rel**. Intermediate between the two, we have the category **Part** of sets and partial functions: a partial function $A \rightharpoonup B$ is a relation which satisfies the 'uniqueness' but not the 'existence' condition in the definition of a function; equivalently, it is a function from some subset of A to B.

(f) For any category $\mathcal{C}$, we obtain a new category $\mathcal{C}^{\mathrm{op}}$ (its *dual* or *opposite* category) by taking the same objects and morphisms, interchanging the operations dom and cod, and reversing the order of composition. Thus we obtain a *duality principle* for categories; whenever we have proved a proposition P about categories, we have also proved the proposition P* obtained from P by reversing all the arrows. (One should think of category theory as being like a supermarket whose shelves are piled high with 'buy one, get one free' offers.)

(g) So far, all the categories we have looked at have been large (i.e., they have had proper classes of objects); we should see some small ones. If there are no objects, there can be no morphisms either, so that is not very interesting. But if there is just one object ($*$, say), then all composites must be defined; so (provided we impose the restriction of smallness) what we have is just a *monoid*, i.e. a semigroup with identity element. In particular, any group is a monoid, so we can regard groups as small categories with one object in which every morphism is an isomorphism. (We haven't defined 'isomorphism', but it's obvious what it has to mean.)

(h) More generally, we define a *groupoid* to be a category (with several objects) in which every morphism is an isomorphism. One example will be familiar to those of you who have done a first course in algebraic topology: given a space X and a base-point $x \in X$, we have a notion of *loop* in X based at x, and a composition operation on these loops. This operation is not associative, but it becomes so if we quotient by the relation of homotopy, giving rise to a group $\Pi_1(X, x)$, the *fundamental group* of X based at x. But we can avoid the arbitrary choice of base-point by taking *all* the points of X as objects of a groupoid $\Pi_1(X)$, whose morphisms $x \to y$ are homotopy classes of paths starting at x and ending at y. (If you look at the proof in any topology textbook that the fundamental group is a group, you will see that it never uses the fact that all the paths under consideration begin and end at the same point, so it actually proves that $\Pi_1(X)$ is a groupoid.)

(i) In the other direction, we can restrict the number of morphisms as far as possible. If the only morphisms are identities, we have what is called a *discrete* category; these are not very interesting. Less rigidly, we may suppose that for each pair of objects (A, B) there is at most one morphism $A \to B$; then mor $\mathcal{C}$ becomes a binary relation on ob $\mathcal{C}$, which must be reflexive (since identity morphisms exist) and transitive (since composites exist); in other words, it is what we commonly call a *preorder*. In particular, a partial order is a special case of a preorder; so we may regard a partially ordered set (or poset) as a small preorder whose only isomorphisms are identities. Thus the theory of posets, like that of groups, is subsumed in the theory of categories. (It is often helpful, on encountering a new categorical concept, to ask oneself 'What does this mean in the particular case when the categories involved are posets?')

(j) Finally, an example which most of you have known for a long time, without being aware that it was an example: the category of matrices over a field. Given a field k, the category $\mathbf{Mat}_k$ has natural numbers as objects; for compatibility with the usual definition of matrix multiplication, the morphisms $n \to p$ are $(p \times n)$ matrices with entries from k, and composition is matrix multiplication. Almost certainly, when you first learned how to multiply matrices, you verified that matrix multiplication is associative to the extent that it's defined; when you did that, you were precisely verifying that $\mathbf{Mat}_k$ is a category.

Since the whole philosophy behind category theory is that the objects we study are not more important than the mappings between them, we obviously need a definition of morphism of categories. In fact it is entirely straightforward (exercise: try writing down the following definition before you read it):

Definition 1.4 Let $\mathcal{C}$ and $\mathcal{D}$ be categories. A *functor* $F : \mathcal{C} \to \mathcal{D}$ consists of two mappings (both denoted F; this usually causes no confusion) from ob $\mathcal{C}$ to ob $\mathcal{D}$ and from mor $\mathcal{C}$ to mor $\mathcal{D}$, satisfying $\mathrm{dom}(Ff) = F(\mathrm{dom}\ f)$, $\mathrm{cod}(Ff) = F(\mathrm{cod}\ f)$, $1_{FA} = F(1_A)$ and $F(fg) = (Ff)(Fg)$ whenever fg is defined.

Of course, functors can be composed, and so we can add another example to the list in 1.3; we write **Cat** for the category of all small categories and functors between them. (We refrain from considering 'the category of *all* categories' in order to avoid getting into trouble with the Russell paradox; clearly, **Cat** is not itself a small category.)

Examples 1.5 (a) All the categories in 1.3(b) and (c) admit what we call *forgetful functors* to **Set**, in which we send (for example) a group to its underlying set and a homomorphism to itself considered as a mere function. We can also consider functors which 'forget' some but not all of the structure: for example, there is a forgetful functor from **Rng** to **AbGp** (the category of abelian groups) which forgets the multiplicative structure of a ring but remembers the additive one. Similarly, among the topological examples of 1.3(c) we have forgetful functors **Met** $\to$ **Top** and **Mfd** $\to$ **Top**.

(b) Many familiar constructions in algebra and topology may be viewed as functors. As an example, we give the construction of free groups: recall that, for any set A, there is a group FA which is freely generated by A in the sense that it contains a copy of A and any function $A \to G$, where G is the underlying set of a group, extends uniquely to a homomorphism $FA \to G$. To make F into a functor **Set** $\to$ **Gp**, suppose given a function $A \to B$; we define Ff to be the unique homorphism extending the composite of f with the inclusion $B \to FB$. As so often in this subject, the functoriality (that is, compatibility with composition) follows from the uniqueness; given another function $g\colon B \to C$, $F(gf)$ and $(Fg)(Ff)$ are both homomorphisms extending the same mapping $A \to FC$, so they must be equal.

(c) The power-set construction may be made into a functor **Set** $\to$ **Set**. Given a set A, we write PA for the set of all subsets of A; and given $f\colon A \to B$, we define $Pf\colon PA \to PB$ by taking images, i.e. $Pf(A') = \{f(a) \mid a \in A'\} \subseteq B$.

(d) But we may also make the power-set into a functor by taking inverse images. Given A, we define P^*A to be the same set as PA, but given $f : A \to B$ we define $P^*f : PB \to PA$ by $P^*f(B') = \{a \in A \mid f(a) \in B'\}$. This makes P^* into a functor **Set** $\to$ **Set**$^{\mathrm{op}}$ (or **Set**$^{\mathrm{op}} \to$ **Set**; the two things are defined by exactly the same data). We sometimes use the term '*contravariant* functor from $\mathcal{C}$ to $\mathcal{D}$' to mean a functor $\mathcal{C}^{\mathrm{op}} \to \mathcal{D}$; in this context, we also talk of *covariant* functors to distinguish those which do not reverse the direction of morphisms.

(e) Another familiar contravariant functor is the dual-space construction in linear algebra. Given a vector space V over a field k, we write V^* for the space of linear forms on V (that is, linear maps $V \to k$); and given a linear map $f\colon V \to W$, we write $f^*\colon W^* \to V^*$ for the operation of composing linear forms with f. The fact that this defines a functor $\mathbf{Vect}_k^{\mathrm{op}} \to \mathbf{Vect}_k$ should have been proved in any first course on linear algebra.

(f) The mapping $\mathcal{C} \mapsto \mathcal{C}^{\mathrm{op}}$ defines a functor **Cat** $\to$ **Cat**. Note that this functor is covariant; that is, in applying the duality principle of 1.3(f) to a proposition invoving several categories and functors between them, we must reverse the morphisms within each category, but not the functors between them.

(*g*) For the small categories of 1.3(*g*) and (*i*), the notion of functor reduces to what you would expect: a functor between monoids is just a monoid homomorphism, and a functor between posets is just a monotone (i.e. order-preserving) map.

(*h*) But we can also consider functors between these small categories and larger ones. For example, given a group G considered as a category, what is a functor $G \to \mathbf{Set}$? It picks out a particuar set A (the image of the unique object of G), together with, for each $g \in G$, a mapping $a \mapsto g \,.\, a \colon A \to A$, subject to compatibility with composition and identities. In other words, it is just an *action* or *permutation representation* of G on the set A. Similarly, a functor $G \to \mathbf{Vect}_k$ is just a k-linear representation of G. So one can say that the theory of group representations is subsumed in the theory of functors.

(*i*) The fundamental groupoid construction may be viewed as a functor $\mathbf{Top} \to \mathbf{Cat}$ (and the fundamental group as a functor $\mathbf{Top}_* \to \mathbf{Gp}$, where $\mathbf{Top}_*$ denotes the category of spaces with chosen basepoint and continuous maps preserving the basepoints). We shall not have very much to say about these functors in this course, but it was in order to provide a formal framework for the study of such links between topology and algebra that category theory was invented.

One feature of category theory which distinguishes it from other subjects you have studied up to now is the existence of a 'third level' of structure: in addition to objects and morphisms between the objects, we also have 'morphisms between the morphisms'. The formal definition is as follows:

Definition 1.6 Let $\mathcal{C}$ and $\mathcal{D}$ be categories and $F, G \colon \mathcal{C} \rightrightarrows \mathcal{D}$ two functors. A *natural transformation* $\alpha \colon F \to G$ is an operation assigning to each $A \in \mathrm{ob}\,\mathcal{C}$ a morphism $\alpha_A \colon FA \to GA$ in $\mathcal{D}$, such that for every $f \colon A \to B$ in $\mathcal{C}$ the square

$$\begin{array}{ccc} FA & \xrightarrow{Ff} & FB \\ \downarrow{\scriptstyle \alpha_A} & & \downarrow{\scriptstyle \alpha_B} \\ GA & \xrightarrow{Gf} & GB \end{array}$$

commutes (i.e., $(Gf)(\alpha_A) = (\alpha_B)(Ff)$). (We refer to this square as the *naturality square* for f.)

As with functors, we note that natural transformations can be composed: if we are given α as above and a natural $\beta \colon G \to H$, the (pointwise) composite $\beta\alpha$ is a natural transformation $F \to H$. Thus we have a category $[\mathcal{C}, \mathcal{D}]$ whose objects are the functors $\mathcal{C} \to \mathcal{D}$, and whose morphisms are the natural transformations between them. These 'functor categories' will play a very important role in what follows.

Examples 1.7 (*a*) For any vector space V we have a natural mapping $\alpha_V : V \to V^{**}$ sending a vector v to the linear form 'evaluate at v': $V^* \to k$. These mappings form a natural transformation $1_{\mathbf{Vect}_k} \to **$; and it was of course this eample which inspired the name 'natural transformation', since long before category theory existed people were accustomed to say that a finite-dimensional space is naturally isomorphc to its second dual, but only 'unnaturally' isomorphic to its first dual.

(*b*) Let $U : \mathbf{Gp} \to \mathbf{Set}$ denote the forgetful functor. The way in whch we made the free group construction into a functor in 1.5(*b*) precisely ensures that the operation sending a set A to the inclusion of generators $\eta_A \colon A \to FA$ is a natural transformation $1_{\mathbf{Set}} \to UF$.

(*c*) For each set A, we have a mapping $A \to PA$ sending $a \in A$ to the singleton set $\{a\}$. These form a natural transformation $1_{\mathbf{Set}} \to P$, since $Pf(\{a\}) = \{f(a)\}$ for any a.

(*d*) For monotone maps of posets, we have a (unique) natural transformation $f \to g$ iff $f(a) \leq g(a)$ for all a (the 'naturality condition' being vacuous).

(*e*) Suppose given group homomorphisms $u, v\colon G \rightrightarrows H$. What is a natural transformation $u \to v$? It consists of an element $h \in H$ such that $hu(g) = v(g)h$ for all $g \in G$, or equivalently $v(g) = hu(g)h^{-1}$, i.e. u and v are *conjugate* homomorphisms.

There is an obvious notion of isomorphism of categories: for example, **Rel** is isomorphic to $\mathbf{Rel}^{\mathrm{op}}$ via the functor which sends a set to itself and a relation R to its *converse* $R^\circ = \{(b,a) \mid (a,b) \in R\}$, since this functor is its own inverse. However, the third level of structure introduced in 1.6 allows us to introduce a weaker notion of equivalence, which is in many ways more useful. To define it, we need to talk about natural isomorphisms; there are two potentially different meanings of this phrase, but fortunately they coincide.

Lemma 1.8 *Let $\alpha\colon F \to G$ be a natural transformation between functors $\mathcal{C} \rightrightarrows \mathcal{D}$. Then α is an isomorphism in the functor category $[\mathcal{C}, \mathcal{D}]$ iff it is a pointwise isomorphism, i.e. each α_A is an isomorphism in $\mathcal{D}$.*

Proof One direction is obvious since composition in $[\mathcal{C}, \mathcal{D}]$ is pointwise. Conversely, suppose each α_A has an inverse β_A; we must show that the β's form a natural transformation. But, given any $f\colon A \to B$ in $\mathcal{C}$, we have

$$(Ff)\beta_A = \beta_B \alpha_B (Ff) \beta_A = \beta_B (Gf) \alpha_A \beta_A = \beta_B (Gf)$$

using the naturality of α at the middle step. □

Definition 1.9 Let $\mathcal{C}$ and $\mathcal{D}$ be categories. By an *equivalence* between $\mathcal{C}$ and $\mathcal{D}$, we mean a pair of functors $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{C}$ together with natural isomorphisms $\alpha\colon 1_{\mathcal{C}} \to GF$ and $\beta\colon FG \to 1_{\mathcal{D}}$. We write $\mathcal{C} \simeq \mathcal{D}$ if there exists an equivalence between $\mathcal{C}$ and $\mathcal{D}$.

There is a reason why, in the definition, we have chosen to make α and β point in opposite directions; but this reason will not become apparent until chapter 3, when we discuss adjunctions. For the moment, therefore, you should regard it simply as an idiosyncrasy of the author.

We are primarily interested in properties of categories which are invariant under equivalence (that is, if $\mathcal{C}$ has the property and $\mathcal{C} \simeq \mathcal{D}$, then $\mathcal{D}$ has it too); we call these *categorical properties*. For example, being a groupoid and being a preorder are categorical properties; being a group and being a partial order are not.

Examples 1.10 (*a*) Let $\mathbf{Set}_*$ denote the category of sets with a chosen basepoint and functions preserving the basepoints. There is an equivalence between $\mathbf{Set}_*$ and the category **Part** introduced in 1.3(*e*): we define $F\colon \mathbf{Set}_* \to \mathbf{Part}$ by $F(A,a) = A \setminus \{a\}$ and, for $f\colon (A,a) \to (B,b)$, Ff is the partial function defined by $Ff(x) = f(x)$ if $f(x) \neq b$, and undefined otherwise. In the other direction, we define $G(A) = (A \cup \{A\}, A)$ and $Gf(x) = f(x)$ if ($x \in A$ and) $f(x)$ is defined, $Gf(x) = B$ otherwise. It is clear that FG is the identity functor on **Part**; GF is not the identity, since the new basepoint we add via G may not be the same as the one we removed via F; but it is clearly naturally isomorphic to the identity. Note, incidentally, that these two categories are *not* isomorphic: in **Part** there is an isomorphism class of objects, namely $\{\emptyset\}$, with just one member, whereas each isomorphism class in $\mathbf{Set}_*$ has many members. (Obviously, an isomorphism of categories would induce a bijection from each isomorphism class of objects in the first category to the corresponding class in the second.)

(*b*) Let $\mathbf{fdVect}_k$ denote the category of finite-dimensional k-vector spaces and all linear maps between them. Then $\mathbf{fdVect}_k$ is equivalent to its opposite: both functors are the dual-space functor of 1.5(*d*), and both natural isomorphisms are the transformation α of 1.7(*a*) (whch of course becomes an isomorphism when we restrict to finite-dimensional spaces).

(c) $\mathbf{fdVect}_k$ is also equivalent to the category $\mathbf{Mat}_k$ of 1.3(j). We define $F: \mathbf{Mat}_k \to \mathbf{fdVect}_k$ by sending n to the standard vector space k^n, and a matrix to the linear map it represents with respect to the standard bases of these spaces. To obtain a functor in the other direction, we need to make some arbitrary choices: specifically, we choose a basis for every finite-dimensional vector space. Then we define $G(V) = \dim\ V$, and G sends a linear map to the matrix representing it with respect to the chosen bases. The composite GF is the identity (provided we choose the standard bases for the standard spaces k^n); FG is not, but the chosen bases yield for each V an isomorphism $k^{\dim\ V} \to V$, which is the V-component of a natural isomorphism from FG to the identity.

In many examples of equivalences, like 1.10(c), we find that the functor in one direction can be defined without arbitrary choices, but the other one requires such choices. There is a useful lemma which tells us when such choices are possible; but before stating it we need to introduce the notions of faithful and full functors. Note that, since equality of objects is a less interesting notion than isomorphism, it is rarely of interest to ask whether a given functor is injective or surjective on objects; for the same reason, we usually do not want to know whether it is globally injective or surjective on morphisms. What is of interest is injectivity or surjectivity on morphisms with a specified domain and codomain, as in the first two clauses of the following definition.

Definition 1.11 Let $F: \mathcal{C} \to \mathcal{D}$ be a functor.

(a) We say F is *faithful* if, given f and g in mor $\mathcal{C}$, the equations dom f = dom g, cod f = cod g and $Ff = Fg$ together imply $f = g$. (The name is borrowed from representation theory: group theorists call a permutation or linear representation of G faithful precisely if it is a faithful functor $G \to \mathbf{Set}$ or $G \to \mathbf{Vect}_k$ — though in this case the equality of domains and codomains is automatic.)

(b) We say F is *full* if, given $g: FA \to FB$ in $\mathcal{D}$, there exists $f: A \to B$ in $\mathcal{C}$ with $Ff = g$.

(c) We say F is *essentially surjective* if it is surjective on isomorphism classes of objects, i.e. every object B of $\mathcal{D}$ is isomorphic to FA for some $A \in$ ob $\mathcal{C}$.

Note, incidentally, that a full and faithful functor is necessarily injective on isomorphism classes of objects: if $g : FA \to FB$ is an isomorphism, then the unique $f : A \to B$ with $Ff = g$ is also an isomorphism (its inverse being the unique morphism $B \to A$ mapped by F to g^{-1}). We also use the term *full subcategory* for one whose inclusion functor is full: for example, **Gp** is a full subcategory of the category **Mon** of monoids, since a monoid homomorphism between groups automatically preserves inverses, but **Mon** is not a full subcategory of the category **SGp** of semigroups, since a semigroup homomorphism between monoids (i.e. a mapping preserving binary products) need not preserve the identity element.

Lemma 1.12 *A functor $F : \mathcal{C} \to \mathcal{D}$ is part of an equivalence of categories iff it is full, faithful and essentially surjective.*

Proof First suppose G, α and β exist as in 1.9. Essential surjectivity is immediate from the existence of β. Also, for any $f: A \to B$ we have $f = (\alpha_B)^{-1}(GFf)\alpha_A$, so we can recover f from Ff (plus its domain and codomain); hence F is faithful. Finally, if $g: FA \to FB$, set $f = (\alpha_B)^{-1}(Gg)\alpha_A$; then $GFf = Gg$, but the symmetry in the definition of equivalence ensures that G is also faithful, so $Ff = g$.

Conversely, suppose the conditions are satisfied. For each object B of $\mathcal{D}$, choose an object GB of $\mathcal{C}$ and an isomorphism $\beta_B: FGB \to B$; this defines the functor G on obects. To define it on morphisms, suppose $g: B \to C$ in $\mathcal{D}$; we define $Gg: GB \to GC$ to be the unique morphism whose image under F is the composite $(\beta_C)^{-1}g\beta_B$. As before, uniqueness ensures functoriality: if (g, h) is a composable

pair in $\mathcal{D}$, then $G(gh)$ and $(Gg)(Gh)$ have the same image under F, so they are equal. By construction, β is a natural isomorphism $FG \to 1_{\mathcal{D}}$; so it remains to define α. We take $\alpha_A : A \to GFA$ to be the unique morphism whose image under F is $(\beta_{FA})^{-1}$; the faithfulness of F ensures both that α_A is an isomorphism (its inverse is the unique morphism sent to β_{FA}) and that α is natural (its naturality squares are mapped by F to naturality squares of β^{-1}, so they commute). □

Definition 1.13 We say a category $\mathcal{C}$ is *skeletal* if each isomorphism class of objects of $\mathcal{C}$ has just one member. By a *skeleton* of an arbitrary category $\mathcal{C}$, we mean a full subcategory $\mathcal{C}_0$ containing just one member of each isomorphism class of objects.

Thus for example $\mathbf{Mat}_k$ is a skeletal category; it is not literally a skeleton of $\mathbf{fdVect}_k$, but the full subcategory of the latter consisting of the spaces k^n is so. An equivalence functor between skeletal categories is necessarily (bijective on objects, and hence) an isomorphism, and 1.12 tells us that any category is equivalent to any of its skeletons; so it is tempting to think that we might restrict our attention to skeletal categories, and thus get rid of the concept of equivalence. But the problem with this approach is the need to keep making arbitrary choices; in fact almost any general statement one can make about skeletal categories turns out to be equivalent to the axiom of choice (see Exercise 1.18).

In 1.11 we considered notions of injectivity and surjectivity for functors; we also need similar notions for morphisms within a category. There are several possibilities (we shall meet others later), but the simplest are those of monomorphism and epimorphism, which we now define.

Definition 1.14 We say a morphism f is a *monomorphism* if it is left cancellable, i.e. if the equation $fg = fh$ implies $g = h$ whenever fg and fh are defined. (The adjectival form of 'monomorphism' is 'monic', though some authors simply use 'mono'.). Dually, f is an *epimorphism* (or epic) if it is a monomorphism in the opposite category. We say a category $\mathcal{C}$ is *balanced* if every morphism which is both monic and epic is an isomorphism.

In diagrams, we often denote monomorphisms by arrows with tails (i.e. $A \rightarrowtail B$) and epimorphisms by double-headed arrows ($A \twoheadrightarrow B$).

Examples 1.15 (*a*) In **Set**, monomorphisms are just injective functions: one direction follows from the identification of elements $x \in A$ with morphisms $1 \to A$, and the converse is straightforward. Similarly, epimorphisms are just surjections; again, the left-to-right implication is straightforward, and the converse follows from considering morphisms to $2 = \{0, 1\}$. So **Set** is a balanced category.

(*b*) In **Gp**, one again finds that monomorphisms coincide with injective homomorphisms (use the free group $\mathbb{Z}$ on one generator as a substitute for 1), and epimorphisms with surjections (though the proof of the latter is nontrivial — it requires the use of free products with amalgamation). So **Gp** is also balanced.

(*c*) In **Rng**, monomorphisms are again injections, by a similar argument; but epimorphisms need not be surjective — the inclusion $\mathbb{Z} \to \mathbb{Q}$ is easily seen to be epic as well as monic. So **Rng** is not balanced.

(*d*) In **Top**, monos and epis are injections and surjections, by arguments similar to those of (*a*); but there are continuous bijections which are not homeomorphisms, so **Top** is not balanced.

The last two examples point to the need for stronger notions than 'mere' monomorphisms and epimorphisms; we shall supply these in due course.

Exercises on Chapter 1

Exercise 1.16 Let L be a distributive lattice (i.e. a partially ordered set with finite joins and meets — including the empty join 0 and the empty meet 1 — satisfying the distributive law

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$

for all $a, b, c \in L$). Show that there is a category $\mathbf{Mat}_L$ whose objects are the natural numbers, and whose morphisms $n \to p$ are $(p \times n)$ matrices with entries from L, where we define 'multiplication' of such matrices by analogy with that of matrices over a field, interpreting $\wedge$ as multiplication and $\vee$ as addition. Show also that if L is the two-element lattice $\{0, 1\}$, then $\mathbf{Mat}_L$ is equivalent to the category $\mathbf{Rel}_f$ of finite sets and relations between them.

Exercise 1.17 A morphism $e\colon A \to A$ is called *idempotent* if $ee = e$. An idempotent e is said to *split* if it can be factored as fg where gf is an identity morphism.

(i) Let $\mathcal{E}$ be a class of idempotents in a category $\mathcal{C}$: show that there is a category $\mathcal{C}[\check{\mathcal{E}}]$ whose objects are the members of $\mathcal{E}$, whose morphisms $e \to d$ are those morphisms $f\colon \mathrm{dom}\ e \to \mathrm{dom}\ d$ in $\mathcal{C}$ for which $dfe = f$, and whose composition coincides with composition in $\mathcal{C}$. [*Hint*: first show that the single equation $dfe = f$ is equivalent to the two equations $df = f = fe$. Note that the identity morphism on an object e is not $1_{\mathrm{dom}\ e}$, in general.]

(ii) If $\mathcal{E}$ contains all identity morphisms of $\mathcal{C}$, show that there is a full and faithful functor $I\colon \mathcal{C} \to \mathcal{C}[\check{\mathcal{E}}]$, and that an arbitrary functor $T\colon \mathcal{C} \to \mathcal{D}$ can be factored as $\widehat{T}I$ for some $\widehat{T}$ iff it sends the members of $\mathcal{E}$ to split idempotents in $\mathcal{D}$.

(iii) Deduce that if all idempotents split in $\mathcal{D}$, then the functor categories $[\mathcal{C}, \mathcal{D}]$ and $[\widehat{\mathcal{C}}, \mathcal{D}]$ are equivalent, where $\widehat{\mathcal{C}} = \mathcal{C}[\check{\mathcal{E}}]$ for $\mathcal{E}$ the class of all idempotents in $\mathcal{C}$.

Exercise 1.18 (i) Show that the assertion 'Every small category has a skeleton' implies the axiom of choice. [Given a family $(A_i \mid i \in I)$ of nonempty sets, consider a suitable category whose objects are pairs (i, a) with $a \in A_i$.]

(ii) Show that the assertion 'If $\mathcal{C}_0$ is a skeleton of a small category $\mathcal{C}$, then $\mathcal{C} \simeq \mathcal{C}_0$' implies the assertion that, given a family $(A_i \mid i \in I)$ of nonempty sets, we can find a family $(A'_i \mid i \in I)$ where each A'_i is a nonempty finite subset of A_i. [In standard ZF set theory, this is equivalent to the full axiom of choice. Take a category whose objects are the members of $I \times \{0, 1\}$, with the morphisms $(i, j) \to (i, k)$ being formal finite sums $\sum_{s=1}^{t} n_s a_s$ where all the a_s are in A_i and the n_s are integers with $\sum_{s=1}^{t} n_s = k - j$.]