This is a portion of the eBook at doi:10.7551/mitpress/12581.001.0001

1 Categories

In which we meet categories, explore some notable examples, consider distinctions of size, formulate some of the important ways of constructing new categories from old, and reflect on an alternative definition.

The language of category theory is indispensable to the presentation and understanding of the notions of sheaf theory. Chapters 1–3, together with chapters 6 and 7, will motivate and develop all the category theory needed in this book, emphasizing constructions and perspectives that will take center stage in the development of sheaves.

1.1 Categorical Preliminaries

Fundamentally, the specification of a category involves two main tasks: establishing some *data* or series of givens, and then ensuring that this data conforms to two simple axioms or laws. To define, or verify that one has, a category, one should first make sure the right data is present. This first main step of establishing the data of a category really involves doing four things.

First of all, it means identifying a collection of *objects*. Especially when one is assembling a category out of already established mathematical materials, these objects will typically already go by other names, like vertices, sets, vector spaces, topological spaces, types, various algebras or structured sets, and so on.

Second, one must assemble or specify a collection of *morphisms* (also called *arrows* or *maps*), each with two objects associated to it, namely a dedicated "source" object and "target" object. Fundamentally, a morphism is just some principled way of establishing connections between the objects. We depict morphisms with arrows—for example, a morphism f with source object A and target object B is represented diagrammatically by

$$A \xrightarrow{f} B.$$

Again, when dealing with already established structures, morphisms will usually already have names, like directed edges, functions, linear transformations, continuous maps, terms, homomorphisms or structure-preserving maps, and so on. Many of the categories one meets in practice have sets with some structure or supplementary furnishings attached to them for objects and (the corresponding) maps or functions between the underlying sets for morphisms (where these "respect the structure"), so this is a good model to keep in mind at the outset.

Third, and perhaps most important, one must specify an appropriate notion of *composition* for the morphisms, where for the moment this can be thought of in terms of specifying an operation that enables us to form a "composite morphism" that goes directly from object A to object C whenever there is a morphism from A to some B such that this morphism can be juxtaposed with another morphism that lands in C (in particular, whenever the "source" of this second morphism is the same B that was the "target" of the first morphism). In other words, given any pair of morphisms

$$A \xrightarrow{f} B_1 \qquad\qquad B_2 \xrightarrow{g} C$$

such that the target of the first morphism f is in fact the same object as the source of the second morphism g (i.e., $B_1 = B_2$)—in which case, they are *compatible for composition*—then there exists a specified way of combining these mappings to get a resulting morphism $g \circ f : A \to C$, called the *composite* (of f and g). We use the notation $\circ$ to denote composition, and it is read as "following" (or "after"), so that with $g \circ f$, for instance, we would have "g following f." In other words, g gets applied after f—as such, one might parse this by reading right-to-left: first apply f, then run g on the result. As a very simple example, when dealing with sets of numbers equipped with structural mappings corresponding to addition and multiplication, if f is the function defined by $f(x) = x^2$ and g is the function defined by $g(x) = x + 3$, then we know what the composite of f and g must be, namely $(g \circ f)(x) = g(f(x)) = x^2 + 3$.

We shall see that this composition operation in fact already determines the fourth data item of a category: that for each object, there is assigned a unique *identity* morphism that starts out from that object and returns to itself.

$$A \xrightarrow{\mathrm{id}_A} A$$

If you think for the moment of the model supplied by those commonly encountered categories that have structured sets for objects and structure-preserving maps for morphisms, then you might think of the identity morphism as the "trivial" action, the one that trivially preserves the structure by effectively "doing nothing." These four constituents—objects, morphisms, composite morphisms, and identity morphisms—supply us with the data of the (candidate for a) category.

Next, one must show that the data given above conforms to two very natural laws or axioms governing compositions. The first axiom concerns the behavior of the identity morphisms under composition. Consider how, for an arbitrary morphism $A \xrightarrow{f} B$, such a morphism will automatically be compatible for composition with the identity morphisms (on either end). The first axiom effectively says that morphisms that are *supposed* to do nothing (i.e., the designated identity morphisms) *really* do nothing when composed with other morphisms (i.e., really act as identities, or the "trivial" action, with respect to the morphisms with which they can be composed). More specifically, it stipulates that if we have a morphism f (as above) from source object A to target object B, then first applying the identity morphism on A and then traveling along the morphism f should be the same thing as "just" traveling along the morphism f; and the same goes for applying the morphism f straightaway and following this with the identity morphism on the target object B. In short, it is required that the identity morphisms do not do anything to change other morphisms with which they may be composed, in the precise sense that $f \circ \mathrm{id}_A = f = \mathrm{id}_B \circ f$. Observe that this identity axiom is effectively a condition on composition, namely that composing any morphism with the identity morphisms (on either side) is equal to the original morphism.

Second, suppose you have a string of morphisms

$$A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} D,$$

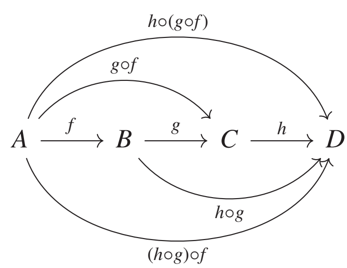
which are compatible for composition, by construction. Given such morphisms, a variety of distinct composites can be formed, yielding (in principle) distinct paths from A to D. For instance, while you could of course get from A to D by stepping through each of the individual morphisms as above, you might instead form the composite map $h \circ g$ and then use this, after taking f to B, to end up in D, as in

$$A \xrightarrow{f} B \xrightarrow{h \circ g} D.$$

Similarly, you could form the composite map $g \circ f$ and use this to get from A to C, and then take h to D, as in

$$A \xrightarrow{g \circ f} C \xrightarrow{h} D.$$

The second axiom just says that if you have a string of morphisms f, g, and h as above, then it should make no difference whether you choose first to go directly from A to C (using the composite map $g \circ f$ that we have by virtue of the third step in the data construction) followed by the map from C to D, or if you go from A to B followed by a direct map from B to D (using the composite map $h \circ g$). This axiom effectively says that composition is associative, in the sense that the outermost arrows of the following diagram are equal:



An entity that has all the data specified above, data that in turn conforms to the two laws described in the preceding two paragraphs, is a category. The mostly informal description given in the preceding paragraphs is given more formally in the following definition.

Definition 1 A *category* **C** consists of the following data:[10]

- A collection of *objects* $A, B, C, \ldots$;[11]

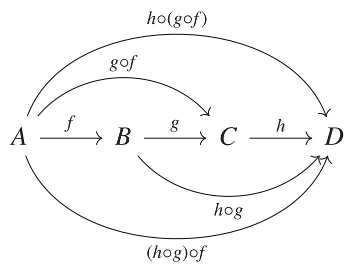
10. Throughout this book, categories are generally designated with bold font. However, sometimes we may use script font instead, especially when dealing with things like orders (discussed below), where each individual order is already a category. Context should always make it clear what category we are working with, so this should not be a problem.

11. The reader who finds themselves worried about what is meant by this apparently somewhat vague word "collection"—which, for now, you may take to mean "set" more or less (though "class" would be better, if that means something to you)—should be reassured that we will address what is going on here in section 1.3.

- A collection of *morphisms* $f, g, h, \ldots$, where each morphism has designated source object and target object—so that, for instance, $f : A \to B$ signifies that f is a morphism with source A and target B;[12]
- To each object A in the collection of objects is assigned a designated morphism in the collection of morphisms from A to A (i.e., where source and target are both the object A), denoted by id_A (or 1_A), called the *identity morphism* on A;
- For any pair of morphisms f, g such that the target of f is equal to the source of g—as in $A \xrightarrow{f} B$ and $B \xrightarrow{g} C$—there exists a *composite morphism* $A \xrightarrow{g \circ f} C$, with source equal to the source of f and target equal to the target of g.

This data gives us a category provided it further satisfies the following two axioms:

- *Associativity* (of composition): for any composable triple of morphisms f, g, h, as in $A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} D$, we have $h \circ (g \circ f) = (h \circ g) \circ f$.



- *Identity*: for any morphism $f : A \to B$, the composites formed by composing f with the identity morphism (on either side, i.e., on source or on target) are equal to f itself—that is, we have $f \circ \mathrm{id}_A = f$ and $\mathrm{id}_B \circ f = f$.

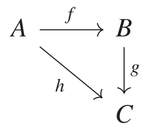
Before turning to examples of categories, let us make a brief observation. The associativity of composites may be expressed with what is called a *commutative diagram*, which is why a diagram appeared in the associativity axiom of the definition. While we will have much more to say about diagrams in later chapters, and be more precise about all this, for now observe how we have been displaying the various objects of a category, together with their morphisms and morphism composition, in the form of *diagrams*, where these are effectively directed graphs with morphisms as directed edges and objects as the names of the implied vertices or nodes attached to such edges. With the display of a morphism f with source A and target B as

$$A \xrightarrow{f} B,$$

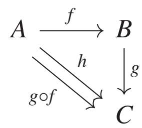
we already have an example of a basic diagram. Now suppose we have a diagram involving three morphisms, as in

12. The term "morphism" comes from *homomorphism*, which is how one refers in abstract algebra to a structure-preserving map between two algebraic structures of the same type (such as groups or rings). The morphisms of a category are also commonly referred to as "arrows" or "maps."

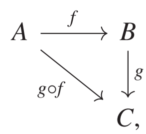
I should also note that while, in this definition, we are using $A, B, C, \ldots$ to range over objects, and $f, g, h, \ldots$ to range over morphisms, this convention will not always be respected. For instance, sometimes we will use $a, b, c, \ldots$, or some other natural names, for objects, and similarly reserve other appropriate notation for the arrows. These notation choices mostly just follow what is customary in the special topic being treated categorically, and this should not cause confusion.



Provided we are indeed working with a category, we know that the composable morphisms $f : A \to B$ and $g : B \to C$ will support the existence of the composite $g \circ f$ morphism, leaving us with a parallel pair of morphisms from A to C, as in



In principle, the two morphisms of such a parallel pair need not be the same—but when they are, so that $h = g \circ f$, we then say that this diagram (the triangle) commutes, and display it with the diagram



where occasionally a checkmark ✓ is drawn in the center if one wants to really stress that a diagram commutes. Other than "degenerate" diagrams involving identities, commutative triangles give us the most basic and paradigmatic instance of commutativity in diagrammatic form. But commutative diagrams one meets in the wild can have many more moving parts. In general, we will say that a diagram *commutes*, or is *commutative*, provided all directed paths with the same start and endpoints compose to give the same morphism—that is, a commutative diagram visualizes equalities between composite morphisms. As such, commutative diagrams are sometimes said to be for the category theorist what equations are for the algebraist. Commutative diagrams, and diagrammatic reasoning in general, are the main tool of the category theorist. They are not merely visual, intuitively spatial aids to the understanding of formal facts, but they even supply new proof techniques—especially that of the "diagram chase," which involves establishing a property of a particular morphism by tracking the components of a commutative diagram. A lot more will be said about diagrams throughout the book. For now, let us consider some examples of categories.

Example 2 **Set** is a category, where this consists of sets for objects and functions (with specified domain and codomain) for morphisms. Composition is given by the usual function composition (which is moreover associative), and identity morphisms are exactly what you imagine (where the identity functions moreover behave as the "units" for composition).

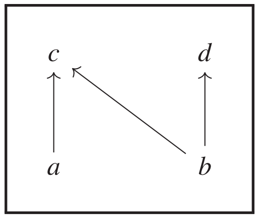
Example 3 (*Order Categories*) A *relation* between sets X and Y is just a subset $R \subseteq X \times Y$, so that a *binary relation* on X is a subset $R \subseteq X \times X$. It is customary to use infix notation for binary relations and use the symbol $\leq_X$ (or just $\leq$, if the carrier set is understood) for the relation on a set X, so that, for instance, one writes $a \leq b$ for $(a, b) \in R$.

We can then define a *preorder* as a set with a binary relation (call it $\leq$) that further satisfies the properties of being *reflexive* and *transitive*. In other words, it is a pair $(X, \leq_X)$ where we have

- $x \leq x$ for all $x \in X$ (reflexivity); and
- if $x \leq y$ and $y \leq z$, then $x \leq z$ (transitivity).

Then a *partially ordered set* (*poset* for short) is a preorder that is additionally *antisymmetric*, where this means that having $x \leq y$ and $y \leq x$ implies that $x = y$.

It is often useful to represent a given poset (or preorder) with a diagram. For instance, suppose we have an order-structure on the set $X = \{a, b, c, d\}$ given by $a \leq c, b \leq c, b \leq d$, together with the obvious identity (reflexivity) $x \leq x$ for all $x \in X$, which we will leave implicit. For a reason that will be better appreciated in a moment, the data of this poset may be displayed by the diagram:



Preorders (posets) can themselves be related to one another, and the right notion here is that of a monotone (or order-preserving) map.

Definition 4 A *monotone* (*order-preserving*) map between preorders (or posets) $(X, \leq_X)$ and $(Y, \leq_Y)$ is a function $f : X \to Y$ on the carrier sets satisfying that for all elements $a, b \in X$,

$$\text{if } a \leq_X b, \text{ then } f(a) \leq_Y f(b).$$

There is then a category **Pre**, the category having preorders for objects and order-preserving functions for morphisms. **Pos** is another category, one having posets for objects and order-preserving functions for morphisms.[13] Each identity arrow will just be the corresponding identity function, regarded as a monotone map. One can verify that for two monotone maps $X \xrightarrow{f} Y$ and $Y \xrightarrow{g} Z$ between orders, the function composition $g \circ f$ is also monotone.

While we are already talking about orders, let us introduce a few other order-theoretic notions, allowing us to further expand our repertoire of order categories. Suppose we further add the property to a poset that for all $a, b \in X$, either $a \leq b$ or $b \leq a$, that is, any two objects are *comparable*. Adding this condition gives us what are called *linear orders* (or, sometimes, *total orders*). If we take finite nonempty linear orders as our objects and monotone (order-preserving) functions between such linear orders as our morphisms, we get another category: **FLin**, the category of finite nonempty linear orders. For a simple example of such an order, take a natural number $n \in \mathbb{N}$, and consider the standard linear order $[n] = (\{0, 1, \ldots, n\}, \leqslant)$, where every such finite linear order may be represented by

$$\overset{0}{\bullet} \xrightarrow{\leq} \overset{1}{\bullet} \xrightarrow{\leq} \overset{2}{\bullet} \xrightarrow{\leq} \overset{3}{\bullet} \xrightarrow{\leq} \cdots \xrightarrow{\leq} \overset{\text{n}}{\bullet}$$

13. As can be seen from the few examples given thus far, it is common to let the *objects* determine the name of a category in question. While this is an entirely sensible practice, it is worth noting that it is at odds with the general spirit or philosophy of category theory, which gives priority to the morphisms (or at least demands that objects be considered together with their morphisms), a matter that is explored further in section 1.5.

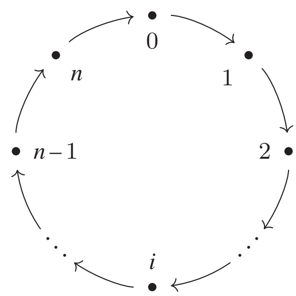
When we take such finite linear orders $[n]$ for our objects (one for each $n \in \mathbb{N}$) and the order-preserving maps between these for our morphisms—defined, for each pair of objects $[m], [n]$, as all the functions $f : \{0, 1, \ldots, m\} \to \{0, 1, \ldots, n\}$ such that for every pair of elements $i, j \in \{0, 1, \ldots, m\}$, if $i \leqslant j$, then $f(i) \leqslant f(j)$—we get what is called the *simplicial category* (or *simplex category*), typically denoted by $\mathbf{\Delta}$. As it turns out, these two categories can be shown to be equivalent, but we postpone making this precise until chapter 2, once we have the resources to do so. For now, this observation may justify thinking of the apparently enlarged category of finite nonempty linear orders in terms of the more manageable objects of the sort depicted above.

For a final example of this sort, we can also consider what is called a *cyclic order*. Just as the names imply, while a linear order is effectively an arrangement of elements along a line (where this has a prescribed direction), a cyclic order can be regarded as an arrangement of elements on a circle (where there is again a direction, such as clockwise). Observe that for elements arranged on a line with a specified direction, it makes sense to say things like "*a* comes before (or after) *b*"; however, for elements arranged in a circle with a direction (say, clockwise), the same sort of thing cannot be said—for instance, it doesn't make sense to say that "*a* is more (or less) clockwise than *b*." This observation motivates the definition of a cyclic order: it is given not in terms of a binary relation, but more naturally as a ternary relation $[a, b, c]$ (read "after a, one arrives at b before c") on a set. The months of the year form such a cyclic order, where we have, for instance, $[\mathit{March}, \mathit{September}, \mathit{February}]$ but not $[\mathit{March}, \mathit{February}, \mathit{September}]$. More formally, a cyclic order on a set is a ternary relation that satisfies the following properties (effectively ternary versions of the properties characterizing a linear order):

1. cyclicity: if $[a, b, c]$, then $[b, c, a]$;
2. asymmetry: if $[a, b, c]$, then not $[c, b, a]$;
3. transitivity: if $[a, b, c]$ and $[a, c, d]$, then $[a, b, d]$;
4. totality: if a, b, and c are distinct, then we have either $[a, b, c]$ or $[c, b, a]$.

The notion of a monotone (order-preserving) function between linear orders has its counterpart for cyclic orders in the following: given the cyclic orders $(X, []_X)$ and $(Y, []_Y)$, we say that a function $f : X \to Y$ is *monotone* provided it preserves the relation for any elements that have pairwise distinct images under f, that is, if given $[a, b, c]$ and the images $f(a), f(b)$, and $f(c)$ are pairwise distinct, then $[f(a), f(b), f(c)]$. This data ultimately gives us a category, one that has for objects all finite nonempty cyclically ordered sets, and for morphisms the monotone maps (in the above sense).

Similar to how we were able to generate a further category of interest by confining attention to the standard linear orders $[n]$, it is also useful to consider the standard cyclic orders, where a cyclic order on a finite set with n elements can be pictured as an (evenly spaced) arrangement of the set on an n-hour clockface.



Such an order is designated Λ_n. If we take such Λ_n as our objects (one for each $n \in \mathbb{N}$), and we take as our morphisms (from an object Λ_m to Λ_n) the monotone functions, then we arrive at what is called the *cyclic category* (or *cycle category*), usually denoted by $\boldsymbol{\Lambda}$. Following Connes and Consani (2015),[14] it is more common to see this category $\boldsymbol{\Lambda}$ given a description that takes the standard orders as objects (one for each $n \in \mathbb{N}$) but where the morphisms (from Λ_m to Λ_n) of $\boldsymbol{\Lambda}$ are instead seen as functions f on the integers that satisfy both of the following properties:

- nondecreasing/order-preserving: if $x \leq y$, then $f(x) \leq f(y)$;
- periodicity/"spiral property": $f(x+m+1) = f(x)+n+1$ for all $x \in \mathbb{Z}$.

Composition is then just the usual composition of functions. To complete the construction, one then takes equivalence classes of such functions under the relation $f \sim g$ whenever their difference is a constant multiple of $n+1$. This equivalence relation is compatible with the composition of functions, and there can be only finitely many such equivalence classes of functions for each pair of natural numbers m, n.

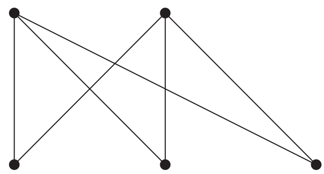
Orders, especially preorders and posets, are very important in category theory, and we will see a lot more of them throughout the book.

Example 5 A graph is typically represented by a bunch of dots or vertices together with edges between certain of the vertices, so that each edge is linking a pair of vertices, supplying what is called a relationship of *incidence* between the vertices and edges. More formally, a *(simple) graph* G consists of a set V of *vertices*, together with a collection of two-element subsets $\{x, y\}$ of V (where we generally assume $x \neq y$), called the *edges*. Sometimes the collection of two-element subsets of V is instead just represented by a set E that consists of the "names" of such pairings, where one then specifies an additional mapping that interprets edges as pairs of vertices. On this approach, a graph effectively consists of two sets—a "vertex set" V and an "edge set" E—together with a map that works to assign each edge to a pair of vertices (via a certain one-to-one function from E to 2-element subsets of V). The relevant notion of a map $G \to G'$ from a graph G to a graph G' is then given by a *graph homomorphism*, where this is a function $f : V \to V'$ on the vertices such that if $\{x, y\}$ is an edge of G, then $\{f(x), f(y)\}$ is an edge of G'.

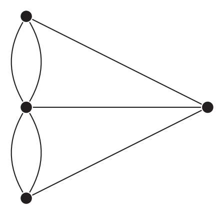
As the pairs of vertices making up the edge set were implicitly defined to be *unordered*, since they were said to be *sets* of the form $\{x, y\}$ rather than ordered pairs, the graphs

14. Alain Connes was the first to describe this category; see Connes (1983), where it is given a somewhat different description.

we just defined are *undirected*. Moreover, assuming that the map interpreting edges as unordered pairs of vertices does so in a one-to-one way amounts to requiring that the graph be *simple* in the sense of having *at most one* edge between two vertices. Taking such unordered simple graphs together with the associated notion of a graph morphism, we have the materials of a category, namely the category of undirected (simple) graphs, **UGrph**, or more commonly **SmpGrph**. The objects of this category are what the graph theorist usually means, by default, when they speak of a "graph." One such graph is pictured below (the names of vertices and edges left out):



If we had instead allowed the function from edges to pairs of vertices to be many-to-one, so that for each unordered pair of distinct vertices there could be an entire set of edges between these, we would be left with undirected *multigraphs*, such as



We can further define *directed graphs* (which are often thought of by the category theorist as *quivers*), where our edges now become arrows. These are the sorts of graphs that will be of most interest to us. Fundamentally, a (directed) *graph* $G = (V, A, s, t)$ can be seen as being comprised of two sets and two functions. Specifically, it consists of a set V of vertices, a set A of directed edges or arcs (arrows), and two functions

$$A \underset{t}{\overset{s}{\rightrightarrows}} V$$

that effectively act to pick out the *source* and *target* of an arc.

Then if $G = (V, A, s, t)$ and $G' = (V', A', s', t')$ are two graphs, the relevant notion of a morphism, namely a *graph homomorphism* $f : G \to G'$, is defined as a pair of morphisms $f_0 : V \to V'$ and $f_1 : A \to A'$ such that sources and targets are preserved, that is, both the diagrams

$$\begin{array}{ccc} A & \xrightarrow{f_1} & A' \\ {\scriptstyle s}\downarrow & & \downarrow{\scriptstyle s'} \\ V & \xrightarrow[f_0]{} & V' \end{array} \qquad\qquad \begin{array}{ccc} A & \xrightarrow{f_1} & A' \\ {\scriptstyle t}\downarrow & & \downarrow{\scriptstyle t'} \\ V & \xrightarrow[f_0]{} & V' \end{array}$$

commute, in the sense that we have

$$s' \circ f_1 = f_0 \circ s \text{ and } t' \circ f_1 = f_0 \circ t.$$

One can verify that composing two such graph homomorphisms f, g will leave us with another morphism $g \circ f$ that is itself a graph homomorphism. Altogether, we have a category called **DirGrph** (or just **Grph**), which has directed graphs as objects and their directed graph homomorphisms as morphisms.

In general, as before, we may further allow there to exist several parallel arrows—that is, with the same source and same target—in which case we would be dealing with directed *multigraphs*. Furthermore, if we were to add closed arrows or loops—that is, arrows whose source and target are identical—then we would be dealing with *looped* (or *reflexive*) graphs. There is a lot more to say about distinctions between different graphs, the distinct categories each gives rise to, and their categorical features of interest; but we will postpone further discussion of such matters.[15]

Example 6 It is probably fair to assume that the reader is already very familiar with vector spaces and the associated ideas of linear algebra. But if not, or in case a refresher is needed, here is the definition.

Let $\mathbf{k}$ be a given field—these supply us with the *scalars* by which we *scale*, by multiplying with, our vectors; so think, for concreteness, of $\mathbf{k} = \mathbb{R}$, the real numbers under addition and multiplication. Let V be a (nonempty) set with addition and scalar multiplications that assigns, to any $u, v \in V$, a *sum* $u + v \in V$, and to any $u \in V$ and $k \in \mathbf{k}$, a *product* $ku \in V$. Then V is said to be a *vector space* (or *linear space*) *over* $\mathbf{k}$, the elements of V being called *vectors*, if the following axioms hold:

- A_1: For any vectors $u, v, w \in V$, we have $(u + v) + w = u + (v + w)$.
- A_2: There exists a vector in V, denoted 0 and called the *zero vector*, for which $u + 0 = u$ for any vector $u \in V$.
- A_3: For each vector $u \in V$, there exists a vector in V, denoted by $-u$, for which we have $u + (-u) = 0$.
- A_4: For any vectors $u, v \in V$, we have $u + v = v + u$.
- M_1: For any scalar $k \in \mathbf{k}$ and any vectors $u, v \in V$, we have $k(u + v) = ku + kv$.
- M_2: For any scalars $a, b \in \mathbf{k}$ and any vector $u \in V$, we have $(a + b)u = au + bu$.
- M_3: For any scalars $a, b \in \mathbf{k}$ and any vector $u \in V$, we have $(ab)u = a(bu)$.
- M_4: For the unit scalar $1 \in \mathbf{k}$, we have $1u = u$ for any $u \in V$.

Matrices furnish us with a basic example of such a thing. Letting $\mathbf{M}_{m,n}$ denote the set of all $m \times n$ matrices over an arbitrary field $\mathbf{k}$, then $\mathbf{M}_{m,n}$ will be a vector space over $\mathbf{k}$, with the usual operations of matrix addition and scalar multiplication. Other important examples are given by the space of all polynomials $P[x]$ with coefficients in some field $\mathbf{k}$, and function spaces, where the elements are all functions from some given nonempty set into some $\mathbf{k}$.

Mappings between two vector spaces V, W over the same field are given by *linear transformations* (or *vector space homomorphisms*), where these are defined as functions $F : V \to W$ that satisfy, for any vectors $v, u \in V$ and any scalar $k \in \mathbf{k}$, the following two conditions:

15. For now, just note that while *quiver* and *directed graph* are often used synonymously, technically the graph theorist expects of its directed graphs that there is at most one arc from one vertex to another, and the notion of a quiver allows for there to be multiple "parallel" arcs between vertices, that is, a quiver is a directed multigraph (where loops are also allowed).

- $F(v+u) = F(v) + F(u)$
- $F(kv) = kF(v)$.

In other words, F is linear if it "preserves" the two fundamental operations of a vector space, vector addition and scalar multiplication.

The category $\mathbf{Vect}_{\mathbf{k}}$ is the category of $\mathbf{k}$-vector spaces (for a given field $\mathbf{k}$, dropping the $\mathbf{k}$ when this is understood), which has vector spaces $V, W, \ldots$ over $\mathbf{k}$ for its objects and linear transformations for its morphisms. To see that this is indeed a category, suppose $F: V \to W$ and $G: W \to U$ are linear transformations between the specified vector spaces. Then, for any $v, v' \in V$ and any $a, b \in \mathbf{k}$, we must have

$$\begin{aligned}(G \circ F)(av + bv') &= G(F(av + bv')) = G(aF(v) + bF(v')) \\ &= aG(F(v)) + bG(F(v')) = a(G \circ F)(v) + b(G \circ F)(v'),\end{aligned}$$

which just shows that the composite $G \circ F: V \to U$ must itself be a linear transformation. Moreover, for any vector space, the identity function will be a linear transformation. Finally, associativity of composition comes for free, since linear transformations are functions and function composition is always associative. If we restrict attention to just finite-dimensional vector spaces, this would yield the category **FinVect**, which is where most of linear algebra takes place.

In vector spaces, the scalars come from the given field and act on the vectors of the space by scalar multiplication (which then obeys the axioms). Every field is a ring, but there are rings that are not fields, as the notion of a ring generalizes that of a field (for a ring, multiplication need not be commutative and multiplicative inverses need not exist); for example, the ring of integers $\mathbb{Z}$ is not a field, since 2, for instance, has no multiplicative inverse in $\mathbb{Z}$. A *module* is just like a vector space, except the scalars need only come from a ring (with identity), and (left or right) multiplication is defined between elements of the ring and elements of the module. So while any ring R that is also a field recaptures the notion of vector spaces, there are plenty of modules that are not vector spaces. We call a module taking its coefficients in the ring R an R-module.

For a concrete illustration of a module, here is one that is often used by music theorists. First consider that, in general, for a set X and a vector space V, we can define the set of all functions from X to V. Then, for f, g in that set and for $c \in \mathbf{k}$ the field of scalars, addition is defined as $(f+g)(x) = f(x) + g(x)$ and scalar multiplication as $f(cx) = c(f(x))$ for all $x \in X$. It can be shown that this set of functions is itself a vector space (over the field $\mathbf{k}$). And in fact, for X the set of n-tuples, it can be shown that this function vector space is isomorphic to the vector space of n-tuples. Applied to music, we might accordingly consider the space $\mathbb{R}^{\{O,P,L,D\}}$, by taking $X = \{O, P, L, D\}$, where O stands for the onset values, P for pitch, L for loudness, and D for duration, and where these give ordered quadruples taking values in O in the first component, P in the second component, and so on. In other words, the members f of this function space can be thought of as the "note event" vector

$$f = (f_O, f_P, f_L, f_D),$$

where, for simplicity, onset f_O may come in units of quarter notes ♩, pitch f_P in units of semitones, loudness f_L in units of cents, and duration f_D in units of quarter notes. This in fact forms an $\mathbb{R}$-module (obviously, since it is in fact a vector space), one that is isomorphic to $\mathbb{R}^4$, the vector space of all ordered quadruples of real numbers (x_1, x_2, x_3, x_4), and it happens to be of some use to music theorists.

Returning to the more general account: a map between two R-modules will be a map that satisfies the same conditions on a linear transformation (except that the scalar is an $r \in R$), that is, it is a function between modules that "preserves" the module structures. More precisely, if we let M and N be R-modules, an R-*homomorphism* (or *module homomorphism*) from M to N is a map $f : M \to N$ that satisfies for every $a, b \in M$ and $r \in R$

- $f(a+b) = f(a) + f(b)$, and
- $f(r \cdot a) = r \cdot f(a)$.

Modules over a fixed ring R, together with such R-module homomorphisms, assemble into a category $\mathbf{Mod}_R$.

Given a homomorphism f from an R-module M to an R-module N, we can define another map $g : M \to N$ by $g(x) = n + f(x)$, where $n \in N$. Such a map is called an *affine transformation*, and when the underlying modules are the same, that is, $M = N$, then such a g captures *symmetries*. Affine maps can be composed just as functions are composed, and thus by taking R-modules together with such maps between them we end up with another category, one that might be denoted by $\mathbf{ModAff}_R$. Incidentally, such a category—and, more broadly, the algebraic theory of rings and modules—supplies a natural setting for the treatment of many aspects of music theory. For instance, if you think of a score of music in terms of the module $\mathbb{R}^{\{O, P, L, D\}}$ representing notes with onset, pitch, loudness, and duration, then morphisms as affine maps or symmetries on this module capture all kinds of musically meaningful (and geometrically representable) transformations—such as pitch transposition, vertical and horizontal inversion and dilation, rhythmic shift, onset-pitch arpeggio, and so on—and a variety a fundamental results concerning things like harmonic analysis and modulation can be derived within this framework.[16]

Example 7 The category **Mon** (**Group**) of monoids (groups) has monoids (groups) for objects and monoid (group) homomorphisms for morphisms. (This example, together with the necessary definitions, will be discussed in more detail in a moment.)

The previous examples involve categories whose objects are sets equipped with some additional structure and whose morphisms are functions that preserve the underlying structure. Categories such as these effectively take given structures of the same sort, together with their structure preserving maps, and package them into a single structure. Conceptually, such categories do what they sound like they do: essentially *categorizing* existing mathematical structures. As such, these categories supply us with a few salient examples of

16. Modules play a large role in much of mathematical music theory, especially variants of $\mathbb{R}^{\{O, P, L, D\}}$ and the module of integers mod 12 (under addition), together with its affine transformations. At least as far back as Lewin (see Lewin (2010), originally published in 1987), transformations between musical elements (instead of musical objects like the C major chord) became the focus of many music theorists, where the transformations in question often form a group. Certain elements of algebra, especially group theory, have accordingly been used for some time to systematize the treatment of common operations on musical chords, and geometrical models of musical structure have also been considered by a few authors (such as Tymoczko (2011)). Approaches to questions of music theory that incorporate category theory (starting with Mazzola (1985)) are somewhat harder to come by, but the reader curious about more category theoretic takes on music may find interesting Popoff, Andreatta, and Ehresmann (2018), Noll (2005), and the work of Guerino Mazzola and followers. The module $\mathbb{R}^{\{O, P, L, D\}}$ presented above, together with further uses of it, is discussed in Mazzola and Andreatta (2006).

what you might think of as *categories of structures*. When Eilenberg and Mac Lane (1945) first defined categories and the related notions allowing categories to be compared (introduced in the next chapter), they stressed how it provided "opportunities for the comparison of constructions... in different branches of mathematics." But with Grothendieck's *Tôhoku* paper a decade later,[17] it started to become evident that category theory was not just a convenient tool for comparing different mathematical structures, but was itself a significant mathematical structure of its own intrinsic interest. While earlier uses of category theory treat categories as largely dispensable tools for helping to identify properties of given mathematical entities such as Abelian groups or certain modules, in Grothendieck's paper categories become objects of mathematics in their own right, whose common properties start to take on an intense mathematical interest. One way of starting to appreciate the sort of shift here is to realize that we do not just have categories consisting *of* mathematical structures of interest, but equally important are those categories that allow us to view categories themselves *as* mathematical structures of interest in their own right. The following two examples supply an initial way into this perspective of *categories as structures* (each of which example reveals crucial features of categories in general and is accordingly often said to supply us with a means of doing "category theory in the miniature").

Example 8 (*Each order is already a category*) Let $(X, \leq_X)$ be a given preorder (or, less generally, a poset). It is easy to verify that we can form the category $\mathcal{X}$ by

- taking for objects of $\mathcal{X}$ the elements of X; and
- declaring that there exists a morphism in $\mathcal{X}$ from a to b exactly when $a \leq b$ (and there is *at most* one such arrow, so this morphism will necessarily be unique).

Notice how transitivity of the relation $\leq$ will automatically give us the required composition morphisms, while reflexivity of $\leq$ just translates to the existence of identity morphisms. Thus, we can regard any given preorder (poset) $(X, \leq_X)$ as a category $\mathcal{X}$ in its own right.

In the other direction, if for every pair of objects A, B in a category $\mathbf{C}$, there is at most one morphism between A and B, then $\mathbf{C}$ in fact defines a presentation of a preorder (poset).

Example 9 (*Each monoid is already a category*) A *monoid* $\mathcal{M} = (M, \cdot, e)$ is a set M equipped with

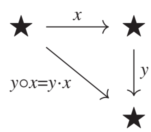
- an associative binary multiplication operation $\cdot : M \times M \to M$, that is, $\cdot$ is a function from $M \times M$ to M (a *binary operation on* M) assigning to each pair $(x, y) \in M \times M$ an element $x \cdot y$ of M, where this operation is moreover associative in the sense that
$$x \cdot (y \cdot z) = (x \cdot y) \cdot z$$
for all $x, y, z \in M$; and
- a designated "unit" element $e \in M$, where this acts as a two-sided identity in that it satisfies
$$e \cdot x = x = x \cdot e$$
for all $x \in M$.

17. See Grothendieck (1957); it is now common to refer to this as the Tôhoku paper.

Comparing this definition to that of a category, it is straightforward to see how any monoid $\mathscr{M}$ can be regarded as a category of its own. Specifically, it is a category with just one object. It does not matter what this object is taken to be, and a priori it has nothing to do with sets and certainly does not come with any structure. To indicate as much, we can just represent it with an arbitrary symbol, such as $\bigstar$. For morphisms of this category we just use the elements $x \in M$, that is, for each $x \in M$ there will be a morphism

$$\bigstar \xrightarrow{x} \bigstar$$

The identity morphism $\mathrm{id}_\bigstar$ is then taken from the monoid unit e and the composition formula for morphisms



from the monoid multiplication. As this operation is associative and e acts as an identity, we do indeed have a category.

In the other direction, notice that if $\mathbf{C}$ is a category with only one object C—or just picking one object out from the category—and we let $\mathrm{Hom}(C, C)$ denote its collection of morphisms (i.e., all the morphisms from C to itself, called "endomorphisms"), then $(\mathrm{Hom}(C, C), \circ, \mathrm{id}_C)$ will be a monoid.

Finally, an element $m \in M$ of a monoid is said to have an *inverse* provided there exists an $m' \in M$ such that $m \cdot m' = e$ and $m' \cdot m = e$. This lets us define a very important mathematical object: that of a group. A *group* is just a monoid for which every element $m \in M$ has an inverse. Moreover, if the group operation does not depend on the order in which two group elements are written, then the group is said to be *abelian*. Similar to what we saw with a monoid, any group itself can be shown to give rise to a category in which there is just one object, but where every morphism (given by the group elements) is now an isomorphism, in the following purely category theoretic sense:

Definition 10 In a category $\mathbf{C}$, a morphism $f : A \to B$ for which there exists a morphism $g : B \to A$ in $\mathbf{C}$ such that $g \circ f = \mathrm{id}_A$ and $f \circ g = \mathrm{id}_B$ is called an *isomorphism*.

Inverses are unique, so we can write $g = f^{-1}$. Observe that the objects A and B are then said to be *isomorphic*, denoted $A \cong B$, whenever there exists an isomorphism between them. Such a notion is used to inform us about when we can regard two objects of some category as *the same*.

The previous two examples—orders and monoids—are not just examples of *any old* categories, but in an important sense, categories in general may be regarded as a sort of fusion of preorders on the one hand and monoids on the other. Over and above the fact that each monoid and each preorder is itself already a category in its own right, these two examples are special in that, through them, we can appreciate how categories more generally are exceptionally "monoid-like" and "preorder-like." We saw that every monoid can be exhibited as a single-object category. Seen from the other side, categories in general may be regarded as the *many-object* version of monoids. We saw that every preorder can be exhibited as a single-arrowed category, as between any two objects there is at most one arrow. Seen from the other side, categories in general may be regarded as the *many-arrowed* version of preorders. Sometimes, like Leibnizian monads, within a small part of the universe, set off on its own, we can see reflections of the whole. In this sense, individual monoids and preorders are like microcosms in which we can glimpse "in the miniature" the essential features of the general notion of a category. Monoids furnish us with not just a study of composition "in the miniature" (by collapsing down to a single object), but in a sense the associative binary operation and neutral or identity element that comprise the data of a monoid seem to provide a prototype for the general associativity and identity *axioms* of a category. Preorders, for their part, furnish us not just with a study of comparison of objects via morphisms "in the miniature" (by collapsing down to at most one morphism from any object to another), but in a sense the reflexivity and transitivity of the order seems to provide the model for the key *data* specifying a category, that is, the assignment of an identity arrow to each object (via reflexivity) and the composition formula (via transitivity).

There is another important (if more philosophical) way in which monoids in particular can shed light on categories. This has the added benefit of introducing the interesting notion of *oidification* and an alternative (philosophically appealing, if somewhat less useful) definition of categories. In order not to unduly distract the reader, we press on with some additional examples of categories and other more pressing matters, leaving a brief section on this topic to the very end of the chapter.

1.2 A Few More Examples

There are many more categories that we might mention, and that are important to mathematicians. However, we will instead just draw attention to a few more categories, and let the rest that will be of particular use to us emerge organically throughout the book.

Example 11 **Top** is the category that has topological spaces for objects and continuous functions for morphisms. This category, and topology in general, is discussed in detail in chapter 4.

Example 12 **Measure** is the category that has measure spaces for objects and (on one definition) appropriate equivalence classes of measurable functions for morphisms.

Example 13 **Man** is the category that has smooth (infinitely differentiable) manifolds for objects, smooth maps for morphisms.

Example 14 Suppose we are given V a vector space. Then we can define a category $\mathbf{V}$ as follows:

- for objects: $\mathbf{V}$ has only one object, called $*$;
- for morphisms (arrows from $*$ to $*$): the vectors v in V;
- for the identity arrow for $*$: the zero vector; and
- for composition of vectors v and v': their sum.

Let us now have a look at an example involving a rather important category, one that starts to make better sense of the idea that, in being visualized by arrows between dots, category theory might be regarded as some sort of graph theory, but with something extra (where this involves some extra structure regarding *composition* of arrows). In our earlier definition of a (directed) graph from example 5, observe that there were no conditions placed on arcs and vertices other than those involving the source and target functions, picking out the source vertex and the target vertex of a given arc a; in particular, there was no requirement regarding the composition of arcs. Thus, it is not the case that a category *is* a graph, for a directed graph in general has no notion of composition of arcs (and does not even have a notion of identity arrows). However, any category—well, any small category[18]—does have an underlying graph. While the converse does not hold, it is an important fact that every directed graph can be *made* into a certain category, via a special construction, discussed in the following example.

Example 15 Given a directed graph G, we first describe the notion of a *path* in G, as any sequence of successive arcs where the target of one arc is the source of the other. More explicitly, for each $n \in \mathbb{N}$, we define a *path* through G of length n as a list of n arcs,

$$i(0) \xrightarrow{e(1)} i(1) \xrightarrow{e(2)} i(2) \longrightarrow \cdots \qquad \xrightarrow{e(n)} i(n)$$

where the target of each arc is the source of the next one. A path of length 1 would then amount to a single arc, while a path of length 0 would be a vertex (node). We can create a category $\mathbf{Pth}(G)$, the *category of paths* through G, with objects the vertices of G and for morphisms from objects x to y all the paths through G from x to y. Given two paths,

$$i(0) \longrightarrow i(1) \longrightarrow i(2) \longrightarrow \cdots \longrightarrow i(n)$$

and

$$j(0) \longrightarrow j(1) \longrightarrow j(2) \longrightarrow \cdots \longrightarrow j(m),$$

with the end node of the first equal to the start node of the second, that is, $i(n) = j(0)$, we form the *composite path* by concatenating or sticking together the two paths along this identical node, that is,

$$i(0) \longrightarrow i(1) \longrightarrow \cdots \longrightarrow i(n) = j(0) \longrightarrow j(1) \longrightarrow \cdots \longrightarrow j(m)$$

resulting in a new path from $i(0)$ to $j(m)$. Then, concatenating paths end to end is associative, making composition in $\mathbf{Pth}(G)$ associative. As for ensuring that each object (vertex of G) has an identity arrow in $\mathbf{Pth}(G)$, we can observe that each vertex has an associated "length 0" path, and sticking such a path at the end of another path does nothing to change that other path. Thus, we can just take the paths of length 0 as our identity arrows, that is, the identity at an object x is given by the path of length 0 from x to x. Moreover, given a graph homomorphism $f : G \to G'$, every path in G will be sent under f to a path in G'.

We will have more to say about this category, and the construction that generates it, in the next chapter.

1.3 Returning to the Definition and Distinctions of Size

Before moving into consideration of how we can get some new categories from old, let us take the opportunity to make an important observation, leading to some important notions and distinctions having to do with size. We have stressed how many of the examples given thus far are categories whose objects are *structured sets* and where the morphisms are

18. The meaning of this distinction is discussed in a moment, in section 1.3.

(some appropriately "structure-preserving") *functions* between the underlying sets. Categories of this sort are sometimes called *concrete categories*. Beyond **Set** itself, of those categories introduced thus far, **Pre**, **Pos**, **Top**, **Mon**(**Group**), $\mathbf{Mod}_R$, $\mathbf{Vect}_k$, **Grph**, and **Man** are all examples of concrete categories. Such examples might lead us to believe that all categories involve just considering some mathematical structures that are fundamentally set-like together with the appropriate notion of morphism between them (where this is some sort of function). However, in an arbitrary category, it cannot be assumed that objects are structured sets and morphisms structure-preserving functions, something that will become more evident especially as we consider more complicated categories. Moreover, the morphisms of a category are *not* always functions. For now, consider "nonconcrete" (or *abstract*) examples furnished by each poset (or preorder) regarded as a category, each group (or monoid) regarded as a category, and example 14. With a given monoid regarded as one object category, recall that the single object has nothing in principle to do with sets, and the morphisms don't carry any structure. Likewise, the morphisms of a given poset regarded as a category are just derived from the order relation, and a priori are not functions.

Considerations of this sort let us raise another important matter, taking us back to the original definition of a category we gave in definition 1. Recall how this definition mentioned a "collection" of objects and a "collection" of morphisms. While I originally suggested that the reader think of a "collection" roughly in terms of a "set," this is not in fact accurate, as the collection of objects may not be a set and arrows need not even be functions. In particular, some collections will not be sets for they are "too big" to be sets—an observation that leads to an important distinction of categories and a few useful definitions.

To better motivate discussion of these matters, first observe that for each of the concrete categories we have looked at, the "collection" of objects will not form a set. To appreciate this, it suffices to consider the case of the category **Set**. We said that this category was concrete, where this meant that the *objects* of the category are each themselves (perhaps structured) sets, and the arrows are functions (perhaps of a certain "structure-preserving" type). But notice what this does *not* say: it does not say that the *collection of all objects* of the category—which is what we are interested in, as far as the first data item of the definition is concerned—forms a set. Suppose that by "collection" in the definition of a category we *had* meant *set*. Then, as far as **Set** goes, the first data item of such a category would be a set of all sets. But for reasons having to do with Russell's paradox, which first emerged in the context of naive set theory, assuming such a set of all sets could even exist leads to a number of fatal foundational problems. So what is the status of the "collection" of objects?

The most direct way this sort of problem was addressed with set theoretical solutions involves prohibiting the unrestricted formation of sets so that such a *set* of all sets does not exist, and to introduce *classes* as a way of retaining "set-like" collections that can still be clearly defined by a property that all its members share, while being distinct from sets and so avoiding paradoxes of the sort first brought to light by Russell. In the Zermelo-Fraenkel (ZF) set theory the reader may be familiar with, one does not see a formal notion of classes, yet it is still common to speak informally of classes, referring to a class that is not a set as a *proper class* (or *large class*), while a class that is a set is called a *small class*.

The precise formal notion of "class"—and the consideration of whether a given entity constitutes a large class or not—ultimately depends on foundational context, that is, on the choice of a set theoretical framework that might allow for the formulation of such things. Strictly speaking, from the axioms of ZF, the class of all small classes (i.e., of all sets) cannot even be formally constructed and attempting to do so leads to contradictions of the sort involved in Russell's paradox. However, adopting certain extensions of ZF, such as the von Neumann-Bernays-Gödel axioms (NBG), allows us to formally distinguish classes and sets, and offers perhaps the most basic technical solution. Classes are taken to be the basic objects of the theory, and a set is formally defined in terms of it: a *set* is just a class that can be an element of other classes (and a *proper class*, then, is just a class that is not a set, in this sense). In such a context, the "collection" of all sets that supplies the objects of **Set** can be seen as a proper class, rather than a set—and the same would go for other similar categories, such as **Group**.

There are other ways of addressing the issues at stake here, including using "Grothendieck universes," but going deep enough into these matters to do them justice would take us too far afield.[19] Without stressing about which foundational framework we are using, let us agree to say that

Definition 16 A class is *small* if it is a set; it is *large* otherwise.

While any individual set—being a *set*!—is small, the point is that (within the foundational contexts that allow for the precise formulation of such notions) there are classes that are large, and the class of all sets is one example of such a thing.

Definition 17 A category **C** is said to be *small* if both the collection of objects of **C** and the collection of morphisms of **C** are sets; it is *large* otherwise.

Occasionally one will see a small category **C** defined as one for which the collection of all *morphisms* in **C** is small (large otherwise), that is, where there is no more than a set's worth of morphisms. Adopting the framework of NBG, then, we would be saying that a large category is one whose class of morphisms is a proper class; otherwise, the category is small. Such definitions that focus on the size of the collection of *morphisms* is ultimately the same as the one given in definition 17, using the fact that objects can be shown to correspond bijectively with identity morphisms, which morphisms of course form a subclass of the morphism class of a category, from which it follows that if **C** is small in the sense that its morphism class is small, then the class of objects of **C** will be small (i.e., a set) as well.

There are many categories that are not small, which can be seen by noting how the concrete categories introduced above each have too many objects to be small, and so are all examples of categories that are large. Yet, a category that is not small may locally—that is, between any pair of fixed objects—look like a small category. This notion is captured by the following definition:

Definition 18 A category is said to be *locally small* whenever between any pair of objects there is only a set's worth of morphisms.

19. Shulman (2008) is a nice reference for the reader interested in pursuing these size issues and the related foundational matters.

We could thus say that a category is small if and only if (iff) it is both locally small and its class of objects is small. Of course, any small category is thus locally small, but many large categories end up being locally small as well. For instance, the concrete categories we have seen thus far, while not small, are indeed locally small in this sense.[20] For a locally small category $\mathbf{C}$, it is customary to write $\mathrm{Hom}_{\mathbf{C}}(A, B)$ or (dropping the subscript) $\mathbf{C}(A, B)$ for the collection of morphisms from A to B in the category $\mathbf{C}$. As this collection is indeed a *set*, it is customary to call the set of morphisms between a pair of fixed objects a *hom-set* (regardless of whether these are literal *homomorphisms*). The reader should also be aware that it is now customary to use this hom-set notation even in the case of categories that are not necessarily locally small, a custom adopted in this book.

The distinctions just introduced will be useful to us going forward and issues related to the size of a category will occasionally resurface throughout the rest of the book. Before ending this section, though, let us address one potentially lingering issue. At this point, the reader who has followed this remark so far might be wondering:

> if removing any vagueness in the word "collection," thereby cashing in on the definition, is to make use of some rigorous foundation allowing for, say, the notion of *classes*, why not just say "class" and specify what is meant by this in the first place?

Adopting a rigorous foundation, such as NBG, capable of supporting the theory of classes, and then writing "class" everywhere we wrote "collection" in the definition is indeed a perfectly acceptable approach, and one will sometimes see authors do this. However, it is even more common to see the definition stated in terms of "collections," as we did. One might viably understand "collection" to encompass both sets and proper classes, as "class" does—but since the precise definition of "class" itself depends on the foundational context, one prefers not to commit to saying "class" (and having to specify the foundational context) and instead uses the more agnostic and deliberately vague "collection." Another reason for preferring the word "collection" is a little more philosophical: a category is really just *anything* that conforms to the definition's conditions. And if any rigorous foundation allowing us to make the relevant distinctions might be used in support of the definition, so that it is effectively independent of chosen foundation, then being overly specific about the particular set theoretical notion of class, for instance, seems to suggest that the concept of a category ultimately rests on some prior set theoretical framework and that any categorical treatment of set theory, for its part, may end up involving some circularity. But this is misleading and misrepresents the power of category theory. There are some thornier issues here, certainly, but let us instead end this section by noting that while **Set** is indeed a very important category, one is not to imagine that category theory somehow lives within set theory. In a way that we can make more precise in chapter 6, set theory itself can be thought of as doing "zero-dimensional" category theory. Moreover, since the 1970s with the work of Lawvere, efforts have been made to show that effectively all the mathematically significant portions of set theory and logic (in the narrow sense) can be seen as part of category theory.

20. In part because so many categories of interest are locally small, some authors accordingly even take local smallness as part of their definition of a category.

1.4 Some New Categories from Old

There are many important things one can do *to* categories, to generate new categories from old ones. Attention is confined, in this final section of the chapter, to those that will be most important for our purposes.

Definition 19 Let **C** be a category. The *dual* (or *opposite*) category $\mathbf{C}^{op}$ is then defined as follows:

- objects: same as the objects of **C**;
- morphisms: given objects A, B, the morphisms from B to A in $\mathbf{C}^{op}$ are exactly the morphisms from A to B in **C** (i.e., just reverse the direction of all arrows in **C**).

Identities for $\mathbf{C}^{op}$ are defined as before, composites are formed by interchanging the order of composition as one would expect, yielding a category. In more detail, for each **C**-arrow $f : A \to B$, introduce an arrow $f^{op} : B \to A$ in $\mathbf{C}^{op}$, so that ultimately, these give all and only the arrows in $\mathbf{C}^{op}$. Then the composite $f^{op} \circ g^{op}$ will be defined precisely when $g \circ f$ is defined in **C**, where for

$$A \underset{f^{op}}{\overset{f}{\rightleftarrows}} B \underset{g^{op}}{\overset{g}{\rightleftarrows}} C,$$

we have that $f^{op} \circ g^{op} = (g \circ f)^{op}$.

In slogan-form,

> *Given a category, just reverse all its morphisms and the order of composition, and you'll get another category (its* dual*)!*

With this seemingly innocuous construction, every result in category theory will have a corresponding dual, essentially got "for free" by simply formally reversing all arrows (and respecting the induced change in the order of composing arrows). In general, given a statement or construction framed in the language of category theory, when we refer to the *dual* of that statement or construction, we simply mean the statement or construction that is obtained by interchanging the source and target of each morphism as well as the order of composition of two morphisms. When a statement is true in a category **C**, then its dual will be true in the dual category $\mathbf{C}^{op}$—"by duality" will refer to this invariance of truth under the operations involved in taking the opposite category. Such duality not only can clarify and simplify relationships that are often hidden in applications or particular contexts but it also *reduces by half* the proof of certain statements (since the other, dual statement will "follow by duality")—or, to see things another way, it *multiplies by two* the number of results, as each theorem will have its corresponding dual. Finally, for any **C**, note that we will have that $(\mathbf{C}^{op})^{op} = \mathbf{C}$.

Next, we consider how, given a category **C**, we can form a new category by taking as our objects all the *arrows* of **C**.

Definition 20 For a category **C**, we define the *arrow category* of **C**, denoted $\mathbf{C}^{\to}$, as having

- objects: the morphisms $A \xrightarrow{f} B$ of **C**;

This is a portion of the eBook at doi:10.7551/mitpress/12581.001.0001

- morphisms: from the $\mathbf{C}^{\rightarrow}$-object $A \xrightarrow{f} B$ to the $\mathbf{C}^{\rightarrow}$-object $A' \xrightarrow{f'} B'$, a morphism is a pair $\langle A \xrightarrow{h} A', B \xrightarrow{k} B' \rangle$ of morphisms from $\mathbf{C}$, making the diagram

$$\begin{array}{ccc} A & \xrightarrow{h} & A' \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f'} \\ B & \xrightarrow[k]{} & B' \end{array}$$

  commute (in $\mathbf{C}$).

Composition of arrows is then carried out by placing commutative squares side-by-side, that is, we put

$$\begin{array}{ccccc} A & \xrightarrow{h} & A' & \xrightarrow{l} & A'' \\ {\scriptstyle f}\downarrow & & {\scriptstyle f'}\downarrow & & \downarrow{\scriptstyle f''} \\ B & \xrightarrow[k]{} & B' & \xrightarrow[m]{} & B'' \end{array}$$

so that $\langle l, m \rangle \circ \langle h, k \rangle = \langle l \circ h, m \circ k \rangle$. The identity arrow for an object $A \xrightarrow{f} B$ is given by the pair $\langle \mathrm{id}_A, \mathrm{id}_B \rangle$.

With the arrow category, we are seeing *all* the arrows of the old category as our objects in the new category. The next construction instead looks at just *some* of the old arrows, where we restrict attention to arrows that have fixed domain (source) or codomain (target).

Definition 21 Given a category $\mathbf{C}$, and an object A of $\mathbf{C}$, we can form the two categories called the *slice* and *co-slice* categories, respectively denoted

$$(\mathbf{C} \downarrow A) \qquad\qquad (A \downarrow \mathbf{C}),$$

also called the category of

objects over A objects under A,

respectively.[21] The objects of the new category are given by

arrows to A arrows from A.

In other words, objects of the slice category are given by all pairs (B,f), where B is an object of $\mathbf{C}$ and $f : A \to B$ an arrow of $\mathbf{C}$, and of the co-slice category by all pairs (B,f) such that $f : B \to A$ is an arrow of $\mathbf{C}$.

Morphisms in the new category are given by $h : (B,f) \to (B',f')$, where this is an arrow $h : B \to B'$ of $\mathbf{C}$ for which the respective triangles

$$\begin{array}{ccc} B & \xrightarrow{h} & B' \\ & {\scriptstyle f}\searrow \quad \swarrow{\scriptstyle f'} & \\ & A & \end{array} \qquad\qquad \begin{array}{ccc} & A & \\ & {\scriptstyle f}\swarrow \quad \searrow{\scriptstyle f'} & \\ B & \xrightarrow[h]{} & B' \end{array}$$

commute in the sense that, for instance, for the triangle on the left, $f' \circ h = f$.

21. These categories are also particular cases of a more general construction, known as *comma categories*. It is not uncommon to see the slice category of objects over $A \in \mathrm{Ob}(\mathbf{C})$ referred to as $\mathbf{C}/A$, and the co-slice category of objects under A referred to as $A/\mathbf{C}$.

Composition in $(\mathbf{C} \downarrow A)$ and $(A \downarrow \mathbf{C})$ is then given by composition in $\mathbf{C}$ of the base arrows h of such triangles.

Categories of this type play an important role in advancing some of the general theory, in addition to being of some intrinsic interest. For now, the slice category of *objects over A* might be thought of as giving something like a view of the category *seen within the context of A* (and the corresponding dual statement for the category of objects under A).

Finally, we define the following notion of a *subcategory*.

Definition 22 A *subcategory* $\mathbf{D}$ of a category $\mathbf{C}$ is got by restricting to a subcollection of the collection of objects of $\mathbf{C}$ (i.e., every $\mathbf{D}$-object is a $\mathbf{C}$-object), and a subcollection of the collection of morphisms of $\mathbf{C}$ (i.e., if A and B are any two $\mathbf{D}$-objects, then all the $\mathbf{D}$-arrows $A \to B$ are present in $\mathbf{C}$), where we further require that

- if the morphism $f : A \to B$ is in $\mathbf{D}$, then A and B are in $\mathbf{D}$ as well;
- if A is in $\mathbf{D}$, then so too is the identity morphism id_A;
- if $f : A \to B$ and $g : B \to C$ are in $\mathbf{D}$, then so too is the composite $g \circ f : A \to C$.

Moreover, we can also define the following:

Definition 23 Let $\mathbf{D}$ be a subcategory of $\mathbf{C}$. Then we say that $\mathbf{D}$ is a *full subcategory* of $\mathbf{C}$ when $\mathbf{C}$ has no arrows $A \to B$ other than the ones already in $\mathbf{D}$, that is, for any $\mathbf{D}$-objects A and B,

$$\mathrm{Hom}_{\mathbf{D}}(A, B) = \mathrm{Hom}_{\mathbf{C}}(A, B).$$

Example 24 The category **FinSet** of finite sets—the category whose objects are all finite sets and whose morphisms are all the functions between them—is a subcategory of **Set**. In fact, it is a *full* subcategory.

The category of abelian groups **Ab** is a (full) subcategory of the category of groups **Group**. $\mathbf{Mod}_R$ is a subcategory of $\mathbf{ModAff}_R$.

If $\mathbf{C}$ is the category that has for objects those parts of $\mathbb{R}^n$ that are open—we will have more to say on this in chapter 4—and for morphisms those mappings between objects that are continuous, then a subcategory $\mathbf{D}$ of $\mathbf{C}$ is formed by restricting to mappings that have a derivative, where a rule of basic calculus shows that $\mathbf{D}$ has composition. A further subcategory of $\mathbf{D}$ could be got by further restricting to those mappings that have *all* derivatives (i.e., the *smooth* ones). There are many other important examples of subcategories that we will encounter throughout this book.

There are a number of other useful things one can do with categories, not to mention the important things one can do and find within categories. Discussion of such matters is taken up in the next chapters, and left to emerge organically throughout the book.

The real power of category theory, however, only really comes into its own once it is realized how, by putting everything on the same "plane," we can consider principled relations *between* categories. This is what we discuss in the next chapter.

1.5 Aside on "No Objects"

Box 1.1
No Objects

It is entirely common to name categories after their objects, and in most cases entirely natural to present categories in a "two-sorted" manner, with the two sorts *objects* and *morphisms*. In a given mathematical context, we are often already very familiar with the objects, and comfortable with seeing the relevant structure-preserving morphisms as entities that sit on top of, or are somehow secondary to, the objects. More generally, as human beings, we seem especially ready to divide up the world into objects, on the one hand, and processes or connections between those objects, on the other, where we take the latter to be somehow parasitic on the more primitive objects.

But as natural as this approach may seem, the objects of a category are in fact in bijective correspondence with (i.e., equivalent to) the identity morphisms—which, on account of one of the axioms, are uniquely determined by how they act as two-sided identities for composition. As such, it is really just the *algebra of morphisms* (without objects) that determines a category. Guided by this, one can give alternative definitions of a category that use only morphisms. The following presents such a single-sorted or "no objects" version of the definition of a category.

Definition 25 (*Category definition again ["no objects" version]*) A *category* (single-sorted) is a collection C, the elements or "individuals" of which are called *morphisms*, together with two endofunctions $s, t: C \to C$ (think "source" and "target") on C and a partial function $\circ: C \times C \to C$, where these satisfy the following axioms:

1. $x \circ y$ is defined iff $s(x) = t(y)$;
2. $s(s(x)) = s(x) = t(s(x))$ and $t(t(x)) = t(x) = s(t(x))$ (so s and t are *idempotent* endofunctions on C with the same image);
3. if $x \circ y$ is defined, then $s(x \circ y) = s(y)$ and $t(x \circ y) = t(x)$;
4. $(x \circ y) \circ z = x \circ (y \circ z)$ (whenever either is defined);
5. $x \circ s(x) = x$ and $t(x) \circ x = x$.

Notice how the elements of the shared image of s and t—that is, the x such that $s(x) = x$ (equivalently, $t(x) = x$)—are the *identities* (or what we would normally construe as the *objects*).

It is likely that the "punch" of this definition is lost on a reader seeing it for the first time. The important thing to realize is that behind this presentation is the idea that each object in the usual definition of a category can in fact be *identified* with its identity morphism, allowing us to realize an arrows-only (or "object-free") definition of a category.

Moreover, referring back to example 9, it is in the context of such an arrows-only version that we can even more easily see how monoids are just one-object categories—which really matters because it ultimately lets us better appreciate how categories in general are just many-object monoids. From a given monoid, we would obtain a category by defining $s(x) = t(x) = e$, where e is the monoid constant (identity) element. Going the other way, given a (nonempty) category satisfying any of

- $s(x) = s(y)$,
- $s(x) = t(y)$, or
- $t(x) = t(y)$,

we can define e as the (unique) identity morphism, and thus obtain a monoid. Note, by the way, how this says that s is a constant function (and thus, so is t, and they are in fact equal).

In this way, single-sorted categories can be seen to emerge via what is sometimes called "*oid*ification" (in this case, of monoids), where this describes a general twofold process whereby

1. some construction is realized as equivalent to a certain category with a *single* object; and then
2. the construction is generalized ("oidified") by moving to a further instantiation or version of that same category type that now has more than one object.

As the nLab highlights, in terms of nomenclature, categories give a pretty notable exception to this general rule of appending "oid" to a concept as we move to its many-object version, and perhaps we should all be greatly relieved that enough people did not succumb to the temptation to replace the term *category* with what this process suggests we should call such many-object monoids: a *monoidoid*!

We will see more examples of this process later on. For now, let us remark briefly on the broader significance of this object-free perspective. Consider how, in the context of graphs and graph theory, the novice will likely see arcs (arrows) as secondary to vertices (objects), for the arcs are frequently construed as just pairs of vertices. It also seems plausible that "psychologically" it is somehow more natural or easier for many of us to begin with objects (as the irreducible "simples") and then move on to *relations* between those objects. But in more general treatments of graphs, dealing with directed multigraphs or quivers for instance, one begins to appreciate that this proclivity really gets things backward: in fact, in more general settings, arcs are more naturally seen as primary and vertices can be seen as degenerate sorts of arcs, or as equivalence classes of arcs under the relations "has the same source (target) as."

In a similar fashion, one might argue that our default object-oriented mindsets can get things backward, in terms of what is really fundamental conceptually. It is often said in category theory that "what matters are the arrows/relations, not objects"—which is the substance of the observation that it is the algebra of morphisms that really determines a category, but it goes far beyond this as well. This is a very powerful idea, one that seems to permeate many aspects of category theory, and it even resurfaces in a particularly poignant way with one of the key results in category theory (the Yoneda results, covered in chapter 6). The object-free definition of a category given above is not typically the one seen in an introduction to categories, perhaps because it seems to complicate the presentation of many classical examples of categories, whose presentation is comparatively more straightforward using the standard two-sorted definition of a category. However, the object-free approach is arguably even more fundamental conceptually, and well attuned to the core philosophy of much of the categorical approach—which insists, in a number of ways, that what really matters is how objects and structures interact or relate—so it is worthwhile to at least be familiar with the existence of such a definition.

Sheaf Theory through Examples

By: Daniel Rosiak

Citation:

Sheaf Theory through Examples
By: Daniel Rosiak
DOI: 10.7551/mitpress/12581.001.0001
ISBN (electronic): 9780262370424
Publisher: The MIT Press
Published: 2022

The open access edition of this book was made possible by generous funding and support from Arcadia – a charitable fund of Lisbet Rausing and Peter Baldwin, and MIT Press Direct to Open

The open access edition of this book was made possible by generous funding from Arcadia—a charitable fund of Lisbet Rausing and Peter Baldwin.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in LaTeX by the author.

Library of Congress Cataloging-in-Publication Data

Names: Rosiak, Daniel, author.
Title: Sheaf theory through examples / Daniel Rosiak.
Description: Cambridge, Massachusetts : The MIT Press, [2022] | Includes bibliographical references and index.
Identifiers: LCCN 2021058949 | ISBN 9780262542159 (paperback)
Subjects: LCSH: Sheaf theory.
Classification: LCC QA612.36 .R67 2022 | DDC 514/.224–dc23/eng20220521
LC record available at https://lccn.loc.gov/2021058949