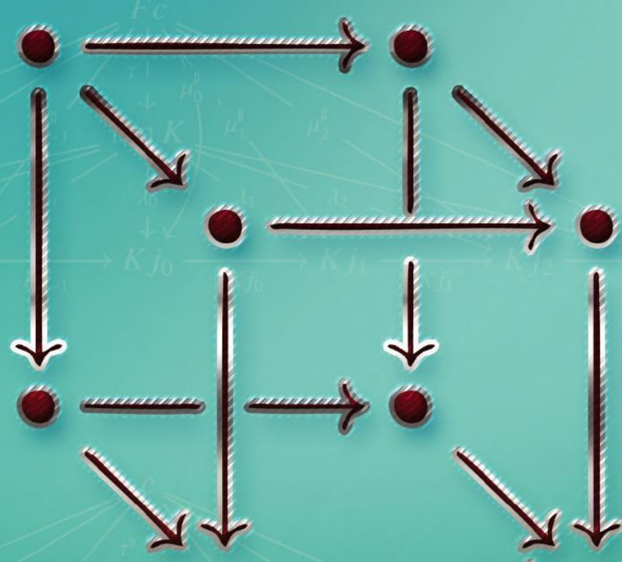
AURORA
DOVER MODERN MATH ORIGINALS

CATEGORY THEORY IN CONTEXT



Emily Riehl

CATEGORY THEORY IN CONTEXT

Emily Riehl

DOVER PUBLICATIONS, INC., Mineola, New York

Chapter 6 is adapted with permission from Chapter 1 of
Categorical Homotopy Theory by Emily Riehl, Cambridge University Press.

Copyright

Bibliographical Note

Category Theory in Context is a new work,
first published by Dover Publications, Inc., in 2016.

International Standard Book Number

ISBN-13: 978-0-486-80903-8
ISBN-10: 0-486-80903-X

Manufactured in the United States by RR Donnelley
80903X01 2016
www.doverpublications.com

Contents

PREFACE

Atiyah described mathematics as the "science of analogy." In this vein, the purview of category theory is *mathematical analogy*. Category theory provides a cross-disciplinary language for mathematics designed to delineate general phenomena, which enables the transfer of ideas from one area of study to another. The category-theoretic perspective can function as a simplifying[1] abstraction, isolating propositions that hold for formal reasons from those whose proofs require techniques particular to a given mathematical discipline.[2]

A subtle shift in perspective enables mathematical content to be described in language that is relatively indifferent to the variety of objects being considered. Rather than characterize the objects directly, the categorical approach emphasizes the transformations between objects of the same general type. A fundamental lemma in category theory implies that any mathematical object can be characterized by its *universal property*—loosely by a representation of the morphisms to or from other objects of a similar form. For example, tensor products, "free" constructions, and localizations are characterized by universal properties in appropriate *categories*, or mathematical contexts. A universal property typically expresses one of the mathematical roles played by the object in question. For instance, one universal property associated to the unit interval identifies self-homeomorphisms of this space with re-parameterizations of paths. Another highlights the operation of gluing two intervals end to end to obtain a new interval, the construction used to define composition of paths.

Certain classes of universal properties define blueprints which specify how a new object may be built out of a collection of existing ones. A great variety of mathematical constructions fit into this paradigm: products, kernels, completions, free products, "gluing" constructions, and quotients are all special cases of the general category-theoretic notion of *limits* or *colimits*, a characterization that makes it easy to define transformations to or from the objects so-defined. The input data for these constructions are *commutative diagrams*, which are themselves a vehicle for mathematical definitions, e.g., of rings or algebras, representations of a group, or chain complexes.

Important technical differences between particular varieties of mathematical objects can be described by the distinctive properties of their categories: that rings have all limits and colimits while fields have few, that a continuous bijection defines an isomorphism of compact Hausdorff spaces but not of generic topological spaces. Constructions that convert mathematical objects of one type into objects of another type often define transformations between categories, called *functors*. Many of the basic objects of study in modern algebraic topology and algebraic geometry involve functors and would be impossible to define without category-theoretic language.

[1]In his mathematical notebooks, Hilbert formulated a "24th problem" (inspired by his work on syzygies) to develop a criterion of simplicity for evaluating competing proofs of the same result [**TW02**].

[2]For example, the standard properties of induced representations (Frobenius reciprocity, transitivity of induction, even the explicit formula) are true of any construction defined as a *left Kan extension*; character tables, however, are non-formal.

Category theory also contributes new proof techniques, such as *diagram chasing* or arguments by duality; Steenrod called these methods "abstract nonsense."[3] The aim of this text is to introduce the language, philosophy, and basic theorems of category theory. A complementary objective is to put this theory into practice: studying functoriality in algebraic topology, naturality in group theory, and universal properties in algebra.

Practitioners often assert that the hard part of category theory is to state the correct definitions. Once these are established and the categorical style of argument is sufficiently internalized, proving the theorems tends to be relatively easy.[4] Indeed, the proofs of several propositions appearing in this text are left as exercises, with confidence that the reader will eventually find it more efficient to supply their own arguments than to read the author's.[5] The relative simplicity of the proofs of major theorems occasionally leads detractors to assert that there are no theorems in category theory. This is not at all the case! Counterexamples abound in the text that follows. A short list of further significant theorems, beyond the scope of a first course but not too far to be out of the reach of comprehension, appears as an epilogue.

Sample corollaries

It is difficult to preview the main theorems in category theory before developing fluency in the language needed to state them. (A reader possessing such fluency might wish to glance ahead to §E.1.) Instead, here are a few corollaries, results in other areas of mathematics that follow trivially as special cases of general categorical results that are proven in this text.

As an application of the theory of equivalence between categories:

COROLLARY 1.5.13. *In a path-connected space, any choice of basepoint yields an isomorphic fundamental group.*

A fundamental lemma in category theory has the following two results as corollaries:

COROLLARY 2.2.9. *Every row operation on matrices with n rows is defined by left multiplication by some $n \times n$ matrix, namely the matrix obtained by performing the row operation on the identity matrix.*

COROLLARY 2.2.10. *Any group is isomorphic to a subgroup of a permutation group.*

A special case of a general result involving the interchange of limits and colimits is:

COROLLARY 3.8.4. *For any pair of sets X and Y and any function $f\colon X \times Y \to \mathbb{R}$*

$$\sup_{x\in X} \inf_{y\in Y} f(x,y) \leq \inf_{y\in Y} \sup_{x\in X} f(x,y)$$

---

[3] Lang's *Algebra* [**Lan02**, p. 759] supports the general consensus that this was not intended as an epithet:

> In the forties and fifties (mostly in the works of Cartan, Eilenberg, MacLane, and Steenrod, see [**CE56**]), it was realized that there was a systematic way of developing certain relations of linear algebra, depending only on fairly general constructions which were mostly arrow-theoretic, and were affectionately called **abstract nonsense** by Steenrod.

[4] A famous exercise in Lang's *Algebra* asks the reader to "Take any book on homological algebra, and prove all the theorems without looking at the proofs given in that book" [**Lan84**, p. 175]. Homological algebra is the subject whose development induced Eilenberg and Mac Lane to introduce the general notions of category, functor, and natural transformation.

[5] In the first iteration of the course that inspired the writing of these lecture notes, the proofs of several major theorems were also initially left to the exercises, with a type-written version appearing only after the problem set was due.

whenever these infima and suprema exist.

The following five results illustrate a few of the many corollaries of a common theorem, which describes one consequence of a type of "duality" enjoyed by certain pairs of mathematical constructions:

COROLLARY 4.5.4. *For any function $f\colon A \to B$, the inverse image function $f^{-1}\colon PB \to PA$ between the power sets of A and B preserves both unions and intersections, while the direct image function $f_*\colon PA \to PB$ only preserves unions.*

COROLLARY 4.5.5. *For any vector spaces U, V, W,*

$$U \otimes (V \oplus W) \cong (U \otimes V) \oplus (U \otimes W).$$

COROLLARY 4.5.6. *For any cardinals α, β, γ, cardinal arithmetic satisfies the laws:*

$$\alpha \times (\beta + \gamma) = (\alpha \times \beta) + (\alpha \times \gamma) \qquad (\beta \times \gamma)^\alpha = \beta^\alpha \times \gamma^\alpha \qquad \alpha^{\beta+\gamma} = \alpha^\beta \times \alpha^\gamma .$$

COROLLARY 4.5.7. *The free group on the set $X \sqcup Y$ is the free product of the free groups on the sets X and Y.*

COROLLARY 4.5.8. *For any R–S bimodule M, the tensor product $M \otimes_S -$ is right exact.*

Finally, a general theorem that recognizes categories whose objects bear some sort of "algebraic" structure has a number of consequences, including:

COROLLARY 5.6.2. *Any bijective continuous function between compact Hausdorff spaces is a homeomorphism.*

This is not to say that category theory necessarily provides a more efficient proof of these results. In many cases, the proof that general consensus designates the "most elegant" reflects the categorical argument. The point is that the category-theoretic perspective allows for an efficient packaging of general arguments that can be used over and over again and eliminates contextual details that can safely be ignored. For instance, our proof that the tensor product commutes with the direct sum of vector spaces will not make use of any bases, but appeals instead to the universal properties of the tensor product and direct sum constructions.

A tour of basic categorical notions

> . . . the science of mathematics exemplifies the interdependence of its parts.
>
> Saunders Mac Lane, "Topology and logic as a source of algebra" [**ML76**]

A *category* is a context for the study of a particular class of mathematical objects. Importantly, a category is not simply a type signature, it has both "nouns" and "verbs," containing specified collections of objects and transformations, called *morphisms*,[6] between them. Groups, modules, topological spaces, measure spaces, ordinals, and so forth form categories, but these classifications are not the main point. Rather, the action of packaging each variety of objects into a category shifts one's perspective from the particularities of each mathematical sub-discipline to potential commonalities between them. A basic observation along these lines is that there is a single categorical definition of *isomorphism*

[6]The term "morphism" is derived from *homomorphism*, the name given in algebra to a structure-preserving function. Synonyms include "arrow" (because of the notation "→") and "map" (adopting the standard mathematical colloquialism).

that specializes to define isomorphisms of groups, homeomorphisms of spaces, order isomorphisms of posets, and even isomorphisms between categories (see Definition 1.1.9).

Mathematics is full of constructions that translate mathematical objects of one kind into objects of another kind. A construction that converts the objects in one category into objects in another category is *functorial* if it can be extended to a mapping on morphisms in such a way that composites and identity morphisms are preserved. Such constructions define morphisms between categories, called *functors*. Functoriality is often a key property: for instance, the chain rule from multivariable calculus expresses the functoriality of the derivative (see Example 1.3.2(x)). In contrast with earlier numerical invariants in topology, functorial invariants (the fundamental group, homology) tend both to be more easily computable and also provide more precise information. While the Euler characteristic can distinguish between the closed unit disk and its boundary circle, an easy proof by contradiction involving the functoriality of their fundamental groups proves that any continuous endomorphism of the disk must have a fixed point (see Theorem 1.3.3).

On occasion, functoriality is achieved by *categorifying* an existing mathematical construction. "Categorification" refers to the process of turning sets into categories by adding morphisms, whose introduction typically demands a re-interpretation of the elements of the sets as related mathematical objects. A celebrated knot invariant called the Jones polynomial must vanish for any knot diagram that presents the unknot, but its *categorification*, a functor[7] called *Khovanov homology*, detects the unknot in the sense that any knot diagram whose Khovanov homology vanishes must represent the unknot. Khovanov homology converts an oriented link diagram into a chain complex whose graded Euler characteristic is the Jones polynomial.

A functor may describe an *equivalence of categories*, in which case the objects in one category can be translated into and reconstructed from the objects of another. For instance, there is an equivalence between the category of finite-dimensional vector spaces and linear maps and a category whose objects are natural numbers and whose morphisms are matrices (see Corollary 1.5.11). This process of conversion from college linear algebra to high school linear algebra defines an equivalence of categories; eigenvalues and eigenvectors can be developed for matrices or for linear transformations, it makes no difference.

Treating categories as mathematical objects in and of themselves, a basic observation is that the process of formally "turning around all the arrows" in a category produces another category. In particular, any theorem proven for all categories also applies to these *opposite categories*; the re-interpretation of the result in the opposite of an opposite category yields the statement of the *dual theorem*. Categorical constructions also admit duals: for instance, in Zermelo–Fraenkel set theory, a function $f\colon X \to Y$ is defined via its *graph*, a subset of $X \times Y$ isomorphic to X. The dual presentation represents a function via its *cograph*, a Y-indexed partition of $X \sqcup Y$. Categorically-proven properties of the graph representation will dualize to describe properties of the cograph representation.

Categories and functors were introduced by Eilenberg and Mac Lane with the goal of giving precise meaning to the colloquial usage of "natural" to describe families of isomorphisms. For example, for any triple of $\Bbbk$-vector spaces U, V, W, there is an isomorphism

$$\mathsf{Vect}_{\Bbbk}(U \otimes_{\Bbbk} V, W) \cong \mathsf{Vect}_{\Bbbk}(U, \mathrm{Hom}(V, W)) \tag{0.0.1}$$

between the set of linear maps $U \otimes_{\Bbbk} V \to W$ and the set of linear maps from U to the vector space $\mathrm{Hom}(V, W)$ of linear maps from V to W. This isomorphism is natural in all three

[7]Morally, one could argue that functoriality is the main innovation in this construction, but making this functoriality precise is somewhat subtle [**CMW09**].

variables, meaning it defines an isomorphism not simply between these sets of maps but between appropriate set-valued functors of U, V, and W. Chapter 1 introduces the basic language of category theory, defining categories, functors, natural transformations, and introducing the principle of duality, equivalences of categories, and the method of proof by diagram chasing.

In fact, the isomorphism (0.0.1) *defines* the vector space $U \otimes_{\Bbbk} V$ by declaring that linear maps $U \otimes_{\Bbbk} V \to W$ correspond to linear maps $U \to \mathrm{Hom}(V, W)$, i.e., to bilinear maps $U \times V \to W$. This definition is sufficiently robust that important properties of the tensor product—for instance its symmetry and associativity—can be proven without reference to any particular construction (see Proposition 2.3.9 and Exercise 2.3.ii). The advantages of this approach compound as the mathematical objects so-described become more complicated.

In Chapter 2, we study such definitions abstractly. A characterization of the morphisms either to or from a fixed object describes its *universal property*; the cases of "to" or "from" are dual. By the Yoneda lemma—which, despite its innocuous statement, is arguably the most important result in category theory—every object is characterized by either of its universal properties. For example, the Sierpinski space is characterized as a topological space by the property that continuous functions $X \to S$ correspond naturally to open subsets of X. The complete graph on n vertices is characterized by the property that graph homomorphisms $G \to K_n$ correspond to n-colorings of the vertices of the graph G with the property that adjacent vertices are assigned distinct colors. The polynomial ring $\mathbb{Z}[x_1, \ldots, x_n]$ is characterized as a commutative unital ring by the property that ring homomorphisms $\mathbb{Z}[x_1, \ldots, x_n] \to R$ correspond to n-tuples of elements $(r_1, \ldots, r_n) \in R$. Modern algebraic geometry begins from the observation that a commutative ring can be identified with the functor that it represents.

The idea of probing a fixed object using morphisms abutting to it from other objects in the category gives rise to a notion of "generalized elements" (see Remark 3.4.15). The elements of a set A are in bijection with functions $* \to A$ with domain a singleton set; a *generalized element* of A is a morphism $X \to A$ with generic domain. In the category of directed graphs, a parallel pair of graph homomorphisms $\phi, \psi \colon A \rightrightarrows B$ can be distinguished by considering generalized elements of A whose domain is the free-living vertex or the free-living directed edge.[8] A related idea leads to the representation of a topological space via its singular complex.

The Yoneda lemma implies that a general mathematical object can be *represented* as a functor valued in the category of sets. A related classical antecedent is a result that comforted those who were troubled by the abstract definition of a group: namely that any group is isomorphic to a subgroup of a permutation group (see Corollary 2.2.10). A deep consequence of these functorial representations is that proofs that general categorically-described constructions are isomorphic reduce to the construction of a bijection between their set-theoretical analogs (for instance, see the proof of Theorem 3.4.12).

Chapter 3 studies a special case of definitions by universal properties, which come in two dual forms, referred to as *limits* and *colimits*. For example, aggregating the data of the cyclic p-groups $\mathbb{Z}/p^n$ and homomorphisms between them, one can build more complicated abelian groups. Limit constructions build new objects in a category by "imposing equations" on existing ones. For instance, the diagram of quotient homomorphisms

$$\cdots \twoheadrightarrow \mathbb{Z}/p^n \twoheadrightarrow \cdots \twoheadrightarrow \mathbb{Z}/p^3 \twoheadrightarrow \mathbb{Z}/p^2 \twoheadrightarrow \mathbb{Z}/p$$

[8]The incidence relation in the graph A can be recovered by also considering the homomorphisms between these graphs.

has a limit, namely the group $\mathbb{Z}_p$ of *p-adic integers*: its elements can be understood as tuples of elements $(a_n \in \mathbb{Z}/p^n)_{n\in\omega}$ that are compatible modulo congruence. There is a categorical explanation for the fact that $\mathbb{Z}_p$ is a commutative ring and not merely an abelian group: each of these quotient maps is a ring homomorphism, and so this diagram and also its limit lifts to the category of rings.[9]

By contrast, colimit constructions build new objects by "gluing together" existing ones. The colimit of the sequence of inclusions

$$\mathbb{Z}/p \hookrightarrow \mathbb{Z}/p^2 \hookrightarrow \mathbb{Z}/p^3 \hookrightarrow \cdots \hookrightarrow \mathbb{Z}/p^n \hookrightarrow \cdots$$

is the *Prüfer p-group* $\mathbb{Z}[\frac{1}{p}]/\mathbb{Z}$, an abelian group which can be presented via generators and relations as

$$\mathbb{Z}[\tfrac{1}{p}]/\mathbb{Z} := \left\langle g_1, g_2, \ldots \;\middle|\; pg_1 = 0, pg_2 = g_1, pg_3 = g_2, \ldots \right\rangle. \tag{0.0.2}$$

The inclusion maps are not ring homomorphisms (failing to preserve the multiplicative identity) and indeed it turns out that the Prüfer p-group does not admit any non-trivial multiplicative structure.

Limits and colimits are accompanied by universal properties that generalize familiar universal properties in analysis. A poset $(A, \leq)$ may be regarded as a category whose objects are the elements $a \in A$ and in which a morphism $a \to a'$ is present if and only if $a \leq a'$. The *supremum* of a collection of elements $\{a_i\}_{i\in I}$, an example of a colimit in the category $(A, \leq)$, has a universal property: namely to prove that

$$\sup_{i\in I} a_i \leq a$$

is equivalent to proving that $a_i \leq a$ for all $i \in I$. The universal property of a generic colimit is a generalization of this, where the collection of morphisms $(a_i \to a)_{i\in I}$ is regarded as data, called a *cone* under the diagram, rather than simply a family of conditions. Limits have a dual universal property that specializes to the universal property of the infimum of a collection of elements in a poset.

Chapter 4 studies a generalization of the notion of equivalence of categories, in which a pair of categories are connected by a pair of opposite-pointing translation functors called an *adjunction*. An adjunction expresses a kind of "duality" between a pair of functors, first recognized in the case of the construction of the tensor product and hom functors for abelian groups (see Example 4.3.11). Any adjunction restricts to define an equivalence between certain subcategories, but categories connected by adjunctions need not be equivalent. For instance, there is an adjunction connecting the poset of subsets of $\mathbb{C}^n$ and the poset of subsets of the ring $\mathbb{C}[x_1, \ldots, x_n]$ that restricts to define an equivalence between Zariski closed subsets and radical ideals (see Example 4.3.2). Another adjunction encodes a duality between the constructions of the suspension and of the loop space of a based topological space (see Example 4.3.14).

When a "forgetful" functor admits an adjoint, that adjoint defines a "free" (or, less commonly, the dual "cofree") construction. Such functors define universal solutions to optimization problems, e.g., of adjoining a multiplicative unit to a non-unital ring. The existence of free groups or free rings have implications for the constructions of limits in these categories (namely, Theorem 4.5.2); the dual properties for colimits do not hold because there are no "cofree" groups or rings in general. A category-theoretic re-interpretation of the construction of the Stone–Čech compactification of a topological space defines a

[9]The lifting of the limit is considerably more subtle than the lifting of the diagram. Results of this nature motivate Chapter 5.

left adjoint to any limit-preserving functor between any pair of categories with similar set-theoretic properties (see Theorem 4.6.10 and Example 4.6.12).

Many familiar varieties of "algebraic" objects—such as groups, rings, modules, pointed sets, or sets acted on by a group—admit a "free–forgetful" adjunction with the category of sets. A special property of these adjoint functors explains many of the common features of the categories of algebras that are presented in this manner. Chapter 5 introduces the categorical approach to universal algebra, which distinguishes the categories of rings, compact Hausdorff spaces, and lattices from the set-theoretically similar categories of fields, generic topological spaces, and posets. The former categories, but not the latter, are *categories of algebras* over the category of sets.

The notion of *algebra* is given a precise meaning in relation to a *monad*, an endofunctor that provides a syntactic encoding of algebraic structure that may be borne by objects in the category on which it acts. Monads are also used to construct categories whose morphisms are partially-defined or non-deterministic functions, such as Markov kernels (see Example 5.2.10), and are separately of interest in computer science. A key result in categorical universal algebra is a vast generalization of the notion of a presentation of a group via generators and relations, such as in (0.0.2), which demonstrates that an algebra of any variety can be presented canonically as a *coequalizer*[10] of a pair of maps from a free algebra on the "relations" to a free algebra on the "generators."

The concluding Chapter 6 introduces a general formalism that can be used to redefine all of the basic categorical notions introduced in the first part of the text. Special cases of *Kan extensions* define representable functors, limits, colimits, adjoint functors, and monads, and their study leads to a generalization of, as well as a dualization of, the Yoneda lemma. In the most important cases, a Kan extension can be computed by a particular formula, which specializes to give the construction of a representation for a group induced from a representation for a subgroup (see Example 6.2.8), to provide a new way to think about the collection of ultrafilters on a set (see Example 6.5.12), and to define an equivalence of categories connecting sheaves on a space with étale spaces over that space (see Exercise 6.5.iii).

A brief detour introduces *derived functors*, which are certain special Kan extensions that are of great importance in homological algebra and algebraic topology. A recent categorical discovery reveals that a common mechanism for constructing "point-set level" derived functors yields total derived functors with superior universal properties (see Propositions 6.4.12 and 6.4.13). A final motivation for the study of Kan extensions reaches beyond the scope of this book. The calculus of Kan extensions facilitates the extension of basic category theory to *enriched*, *internal*, *fibered*, or *higher-dimensional* contexts, which provide natural homes for more sophisticated varieties of mathematical objects whose transformations have some sort of higher-dimensional structure.

Note to the reader

The text that follows is littered with examples drawn from a broad range of mathematical areas. The examples are included for color or historical context but are never essential for understanding the abstract category theory. In principle, one could study category theory immediately after learning some basic set theory and logic, as no other prerequisites are strictly required, but without some level of mathematical maturity it would be difficult to see what the point of it all is. We hope that the majority of examples are comprehensible in outline, even if the details are unfamiliar, but if this is not the case, it is not worth stressing

[10]A coequalizer is a generalization of a cokernel to contexts that may lack a "zero" homomorphism.

over. Inevitably, given the diversity of mathematical tastes and experiences, the examples presented here will seldom be optimized for any particular individual, and indeed, each reader is encouraged to search for their own contexts in which to explore categorical ideas.

Notational conventions

An arrow symbol "$\to$," either in a display or in text, is only ever used to denote a morphism in an appropriate category. In particular, the objects surrounding it necessarily lie in a common category. Double arrows "$\Rightarrow$" are reserved for natural transformations, the notation used to suggest the intuition that these are some variety of "2-dimensional" morphisms. The symbol "$\mapsto$," read as "maps to," appears occasionally when defining a function between sets by specifying its action on particular elements. The symbol "$\rightsquigarrow$" is used in a less technical sense to mean something along the lines of "yields" or "leads to" or "can be used to construct." If the presence of certain morphisms implies the existence of another morphism, the latter is often depicted with a dashed arrow "$\dashrightarrow$" to suggest the correct order of inference.[11]

We use "$\rightrightarrows$" as an abbreviation for a **parallel pair** of morphisms, i.e., for a pair of morphisms with common source and target, and "$\leftrightarrows$" as an abbreviation for an **opposing pair** of morphisms with sources and targets swapped.

Italics are used occasionally for emphasis and to highlight technical terms. Boldface signals that a technical term is being defined by its surrounding text.

The symbol "$=$" is reserved for genuine equality (with "$:=$" used for definitional equality), with "$\cong$" used instead for isomorphism in the appropriate ambient category, by far the more common occurrence.

Acknowledgments

Many of the theorems appearing here are standard fare for a first course on category theory, but the examples are not. Rather than rely solely on my own generative capacity, I consulted a great many people while preparing this text and am grateful for their generosity in sharing ideas.

To begin, I would like to thank the following people who responded to a call for examples on the *n*-Category Café and the categories mailing list: John Baez, Martin Brandenburg, Ronnie Brown, Tyler Bryson, Tim Campion, Yemon Choi, Adrian Clough, Samuel Dean, Josh Drum, David Ellerman, Tom Ellis, Richard Garner, Sameer Gupta, Gejza Jenča, Mark Johnson, Anders Kock, Tom LaGatta, Paul Levy, Fred Linton, Aaron Mazel-Gee, Jesse McKeown, Kimmo Rosenthal, Mike Shulman, Peter Smith, Arnaud Spiwack, Ross Street, John Terilla, Todd Trimble, Mozibur Ullah, Enrico Vitale, David White, Graham White, and Qiaochu Yuan.

In particular, Anders Kock suggested a more general formulation of "the chain rule expresses the functoriality of the derivative" than appears in Example 1.3.2(x). The expression of the fundamental theorem of Galois theory as an isomorphism of categories that appears as Example 1.3.15 is a favorite exercise of Peter May's. Charles Blair suggested Exercise 1.3.iii and a number of expository improvements to the first chapter. I learned about the unnatural isomorphism of Proposition 1.4.4 from Mitya Boyarchenko. Peter Haine suggested Example 1.4.6, expressing the Riesz representation theorem as a natural isomorphism of Banach space-valued functors; Example 3.6.2, constructing the path-components functor, and Example 3.8.6. He also contributed Exercises 1.2.v, 1.5.v, 1.6.iv, 3.1.xii, and 4.5.v and served as my LaTeX consultant.

[11] Readers who dislike this convention can simply connect the dots.

John Baez reminded me that the groupoid of finite sets is a categorification of the natural numbers, providing a suitable framework in which to prove certain basic equations in elementary arithmetic; see Example 1.4.9 and Corollary 4.5.6. Juan Climent Vidal suggested using the axiom of regularity to define the non-trivial part of the equivalence of categories presented in Example 1.5.6 and contributed the equivalence of plane geometries that appears as Exercise 1.5.viii. Samuel Dean suggested Corollary 1.5.13, Fred Linton suggested Corollary 2.2.9, and Ronnie Brown suggested Example 3.5.8. Ralf Meyer suggested Example 2.1.5(vi), Example 5.2.6(ii), Corollary 4.5.8, and a number of exercises including 2.1.ii and 2.4.v, which he used when teaching a similar course. He also pointed me toward a simpler proof of Proposition 6.4.12.

Mozibur Ullah suggested Exercise 3.5.vii. I learned of the non-natural objectwise isomorphism appearing in Example 3.6.5 from Martin Brandenburg who acquired it from Tom Leinster. Martin also suggested the description of the real exponential function as a Kan extension appearing in Example 6.2.7. Andrew Putman pointed out that Lang's *Algebra* constructs the free group on a set using the construction of the General Adjoint Functor Theorem, recorded as Example 4.6.6. Paul Levy suggested using affine spaces to motivate the category of algebras over a monad, as discussed in Section 5.2; a similar example was suggested by Enrico Vitale. Dominic Verity suggested something like Exercise 5.5.vii. Vladimir Sotirov pointed out that the appropriate size hypotheses were missing from the original statement of Theorem 6.3.7 and directed me toward a more elegant proof of Lemma 4.6.5.

Marina Lehner, while writing her undergraduate senior thesis under my direction, showed me that an entirely satisfactory account of Kan extensions can be given without the calculus of ends and coends. I have enjoyed, and this book has been enriched by, several years of impromptu categorical conversations with Omar Antolín Camarena. I am extremely appreciative of the careful readings undertaken by Tobias Barthel, Martin Brandenburg, Benjamin Diamond, Darij Grinberg, Peter Haine, Ralf Meyer, Peter Smith, and Juan Climent Vidal, who each sent detailed lists of corrections and suggestions. I would also like to thank John Grafton and Janet Kopito, the acquisitions and in-house editors at Dover Publications, and David Gargaro for his meticulous copyediting.

I am grateful for the perspicacious comments and questions from those who attended the first iteration of this course at Harvard—Paul Bamberg, Nathan Gupta, Peter Haine, Andrew Liu, Wyatt Mackey, Nat Mayer, Selorm Ohene, Jacob Seidman, Aaron Slipper, Alma Steingart, Nithin Tumma, Jeffrey Yan, Liang Zhang, and Michael Fountaine, who served as the course assistant—and the second iteration at Johns Hopkins—Benjamin Diamond, Nathaniel Filardo, Alex Grounds, Hanveen Koh, Alex Rozenshteyn, Xiyuan Wang, Shengpei Yan, and Zhaoning Yang. I would also like to thank the Departments of Mathematics at both institutions for giving me opportunities to teach this course and my colleagues there, who have created two extremely pleasant working environments. While these notes were being revised, I received financial support from the National Science Foundation Division of Mathematical Sciences DMS-1509016.

My enthusiasm for mathematical writing and patience for editing were inherited from Peter May, my PhD supervisor. I appreciate the indulgence of my collaborators, my friends, and my family while this manuscript was in its final stages of preparation. This book is dedicated to Peter Johnstone and Martin Hyland, whose tutelage I was fortunate to come under in a pivotal year spent at Cambridge supported by the Winston Churchill Foundation of the United States.

CATEGORY THEORY IN CONTEXT

CHAPTER 1

Categories, Functors, Natural Transformations

> Frequently in modern mathematics there occur phenomena of "naturality".
>
> Samuel Eilenberg and Saunders Mac Lane, "Natural isomorphisms in group theory" [**EM42b**]

A **group extension** of an abelian group H by an abelian group G consists of a group E together with an inclusion of $G \hookrightarrow E$ as a normal subgroup and a surjective homomorphism $E \twoheadrightarrow H$ that displays H as the quotient group E/G. This data is typically displayed in a diagram of group homomorphisms:

$$0 \to G \to E \to H \to 0.$$[1]

A pair of group extensions E and E' of G and H are considered to be equivalent whenever there is an isomorphism $E \cong E'$ that *commutes with* the inclusions of G and quotient maps to H, in a sense that is made precise in §1.6. The set of equivalence classes of *abelian* group extensions E of H by G defines an abelian group $\mathrm{Ext}(H, G)$.

In 1941, Saunders Mac Lane gave a lecture at the University of Michigan in which he computed for a prime p that $\mathrm{Ext}(\mathbb{Z}[\frac{1}{p}]/\mathbb{Z}, \mathbb{Z}) \cong \mathbb{Z}_p$, the group of p-adic integers, where $\mathbb{Z}[\frac{1}{p}]/\mathbb{Z}$ is the Prüfer p-group. When he explained this result to Samuel Eilenberg, who had missed the lecture, Eilenberg recognized the calculation as the homology of the 3-sphere complement of the p-adic solenoid, a space formed as the infinite intersection of a sequence of solid tori, each wound around p times inside the preceding torus. In teasing apart this connection, the pair of them discovered what is now known as the **universal coefficient theorem** in algebraic topology, which relates the *homology* H_* and *cohomology groups* H^* associated to a space X via a group extension [**ML05**]:

$$0 \to \mathrm{Ext}(H_{n-1}(X), G) \to H^n(X, G) \to \mathrm{Hom}(H_n(X), G) \to 0\,. \tag{1.0.1}$$

To obtain a more general form of the universal coefficient theorem, Eilenberg and Mac Lane needed to show that certain isomorphisms of abelian groups expressed by this group extension extend to spaces constructed via direct or inverse limits. And indeed this is the case, precisely because the homomorphisms in the diagram (1.0.1) are *natural* with respect to continuous maps between topological spaces.

The adjective "natural" had been used colloquially by mathematicians to mean "defined without arbitrary choices." For instance, to define an isomorphism between a finite-dimensional vector space V and its **dual**, the vector space of linear maps from V to the

[1] The zeros appearing on the ends provide no additional data. Instead, the first zero implicitly asserts that the map $G \to E$ is an inclusion and the second that the map $E \to H$ is a surjection. More precisely, the displayed sequence of group homomorphisms is **exact**, meaning that the kernel of each homomorphism equals the image of the preceding homomorphism.

ground field $\mathbb{k}$, requires a choice of basis. However, there is an isomorphism between V and its double dual that requires no choice of basis; the latter, but not the former, is *natural*.

To give a rigorous proof that their particular family of group isomorphisms extended to inverse and direct limits, Eilenberg and Mac Lane sought to give a mathematically precise definition of the informal concept of "naturality." To that end, they introduced the notion of a *natural transformation*, a parallel collection of homomorphisms between abelian groups in this instance. To characterize the source and target of a natural transformation, they introduced the notion of a *functor*.[2] And to define the source and target of a functor in the greatest generality, they introduced the concept of a *category*. This work, described in "The general theory of natural equivalences" [**EM45**], published in 1945, marked the birth of category theory.

While categories and functors were first conceived as auxiliary notions, needed to give a precise meaning to the concept of naturality, they have grown into interesting and important concepts in their own right. Categories suggest a particular perspective to be used in the study of mathematical objects that pays greater attention to the maps between them. Functors, which translate mathematical objects of one type into objects of another, have a more immediate utility. For instance, the Brouwer fixed point theorem translates a seemingly intractable problem in topology to a trivial one ($0 \neq 1$) in algebra. It is to these topics that we now turn.

Categories are introduced in §1.1 in two guises: firstly as universes categorizing mathematical objects and secondly as mathematical objects in their own right. The first perspective is used, for instance, to define a general notion of *isomorphism* that can be specialized to mathematical objects of every conceivable variety. The second perspective leads to the observation that the axioms defining a category are self-dual.[3] Thus, as explored in §1.2, for any proof of a theorem about all categories from these axioms, there is a dual proof of the dual theorem obtained by a syntactic process that is interpreted as "turning around all the arrows."

Functors and natural transformations are introduced in §1.3 and §1.4 with examples intended to shed light on the linguistic and practical utility of these concepts. The category-theoretic notions of *isomorphism*, *monomorphism*, and *epimorphism* are invariant under certain classes of functors, including in particular the *equivalences of categories*, introduced in §1.5. At a high level, an equivalence of categories provides a precise expression of the intuition that mathematical objects of one type are "the same as" objects of another variety: an equivalence between the category of matrices and the category of finite-dimensional vector spaces equates high school and college linear algebra.

In addition to providing a new language to describe emerging mathematical phenomena, category theory also introduced a new proof technique: that of the diagram chase. The introduction to the influential book [**ES52**] presents *commutative diagrams* as one of the "new techniques of proof" appropriate for their axiomatic treatment of homology theory. The technique of diagram chasing is introduced in §1.6 and applied in §1.7 to construct new natural transformations as *horizontal* or *vertical composites* of given ones.

[2] A brief account of functors and natural isomorphisms in group theory appeared in a 1942 paper [**EM42b**].

[3] As is the case for the duality in projective plane geometry, this duality can be formulated precisely as a feature of the first-order theories that axiomatize these structures.

1.1. Abstract and concrete categories

> It frames a possible template for any mathematical theory: the theory should have *nouns* and *verbs*, i.e., objects, and morphisms, and there should be an explicit notion of composition related to the morphisms; the theory should, in brief, be packaged by a category.
>
> Barry Mazur, "When is one thing equal to some other thing?" [**Maz08**]

Definition 1.1.1. A **category** consists of

- a collection of **objects** $X, Y, Z, \ldots$
- a collection of **morphisms** $f, g, h, \ldots$

so that:

- Each morphism has specified **domain** and **codomain** objects; the notation $f \colon X \to Y$ signifies that f is a morphism with domain X and codomain Y.
- Each object has a designated **identity morphism** $1_X \colon X \to X$.
- For any pair of morphisms f, g with the codomain of f equal to the domain of g, there exists a specified **composite morphism**[4] gf whose domain is equal to the domain of f and whose codomain is equal to the codomain of g, i.e.,:

$$f \colon X \to Y, \quad g \colon Y \to Z \qquad \rightsquigarrow \qquad gf \colon X \to Z\,.$$

This data is subject to the following two axioms:

- For any $f \colon X \to Y$, the composites $1_Y f$ and $f 1_X$ are both equal to f.
- For any composable triple of morphisms f, g, h, the composites $h(gf)$ and $(hg)f$ are equal and henceforth denoted by hgf.

$$f \colon X \to Y, \quad g \colon Y \to Z, \quad h \colon Z \to W \qquad \rightsquigarrow \qquad hgf \colon X \to W\,.$$

That is, the composition law is associative and unital with the identity morphisms serving as two-sided identities.

Remark 1.1.2. The objects of a category are in bijective correspondence with the identity morphisms, which are uniquely determined by the property that they serve as two-sided identities for composition. Thus, one can define a category to be a collection of morphisms with a partially-defined composition operation that has certain special morphisms, which are used to recognize composable pairs and which serve as two-sided identities; see [**Ehr65**, §I.1] or [**FS90**, §1.1]. But in practice it is not so hard to specify both the objects and the morphisms and this is what we shall do.

It is traditional to name a category after its objects; typically, the preferred choice of accompanying structure-preserving morphisms is clear. However, this practice is somewhat contrary to the basic philosophy of category theory: that mathematical objects should always be considered in tandem with the morphisms between them. By Remark 1.1.2, the algebra of morphisms determines the category, so of the two, the objects and morphisms, the morphisms take primacy.

Example 1.1.3. Many familiar varieties of mathematical objects assemble into a category.

[4]The composite may be written less concisely as $g \cdot f$ when this adds typographical clarity.

(i) Set has sets as its objects and functions, with specified domain and codomain,[5] as its morphisms.
(ii) Top has topological spaces as its objects and continuous functions as its morphisms.
(iii) Set_* and Top_* have sets or spaces with a specified basepoint[6] as objects and basepoint-preserving (continuous) functions as morphisms.
(iv) Group has groups as objects and group homomorphisms as morphisms. This example lent the general term "morphisms" to the data of an abstract category. The categories Ring of associative and unital rings and ring homomorphisms and Field of fields and field homomorphisms are defined similarly.
(v) For a fixed unital but not necessarily commutative ring R, Mod_R is the category of left R-modules and R-module homomorphisms. This category is denoted by $\mathsf{Vect}_{\Bbbk}$ when the ring happens to be a field $\Bbbk$ and abbreviated as Ab in the case of $\mathsf{Mod}_{\mathbb{Z}}$, as a $\mathbb{Z}$-module is precisely an abelian group.
(vi) Graph has graphs as objects and graph morphisms (functions carrying vertices to vertices and edges to edges, preserving incidence relations) as morphisms. In the variant $\mathsf{DirGraph}$, objects are directed graphs, whose edges are now depicted as arrows, and morphisms are directed graph morphisms, which must preserve sources and targets.
(vii) Man has smooth (i.e., infinitely differentiable) manifolds as objects and smooth maps as morphisms.
(viii) Meas has measurable spaces as objects and measurable functions as morphisms.
(ix) Poset has partially-ordered sets as objects and order-preserving functions as morphisms.
(x) Ch_R has chain complexes of R-modules as objects and chain homomorphisms as morphisms.[7]
(xi) For any *signature* σ, specifying constant, function, and relation symbols, and for any collection of well formed sentences $\mathbb{T}$ in the first-order language associated to σ, there is a category $\mathsf{Model}_{\mathbb{T}}$ whose objects are σ-structures that *model* $\mathbb{T}$, i.e., sets equipped with appropriate constants, relations, and functions satisfying the axioms $\mathbb{T}$. Morphisms are functions that preserve the specified constants, relations, and functions, in the usual sense.[8] Special cases include (iv), (v), (vi), (ix), and (x).

The preceding are all examples of *concrete categories*, those whose objects have underlying sets and whose morphisms are functions between these underlying sets, typically the "structure-preserving" morphisms. A more precise definition of a concrete category is given in 1.6.17. However, "abstract" categories are also prevalent:

EXAMPLE 1.1.4.

(i) For a unital ring R, Mat_R is the category whose objects are positive integers and in which the set of morphisms from n to m is the set of $m \times n$ matrices with values in

[5][**EM45**, p. 239] emphasizes that the data of a function should include specified sets of inputs and potential outputs, a perspective that was somewhat radical at the time.

[6]A **basepoint** is simply a chosen distinguished point in the set or space.

[7]A **chain complex** $C_\bullet$ is a collection $(C_n)_{n\in\mathbb{Z}}$ of R-modules equipped with R-module homomorphisms $d\colon C_n \to C_{n-1}$, called **boundary homomorphisms**, with the property that $d^2 = 0$, i.e., the composite of any two boundary maps is the zero homomorphism. A map of chain complexes $f\colon C_\bullet \to C'_\bullet$ is comprised of a collection of homomorphisms $f_n\colon C_n \to C'_n$ so that $df_n = f_{n-1}d$ for all $n \in \mathbb{Z}$.

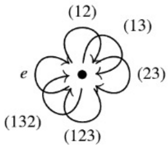
[8]Model theory pays greater attention to other types of morphisms, for instance the *elementary embeddings*, which are (automatically injective) functions that preserve and reflect satisfaction of first-order formulae.

R. Composition is by matrix multiplication

$$n \xrightarrow{A} m, \quad m \xrightarrow{B} k \qquad \rightsquigarrow \qquad n \xrightarrow{B \cdot A} k$$

with identity matrices serving as the identity morphisms.

(ii) A group G (or, more generally, a monoid[9]) defines a category $\mathsf{B}G$ with a single object. The group elements are its morphisms, each group element representing a distinct endomorphism of the single object, with composition given by multiplication. The identity element $e \in G$ acts as the identity morphism for the unique object in this category.

$\mathsf{B}S_3 =$ 

(iii) A poset $(\mathsf{P}, \leq)$ (or, more generally, a preorder[10]) may be regarded as a category. The elements of P are the objects of the category and there exists a unique morphism $x \to y$ if and only if $x \leq y$. Transitivity of the relation "$\leq$" implies that the required composite morphisms exist. Reflexivity implies that identity morphisms exist.

(iv) In particular, any ordinal $\alpha = \{\beta \mid \beta < \alpha\}$ defines a category whose objects are the smaller ordinals. For example, $\mathbb{0}$ is the category with no objects and no morphisms. $\mathbb{1}$ is the category with a single object and only its identity morphism. $\mathbb{2}$ is the category with two objects and a single non-identity morphism, conventionally depicted as $0 \to 1$. ω is the category *freely generated by the graph*

$$0 \to 1 \to 2 \to 3 \to \cdots$$

in the sense that every non-identity morphism can be uniquely factored as a composite of morphisms in the displayed graph; a precise definition of the notion of free generation is given in Example 4.1.13.

(v) A set may be regarded as a category in which the elements of the set define the objects and the only morphisms are the required identities. A category is **discrete** if every morphism is an identity.

(vi) Htpy, like Top, has spaces as its objects but morphisms are homotopy classes of continuous maps. Htpy_* has based spaces as its objects and basepoint-preserving homotopy classes of based continuous maps as its morphisms.

(vii) $\mathsf{Measure}$ has measure spaces as objects. One reasonable choice for the morphisms is to take equivalence classes of measurable functions, where a parallel pair of functions are equivalent if their domain of difference is contained within a set of measure zero.

Thus, the philosophy of category theory is extended. The categories listed in Example 1.1.3 suggest that mathematical objects ought to be considered together with the appropriate notion of morphism between them. The categories listed in Example 1.1.4 illustrate that

[9] A **monoid** is a set M equipped with an associative binary multiplication operation $M \times M \xrightarrow{\cdot} M$ and an identity element $e \in M$ serving as a two-sided identity. In other words, a monoid is precisely a one-object category.

[10] A **preorder** is a set with a binary relation $\leq$ that is reflexive and transitive. In other words, a preorder is precisely a category in which there are no parallel pairs of distinct morphisms between any fixed pair of objects. A **poset** is a preorder that is additionally antisymmetric: $x \leq y$ and $y \leq x$ implies that $x = y$.

these morphisms are not always functions.[11] The morphisms in a category are also called **arrows** or **maps**, particularly in the contexts of Examples 1.1.4 and 1.1.3, respectively.

REMARK 1.1.5. Russell's paradox implies that there is no set whose elements are "all sets." This is the reason why we have used the vague word "collection" in Definition 1.1.1. Indeed, in each of the examples listed in 1.1.3, the collection of objects is not a set. Eilenberg and Mac Lane address this potential area of concern as follows:

> . . . the whole concept of a category is essentially an auxiliary one; our basic concepts are essentially those of a *functor* and of a natural transformation . . . . The idea of a category is required only by the precept that every function should have a definite class as domain and a definite class as range, for the categories are provided as the domains and ranges of functors. Thus one could drop the category concept altogether and adopt an even more intuitive standpoint, in which a functor such as "Hom" is not defined over the category of "all" groups, but for each particular pair of groups which may be given. [**EM45**]

The set-theoretical issues that confront us while defining the notion of a category will compound as we develop category theory further. For that reason, common practice among category theorists is to work in an extension of the usual Zermelo–Fraenkel axioms of set theory, with new axioms allowing one to distinguish between "small" and "large" sets, or between sets and classes. The search for the most useful set-theoretical foundations for category theory is a fascinating topic that unfortunately would require too long of a digression to explore.[12] Instead, we sweep these foundational issues under the rug, not because these issues are not serious or interesting, but because they distract from the task at hand.[13]

For the reasons just discussed, it is important to introduce adjectives that explicitly address the size of a category.

DEFINITION 1.1.6. A category is **small** if it has only a set's worth of arrows.

By Remark 1.1.2, a small category has only a set's worth of objects. If C is a small category, then there are functions

$$\mathrm{mor}\,\mathsf{C} \;\substack{\xrightarrow{\mathrm{dom}} \\ \xleftarrow{\mathrm{id}} \\ \xrightarrow[\mathrm{cod}]{}}\; \mathrm{ob}\,\mathsf{C}$$

that send a morphism to its domain and its codomain and an object to its identity.

[11] Reid's *Undergraduate algebraic geometry* emphasizes that the morphisms are not always functions, writing "Students who disapprove are recommended to give up at once and take a reading course in category theory instead" [**Rei88**, p. 4].

[12] The preprint [**Shu08**] gives an excellent overview, though it is perhaps better read after Chapters 1–4.

[13] If pressed, let us assume that there exists a countable sequence of **inaccessible cardinals**, meaning uncountable cardinals that are **regular** and **strong limit**. A cardinal κ is **regular** if every union of fewer than κ sets each of cardinality less than κ has cardinality less than κ, and **strong limit** if $\lambda < \kappa$ implies that $2^\lambda < \kappa$. Inaccessibility means that sets of size less than κ are closed under power sets and κ-small unions. If κ is inaccessible, then the κ-stage of the *von Neumann hierarchy*, the set V_κ of sets of rank less than κ, is a model of Zermelo–Fraenkel set theory with choice (ZFC); the set V_κ is a *Grothendieck universe*. The assumption that there exists a countable sequence of inaccessible cardinals means that we can "do set theory" inside the universe V_κ, and then enlarge the universe if necessary as often as needed.

If ZFC is consistent, these axioms cannot prove the existence of an inaccessible cardinal or the consistency of the assumption that one exists (by Gödel's second incompleteness theorem). Nonetheless, from the perspective of the hierarchy of large cardinal axioms, the existence of inaccessibles is a relatively mild hypothesis.

None of the categories in Example 1.1.3 are small—each has too many objects—but "locally" they resemble small categories in a sense made precise by the following notion:

DEFINITION 1.1.7. A category is **locally small** if between any pair of objects there is only a set's worth of morphisms.

It is traditional to write

$$\mathsf{C}(X,Y) \qquad \text{or} \qquad \mathrm{Hom}(X,Y) \tag{1.1.8}$$

for the set of morphisms from X to Y in a locally small category C.[14] The set of arrows between a pair of fixed objets in a locally small category is typically called a **hom-set**, whether or not it is a set of "homomorphisms" of any particular kind. Because the notation (1.1.8) is so convenient, it is also adopted for the collection of morphisms between a fixed pair of objects in a category that is not necessarily locally small.

A category provides a context in which to answer the question "When is one thing the same as another thing?" Almost universally in mathematics, one regards two objects of the same category to be "the same" when they are isomorphic, in a precise categorical sense that we now introduce.

DEFINITION 1.1.9. An **isomorphism** in a category is a morphism $f\colon X \to Y$ for which there exists a morphism $g\colon Y \to X$ so that $gf = 1_X$ and $fg = 1_Y$. The objects X and Y are **isomorphic** whenever there exists an isomorphism between X and Y, in which case one writes $X \cong Y$.

An **endomorphism**, i.e., a morphism whose domain equals its codomain, that is an isomorphism is called an **automorphism**.

EXAMPLE 1.1.10.

(i) The isomorphisms in Set are precisely the **bijections**.
(ii) The isomorphisms in Group, Ring, Field, or Mod_R are the bijective homomorphisms.
(iii) The isomorphisms in the category Top are the **homeomorphisms**, i.e., the continuous functions with continuous inverse, which is a stronger property than merely being a bijective continuous function.
(iv) The isomorphisms in the category Htpy are the **homotopy equivalences**.
(v) In a poset $(\mathsf{P}, \leq)$, the axiom of antisymmetry asserts that $x \leq y$ and $y \leq x$ imply that $x = y$. That is, the only isomorphisms in the category $(\mathsf{P}, \leq)$ are identities.

Examples 1.1.10(ii) and (iii) suggest the following general question: In a concrete category, when are the isomorphisms precisely those maps in the category that induce bijections between the underlying sets? We will see an answer in Lemma 5.6.1.

DEFINITION 1.1.11. A **groupoid** is a category in which every morphism is an isomorphism.

EXAMPLE 1.1.12.

(i) A **group** is a groupoid with one object.[15]
(ii) For any space X, its **fundamental groupoid** $\Pi_1(X)$ is a category whose objects are the points of X and whose morphisms are endpoint-preserving homotopy classes of paths.

---

[14]Mac Lane credits Emmy Noether for emphasizing the importance of homomorphisms in abstract algebra, particularly the homomorphism onto a quotient group, which plays an integral role in the statement of her first isomorphism theorem. His recollection is that the arrow notation first appeared around 1940, perhaps due to Hurewicz [**ML88**]. The notation $\mathrm{Hom}(X,Y)$ was first used in [**EM42a**] for the set of homomorphisms between a pair of abelian groups.

[15]This is not simply an example; it is a definition.

A **subcategory** D of a category C is defined by restricting to a subcollection of objects and subcollection of morphisms subject to the requirements that the subcategory D contains the domain and codomain of any morphism in D, the identity morphism of any object in D, and the composite of any composable pair of morphisms in D. For example, there is a subcategory $\mathsf{CRing} \subset \mathsf{Ring}$ of commutative unital rings. Both of these form subcategories of the category Rng of not-necessarily unital rings and homomorphisms that need not preserve the multiplicative unit.[16]

LEMMA 1.1.13. *Any category* C *contains a* ***maximal groupoid****, the subcategory containing all of the objects and only those morphisms that are isomorphisms.*

PROOF. Exercise 1.1.ii. □

For instance, $\mathsf{Fin}_{\mathrm{iso}}$, the category of finite sets and bijections, is the maximal subgroupoid of the category Fin of finite sets and all functions. Example 1.4.9 will explain how this groupoid can be regarded as a categorification of the natural numbers, providing a vantage point from which to prove the laws of elementary arithmetic.

Exercises.

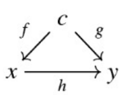
EXERCISE 1.1.i.

(i) Show that a morphism can have at most one inverse isomorphism.
(ii) Consider a morphism $f\colon x \to y$. Show that if there exists a pair of morphisms $g, h\colon y \rightrightarrows x$ so that $gf = 1_x$ and $fh = 1_y$, then $g = h$ and f is an isomorphism.

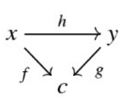
EXERCISE 1.1.ii. Let C be a category. Show that the collection of isomorphisms in C defines a subcategory, the **maximal groupoid** inside C.

EXERCISE 1.1.iii. For any category C and any object $c \in \mathsf{C}$, show that:

(i) There is a category c/C whose objects are morphisms $f\colon c \to x$ with domain c and in which a morphism from $f\colon c \to x$ to $g\colon c \to y$ is a map $h\colon x \to y$ between the codomains so that the triangle



commutes, i.e., so that $g = hf$.
(ii) There is a category C/c whose objects are morphisms $f\colon x \to c$ with codomain c and in which a morphism from $f\colon x \to c$ to $g\colon y \to c$ is a map $h\colon x \to y$ between the domains so that the triangle



commutes, i.e., so that $f = gh$.

The categories c/C and C/c are called **slice categories** of C **under** and **over** c, respectively.

[16]To justify our default notion of ring, see Poonen's "Why all rings should have a 1" [**Poo14**]. The relationship between unital and non-unital rings is explored in greater depth in §4.6.

1.2. Duality

> The dual of any axiom for a category is also an axiom . . . A simple metamathematical argument thus proves the *duality principle*. If any statement about a category is deducible from the axioms for a category, the dual statement is likely deducible.
>
> Saunders Mac Lane, "Duality for groups" [**ML50**]

Upon first acquaintance, the primary role played by the notion of a category might appear to be taxonomic: vector spaces and linear maps define one category, manifolds and smooth functions define another. But a category, as defined in 1.1.1, is also a mathematical object in its own right, and as with any mathematical definition, this one is worthy of further consideration. Applying a mathematician's gaze to the definition of a category, the following observation quickly materializes. If we visualize the morphisms in a category as arrows pointing from their domain object to their codomain object, we might imagine simultaneously reversing the directions of every arrow. This leads to the following notion.

Definition 1.2.1. Let C be any category. The **opposite category** C^{op} has

- the same objects as in C, and
- a morphism f^{op} in C^{op} for each a morphism f in C so that the domain of f^{op} is defined to be the codomain of f and the codomain of f^{op} is defined to be the domain of f: i.e.,

$$f^{\mathrm{op}} \colon X \to Y \quad \in \mathsf{C}^{\mathrm{op}} \qquad \leftrightsquigarrow \qquad f \colon Y \to X \quad \in \mathsf{C}\,.$$

That is, C^{op} has the same objects and morphisms as C, except that "each morphism is pointing in the opposite direction." The remaining structure of the category C^{op} is given as follows:

- For each object X, the arrow 1_X^{op} serves as its identity in C^{op}.
- To define composition, observe that a pair of morphisms $f^{\mathrm{op}}, g^{\mathrm{op}}$ in C^{op} is composable precisely when the pair g, f is composable in C, i.e., precisely when the codomain of g equals the domain of f. We then define $g^{\mathrm{op}} \cdot f^{\mathrm{op}}$ to be $(f \cdot g)^{\mathrm{op}}$: i.e.,

$$\begin{array}{ccc} f^{\mathrm{op}} \colon X \to Y,\ g^{\mathrm{op}} \colon Y \to Z \quad \in \mathsf{C}^{\mathrm{op}} & \rightsquigarrow & g^{\mathrm{op}} f^{\mathrm{op}} \colon X \to Z \quad \in \mathsf{C}^{\mathrm{op}} \\ \updownarrow & & \updownarrow \\ g \colon Z \to Y,\ f \colon Y \to X \quad \in \mathsf{C} & \rightsquigarrow & fg \colon Z \to X \quad \in \mathsf{C} \end{array}$$

The data described in Definition 1.2.1 defines a category C^{op}—i.e., the composition law is associative and unital—if and only if C defines a category. In summary, the process of "turning around the arrows" or "exchanging domains and codomains" exhibits a syntactical self-duality satisfied by the axioms for a category. Note that the category C^{op} contains precisely the same information as the category C. Questions about the one can be answered by examining the other.

Example 1.2.2.

(i) $\mathsf{Mat}_R^{\mathrm{op}}$ is the category whose objects are non-zero natural numbers and in which a morphism from m to n is an $m \times n$ matrix with values in R. The upshot is that a reader who would have preferred the opposite handedness conventions when defining Mat_R would have lost nothing by adopting them.

(ii) When a preorder $(\mathsf{P}, \leq)$ is regarded as a category, its opposite category is the category that has a morphism $x \to y$ if and only if $y \leq x$. For example, ω^{op} is the category freely generated by the graph

$$\cdots \to 3 \to 2 \to 1 \to 0\,.$$

(iii) If G is a group, regarded as a one-object groupoid, the category $(\mathsf{B}G)^{\mathrm{op}} \cong \mathsf{B}(G^{\mathrm{op}})$ is again a one-object groupoid, and hence a group. The group G^{op} is called the **opposite group** and is used to define right actions as a special case of left actions; see Example 1.3.9.

This syntactical duality has a very important consequence for the development of category theory. Any theorem containing a universal quantification of the form "for all categories C" also necessarily applies to the opposites of these categories. Interpreting the result in the dual context leads to a **dual theorem**, proven by the dual of the original proof, in which the direction of each arrow appearing in the argument is reversed. The result is a two-for-one deal: any proof in category theory simultaneously proves two theorems, the original statement and its dual.[17] For example, the reader may have found Exercise 1.1.iii redundant, precisely because the statements (i) and (ii) are dual; see Exercise 1.2.i.

To illustrate the principle of duality in category theory, let us consider the following result, which provides an important characterization of the isomorphisms in a category.

LEMMA 1.2.3. *The following are equivalent:*

(i) $f\colon x \to y$ *is an isomorphism in* C.

(ii) For all objects $c \in \mathsf{C}$, *post-composition with* f *defines a bijection*

$$f_*\colon \mathsf{C}(c,x) \to \mathsf{C}(c,y)\,.$$

(iii) For all objects $c \in \mathsf{C}$, *pre-composition with* f *defines a bijection*

$$f^*\colon \mathsf{C}(y,c) \to \mathsf{C}(x,c)\,.$$

REMARK 1.2.4. In language introduced in Chapter 2, Lemma 1.2.3 asserts that isomorphisms in a locally small category are defined *representably* in terms of isomorphisms in the category of sets. That is, a morphism $f\colon x \to y$ in an arbitrary locally small category C is an isomorphism if and only if the post-composition function $f_*\colon \mathsf{C}(c,x) \to \mathsf{C}(c,y)$ between hom-sets defines an isomorphism in Set for each object $c \in \mathsf{C}$.

In set theoretical foundations that permit the definition of functions between large sets, the proof given here applies also to non-locally small categories. In our exposition, the set theoretical hypotheses of smallness and local smallness will only appear when there are essential subtleties concerning the sizes of the categories in question. This is not one of those occasions.

PROOF OF LEMMA 1.2.3. We will prove the equivalence (i) $\Leftrightarrow$ (ii) and conclude the equivalence (i) $\Leftrightarrow$ (iii) by duality.

Assuming (i), namely that $f\colon x \to y$ is an isomorphism with inverse $g\colon y \to x$, then, as an immediate application of the associativity and identity laws for composition in a category, post-composition with g defines an inverse function

$$g_*\colon \mathsf{C}(c,y) \to \mathsf{C}(c,x)$$

to f_* in the sense that the composites

$$g_*f_*\colon \mathsf{C}(c,x) \to \mathsf{C}(c,x) \quad \text{and} \quad f_*g_*\colon \mathsf{C}(c,y) \to \mathsf{C}(c,y)$$

[17] More generally, the proof of a statement of the form "for all categories $\mathsf{C}_1, \mathsf{C}_2, \ldots, \mathsf{C}_n$" leads to 2^n dual theorems. In practice, however, not all of the dual statements will differ meaningfully from the original; see e.g., §4.3.

are both the identity function: for any $h\colon c \to x$ and $k\colon c \to y$, $g_*f_*(h) = gfh = h$ and $f_*g_*(k) = fgk = k$.

Conversely, assuming (ii), there must be an element $g \in \mathsf{C}(y,x)$ whose image under $f_*\colon \mathsf{C}(y,x) \to \mathsf{C}(y,y)$ is 1_y. By construction, $1_y = fg$. But now, by associativity of composition, the elements $gf, 1_x \in \mathsf{C}(x,x)$ have the common image f under the function $f_*\colon \mathsf{C}(x,x) \to \mathsf{C}(x,y)$, whence $gf = 1_x$. Thus, f and g are inverse isomorphisms.

We have just proven the equivalence (i) ⇔ (ii) for all categories and in particular for the category C^{op}: i.e., a morphism $f^{\mathrm{op}}\colon y \to x$ in C^{op} is an isomorphism if and only if

(1.2.5) $$f^{\mathrm{op}}_*\colon \mathsf{C}^{\mathrm{op}}(c,y) \to \mathsf{C}^{\mathrm{op}}(c,x) \text{ is an isomorphism for all } c \in \mathsf{C}^{\mathrm{op}}.$$

Interpreting the data of C^{op} in its opposite category C, the statement (1.2.5) expresses the same mathematical content as

(1.2.6) $$f^*\colon \mathsf{C}(y,c) \to \mathsf{C}(x,c) \text{ is an isomorphism for all } c \in \mathsf{C}.$$

That is: $\mathsf{C}^{\mathrm{op}}(c,x) = \mathsf{C}(x,c)$, post-composition with f^{op} in C^{op} translates to pre-composition with f in the opposite category C. The notion of isomorphism, as defined in 1.1.9, is self-dual: $f^{\mathrm{op}}\colon y \to x$ is an isomorphism in C^{op} if and only if $f\colon x \to y$ is an isomorphism in C. So the equivalence (i) ⇔ (ii) in C^{op} expresses the equivalence (i) ⇔ (iii) in C.[18] □

Concise expositions of the duality principle in category theory may be found in [**Awo10**, §3.1] and [**HS97**, §II.3]. As we become more comfortable with arguing by duality, dual proofs and eventually also dual statements will seldom be described in this much detail.

Categorical definitions also have duals; for instance:

Definition 1.2.7. A morphism $f\colon x \to y$ in a category is

(i) a **monomorphism** if for any parallel morphisms $h, k\colon w \rightrightarrows x$, $fh = fk$ implies that $h = k$; or

(ii) an **epimorphism** if for any parallel morphisms $h, k\colon y \rightrightarrows z$, $hf = kf$ implies that $h = k$.

Note that a monomorphism or epimorphism in C is, respectively, an epimorphism or monomorphism in C^{op}. In adjectival form, a monomorphism is **monic** and an epimorphism is **epic**. In common shorthand, a monomorphism is a **mono** and an epimorphism is an **epi**. For graphical emphasis, monos are often decorated with a tail "$\rightarrowtail$" while epis may be decorated at their head "$\twoheadrightarrow$."

The following dual statements re-express Definition 1.2.7:

(i) $f\colon x \to y$ is a monomorphism in C if and only if for all objects $c \in \mathsf{C}$, post-composition with f defines an injection $f_*\colon \mathsf{C}(c,x) \to \mathsf{C}(c,y)$.

(ii) $f\colon x \to y$ is an epimorphism in C if and only if for all objects $c \in \mathsf{C}$, pre-composition with f defines an injection $f^*\colon \mathsf{C}(y,c) \to \mathsf{C}(x,c)$.

Example 1.2.8. Suppose $f\colon X \to Y$ is a monomorphism in the category of sets. Then, in particular, given any two maps $x, x'\colon 1 \rightrightarrows X$, whose domain is the singleton set, if $fx = fx'$ then $x = x'$. Thus, monomorphisms are injective functions. Conversely, any injective function can easily be seen to be a monomorphism.

Similarly, a function $f\colon X \to Y$ is an epimorphism in the category of sets if and only if it is surjective. Given functions $h, k\colon Y \rightrightarrows Z$, the equation $hf = kf$ says exactly that h is equal to k on the image of f. This only implies that $h = k$ in the case where the image is all of Y.

[18]A similar translation, as just demonstrated between the statements (1.2.5) and (1.2.6), transforms the proof of (i) ⇔ (ii) into a proof of (i) ⇔ (iii).