The Category of Sets

Nic Ford

Mathcamp 2022

Introduction

These are notes from a class called "The Category of Sets" that I taught at Canada/USA Mathcamp 2022. The target audience is someone who has encountered set theory before and is reasonably comfortable with abstract proofs, but who knows nothing about category theory and might be curious what all the fuss is about. These notes were designed to be somewhat readable on their own, but I make no promises that they'll make nearly as much sense outside the context of that class — some parts stand a good chance of seeming quite unmotivated! I'm heavily indebted to Eric Wofsey for almost everything about the structure of this class, including the title. I had his notes from 2010 open while writing these, and this class would not have turned out nearly as well if I hadn't had the opportunity to crib from them.

1 Categories

If you've heard of category theory at all, there's a decent chance that you've picked up on the fact that it has a reputation for being mystifying and impenetrable. (Indeed, a common nickname for category theory among mathematicians is "abstract nonsense," and I'm pretty sure the people who say this are only halfway joking.) If this class has a single goal, it is to try to dispel this reputation somewhat by exploring the big ideas in a concrete, familiar setting: set theory.

What is category theory? There are many perspectives one could take on this question, but I want to focus on just one for now: category theory is about the properties of mathematical objects that be described purely in terms of the *functions* or *maps* between them, rather than the internal structure of the objects themselves. In our central example, that will mean trying to describe as many of the core concepts from set theory purely in terms of functions, without referring directly to elements at all.

It would be very natural at this point to ask why anyone would ever want to do this. If you only cared about set theory itself, my answer would be pretty weak; there are a few places where the categorical version of some concept might offer a fun and interesting new perspective, but that's about it.

The real power comes from the fact that large swaths of mathematics can *also* be described in a similar "functions-first" way, including groups, rings, fields, vector spaces, topological spaces, manifolds, graphs, and many, many more. (Don't worry — you don't need to know what any of these are to follow these notes!) When you can manage to do a piece of set theory while only talking about functions, often what you have actually done is prove a theorem that applies equally well to all of these other objects simultaneously.

1.1 What is a category?

We'll start with the formal definition, and then see how it applies to sets.

Definition 1.1. A **category** $\mathscr{C}$ consists of the following pieces of information:

- A collection of **objects**.
- For any two objects A, B, a set called $\text{Maps}(A, B)$. (Elements of this set will be called **maps from** A **to** B. A map from A to B will often be written with notation like $f : A \to B$.)
- A **composition operation** $\circ$, which takes in two maps $f : A \to B$ and $g : B \to C$ and produces a map we'll call $g \circ f : A \to C$.

(Notice, in particular, that the composition of two maps is only defined if the "head" of the second map's arrow lines up with the "tail" of the first.) We also need the composition operation to satisfy the following two properties:

- Composition is *associative*, that is, given maps $f : A \to B$, $g : B \to C$, and $h : C \to D$, we have $h \circ (g \circ f) = (h \circ g) \circ f$.
- Every object A has an *identity map* $\text{Id}_A : A \to A$ for which $\text{Id}_A \circ f = f$ and $g \circ I = \text{Id}_A = g$ for any f or g for which those compositions are defined.

This definition will probably make a lot more sense once we see what it means for sets — in fact, I think it's fair to say that this definition was basically written with the example of sets in mind!

Example 1.2. In the **category of sets**, which we'll call Set, the objects are sets, and for any two sets A, B, the maps from A to B are just the functions from A to B. Composition of maps is given by the usual composition of functions $(g \circ f)(x) = g(f(x))$. (This is the reason for the otherwise funny order in which we write composition!) Composition of functions is indeed associative — which I encourage you to check — and for any set A the identity map is the identity function $\text{Id}_A(x) = x$.

This category will be far more important to us than any other, but the definition of a category might be easier to absorb with a couple other examples. We won't use any of these except in some optional exercises, so feel free to skip them.

Example 1.3. In the **category of pointed sets**, written Set^*, an object is a pair (S, s) where S is a set and s is an element of S. (In particular, this means that the set can't be empty!) A map from (S, s) to (T, t) is a function $f : S \to T$ for which $f(s) = t$ — in other words, a map of pointed sets is a map of sets which "preserves the chosen element." Composition and identities work just like in Set.

Example 1.4. In the **category of vector spaces**, written Vect, the objects are vector spaces (let's say over the real numbers) and the maps are linear maps. If you've seen some linear algebra, I encourage you to check that this can be made into a category with the appropriate choice of composition law.

Example 1.5. We can form a category, which I'll call Nat, whose objects are natural numbers and whose maps are given by the following rule: for any two natural numbers m, n, if $m \leq n$ then there is a single map from m to n; and if $m > n$ there are no maps from m to n. How should composition and identity maps work in this category? (You should find that there is actually only one possible answer to this question!)

This last example illustrates an important point: while the definition of a category was certainly *inspired* by sets, the objects of a category don't have to have anything at all to do with sets, and indeed don't need to have "elements" at all! Similarly, the fact that we call the elements of $\mathrm{Maps}(A,B)$ "maps" is inspired by the example of functions between sets, but nothing about the definition of a category requires them to actually map anything to anything else. This is part of the power of category theory: a lot of the same constructions that we can perform in set theory can be applied equally well in a wide variety of other settings, some of which (like the pointed sets and vector spaces above) look at least something like sets and some of which (like this last example) don't. The vast majority of the examples we'll care about in this class will be of the set-like type, but it's good to keep in the back of your mind that the other ones exist.

In order to help with this possible confusion at least a little bit, *in this class, we will always use the words "set" and "function" when we are actually working with sets, and use "object" and "map" when we're talking about an arbitrary category.*

1.2 Isomorphisms

The first category-theoretic concept we'll discuss is a condition for treating two objects in a category as "equivalent":

Definition 1.6. Suppose A and B are objects in some category $\mathscr{C}$. A map $f: A \to B$ is called an **isomorphism** if there is a map $f^{-1}: B \to A$ such that $f^{-1} \circ f = \mathrm{Id}_A$ and $f \circ f^{-1} = \mathrm{Id}_B$. (As the notation suggests, f^{-1} is called the **inverse** of f.) If there is an isomorphism from A to B, we say that A and B are **isomorphic**, and we'll write $A \cong B$.

Notice that if $\mathscr{C} = \mathsf{Set}$, then this is exactly the condition for being a *bijection*, which means that two sets are isomorphic if and only if they have the same cardinality. In some of the exercises you'll think about what isomorphisms look like in some other categories.

In general, though, a good way to think about isomorphic objects is that if A and B are isomorphic, then the maps into or out of A "look the same" as the maps into or out of B. For example, if f is an isomorphism, then composing with f on the right gives a way to turn maps $C \to A$ into maps $C \to B$, and this process can be reversed by composing with f^{-1}. (In particular, every object is isomorphic to itself; do you see why?) You'll make this more precise in Exercise 1.9.

Category theory is all about characterizing mathematical objects in terms of their maps to and from other objects, and so isomorphic objects are often thought of as "interchangeable"; anything map-related you can say about one will be equally true of the other. (It might be helpful to think through why this is true of two sets that are connected by a bijection.) This means, in particular, that when we write down a definition of some type of object in category theory, we will usually be happy if our definition only picks out a unique object *up to isomorphism*. We'll see this in action in the definition we're about to consider.

1.3 Terminal Objects

We're finally ready for our first example of recasting a set-theoretic construction in categorical language.

Definition 1.7. An object A in a category $\mathscr{C}$ is called **terminal** if, for any other object B of $\mathscr{C}$, there is exactly one map $B \to A$.

In the category of sets, *one-element sets* are terminal, because if A has only one element and B is any other set, the only way to construct a function from B to A is to send every element of B to the only element A has.

Notice that all one-elements sets are isomorphic to each other, because isomorphism in the category of sets is just about the cardinality of the set. Moreover, in this particular case, it's not too difficult to show that one-element sets are the *only* terminal objects; I encourage you to take a moment to check this for yourself. (If you do, remember that there are two different ways a set might not be a one-element set: it could have more the one element, or it could be empty! Does your proof cover both cases?)

Perhaps surprisingly, this can also be seen to be true for a reason that, in a sense, has nothing at all to do with sets! In *any* category, terminal objects are all isomorphic to each other:

Theorem 1.8. *Suppose A and A' are two terminal objects in the same category. Then A and A' are isomorphic, and there is only one isomorphism between them.*

Proof. We can use the fact that A' is terminal to get a map $f : A \to A'$. Notice that, by Definition 1.7, f is actually the *only* map from A to A', so if A and A' are going to be isomorphic, this map we have here had better be the isomorphism! We can get a potential inverse for it by using the fact that A is terminal to get a map $g : A' \to A$; we just have to show that $g \circ f = \mathrm{Id}_A$ and $f \circ g = \mathrm{Id}_{A'}$.

Notice that $g \circ f$ is a map from A to A. A is terminal, so (applying the definition with $B = A$) there is only *one* map from A to A. But Id_A is *also* a map from A to A, which means that we must have $g \circ f = \mathrm{Id}_A$! A basically identical argument shows that $f \circ g = \mathrm{Id}_{A'}$. (I'll leave it to you to check that the isomorphism is unique.) □

This argument will come up over and over again in this class. A property like the one in Definition 1.7 is called a **universal property**, and the argument from the proof of Theorem 1.8 can be adapted to show that *any* object defined by a universal property is unique up to a unique isomorphism.

Because of Theorem 1.8, it's very common to talk about "the" terminal object in a category rather than just "a" terminal object. This language takes for granted the philosophy we described earlier where we treat isomorphic objects as interchangeable. It's important to remember, though, that none of what we've said so far guarantees that a category will have a terminal object at all! We've seen that Set has terminal objects, but in at least one of the exercises you'll see some examples of a category that doesn't.

All together, we have found a way to take a concept from set theory — in this case, the concept of a one-element set — and define it without mentioning elements at all! While this does give us a nice new perspective on one-element sets, the real power of this move comes from the fact that the concept of terminal objects applies to lots of other categories as well, even ones whose objects don't have anything like elements. This will be equally true of the other universal properties we'll be studying throughout this class.

Exercises

Exercises that are especially *important* are marked with ⚠. Exercises that are especially *challenging* are marked with (⋆).

1.1. Prove that, for any object A in a category $\mathscr{C}$, the identity map $\mathrm{Id}_A : A \to A$ is unique. That is, if $I : A \to A$ is a map with the property that $I \circ f = f$ and $g \circ I = g$ whenever these compositions are defined, then $I = \mathrm{Id}_A$.

1.2. Suppose A and B are two objects in some category and $f : A \to B$ is an isomorphism. Show that the inverse of f is unique. That is, if $g : B \to A$ is another map and $g \circ f = \mathrm{Id}_A$ and $f \circ g = \mathrm{Id}_B$, then $g = f^{-1}$.

1.3. Give a complete description of a category $\mathscr{C}$ with the following properties:

- $\mathscr{C}$ has infinitely many objects.
- There are two objects A and B in $\mathscr{C}$ which are not isomorphic to each other, but every other object of $\mathscr{C}$ is isomorphic to either A or B.

Make sure you explain what all the objects and all the arrows are!

1.4. Show that "being isomorphic" is an equivalence relation on the objects in any category. (That is, for any objects A, B, C, show that $A \cong A$; that if $A \cong B$ then $B \cong A$; and that if $A \cong B$ and $B \cong C$, then $A \cong C$.)

1.5. *[Don't worry about this problem unless you've seen enough group theory to know what a group homomorphism is.]* This problem is about the **category of groups**, which we'll denote Grp. The objects of Grp are groups, and, for any two groups G, H, the maps from G to H are the group homomorphisms.

(a) Check that this indeed defines a category.

(b) Prove that the isomorphisms in Grp are the same as the group isomorphisms as they might be defined in a group theory class.

(c) Does Grp have a terminal object? If so, what is it?

1.6. *[Don't worry about this problem unless you've seen some linear algebra.]* Does the category Vect from Example 1.4 have a terminal object? If so, what is it?

1.7. *[Don't worry about this problem unless you've seen some linear algebra.]* Form the category FinMat in the following way. The objects of FinMat are natural numbers (including 0). For any two natural numbers m, n, the maps $m \to n$ are all $n \times m$ matrices with real entries, and composition is matrix multiplication. We'll use the convention that, for any k, there is exactly one $0 \times k$ matrix and one $k \times 0$ matrix, which is "empty."

(a) Check that this indeed defines a category.

(b) Does this category have a terminal object? If so, what is it?

(c) This category is closely related to one of the examples from this section. Which one? Describe the relationship as precisely as you can.

1.8. (a) What do isomorphisms look like in the category Nat from Example 1.5?

(b) Does Nat have a terminal object? If so, what is it?

1.9. ⚠Let A, B be two objects in some category, and suppose $f : A \to B$ is a map between them.

(a) If C is another object in the same category, use f to construction a function $f_* : \mathrm{Maps}(C, A) \to \mathrm{Maps}(C, B)$. *[Remember that, no matter what category we're working in, these map sets are actual sets!]*

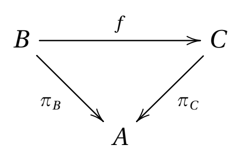
(b) You can do something similar to construct a function f^* involving maps into C. What should the source and target of f^* be?

(c) Show that f is an isomorphism if and only if, for every object C, the function $f_* : \text{Maps}(C, A) \to \text{Maps}(C, B)$ you constructed above is a bijection. *[Hint: Consider the case $C = B$, and then consider the case $C = A$.]*

(d) Do the same for f^*.

[This is one way to make precise the idea that, when two objects are isomorphic, they "look the same" with respect to maps into or out of them.]

1.10. Let A be an object in a category $\mathscr{C}$. You can form a new category called the **slice category over** A, written $\mathscr{C}/A$, as follows. An object of $\mathscr{C}/A$ is a pair (B, π_B) where B is an object of $\mathscr{C}$ and $\pi_B : B \to A$ is a map from $\mathscr{C}$. A morphism from (B, π_B) to (C, π_C) is map $f : B \to C$ (from $\mathscr{C}$) with the property that $\pi_C \circ f = \pi_B$:



(a) Show that this defines a category.

(b) Slice categories always have terminal objects. What are they?

2 Products

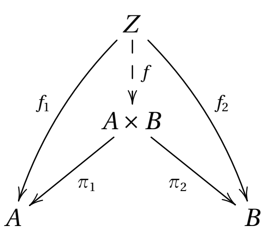
In the last section we saw that one-element sets can be characterized by describing what the maps into them look like. (In particular, we showed that a set has one element if and only if there is exactly one map into it from every other set.) It turns out that a very similar idea will give us a way to describe products of sets in a map-focused way.

2.1 The Universal Property of Products

If A and B are sets, what does it mean to define a function $f : Z \to A \times B$? For any element $z \in Z$, we need an ordered pair of elements, one from A and one from B. In other words, if we had two functions $f_1 : Z \to A$ and $f_2 : Z \to B$, we could define $f(z) = (f_1(z), f_2(z))$. Moreover, this process is reversible, since you can recover f_1 and f_2 from f by composing with the *projection maps* $\pi_1 : A \times B \to A$ and $\pi_2 : A \times B \to B$ defined by $\pi_1((a,b)) = a$ and $\pi_2((a,b)) = b$; I encourage you to check that, if f is defined as above, then $f_1 = \pi_1 \circ f$ and $f_2 = \pi_2 \circ f$.

All together, this means that *defining a map into $A \times B$ is the same as defining two maps, one into A and one into B.* This inspires the following definition:

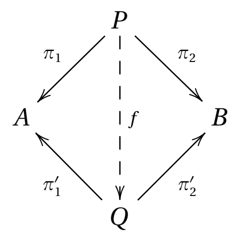
Definition 2.1. If A and B are two objects in a category, a **product** of A and B is an object $A \times B$, together with "projection maps" $\pi_1 : A \times B \to A$ and $\pi_2 : A \times B \to B$, satisfying the following property: for any other object Z with maps $f_1 : Z \to A$ and $f_2 : Z \to B$, there is a unique map $f : Z \to A \times B$ such that $f_1 = \pi_1 \circ f$ and $f_2 = \pi_2 \circ f$.



It's very common in category theory to express universal properties in terms of diagrams like this one. When all chains of arrows connecting the same pair of objects are equal to each other, we say that the diagram **commutes** — in this case, that amounts to the two equalities $f_1 = \pi_1 \circ f$ and $f_2 = \pi_2 \circ f$. So we could have described the property that f has to have simply by saying something like "there is a unique map $f : Z \to A \times B$ making this diagram commute."

The discussion right before Definition 2.1 shows that in the category of sets, $A \times B$, together with the projection maps, is indeed a product — in the categorical sense — of A and B. (If it wasn't, we probably would have had to give the categorical version a different name!) Like Definition 1.7, our definition of terminal objects, this is a universal property, and, as will be true of all the universal properties in this class, this means it's unique. The proof will be pretty similar.

Theorem 2.2. *Let A and B be two objects in a category, and suppose P and Q are both products of A and B, with projection maps $\pi_1 : P \to A$, $\pi_2 : P \to B$, $\pi_1' : Q \to A$, and $\pi_2' : Q \to B$. Then P and Q are isomorphic, and there is a unique isomorphism $f : P \to Q$ which makes the following* *diagram commute (that is, for which* $\pi_1 = \pi_1' \circ f$ *and* $\pi_2 = \pi_2' \circ f$*):*

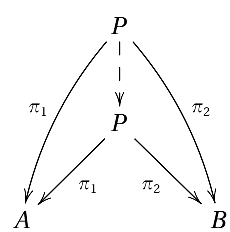


Before diving into the proof, it's worth thinking just a bit about what this last condition is saying and why we might want it. It's useful to think of the projection maps π_1 and π_2 as a part of the structure of the product — they are, after all, the only connection between the product and the two objects that it's supposed to be the product *of*! A nice way to think about this last condition in the theorem statement, then, is that our isomorphism is required to also preserve these projection maps. That is, when we say "products are unique up to isomorphism," we mean not only that the object is unique but that the projection maps are unique as well.

Proof of Theorem 2.2. Just like in the proof of Theorem 1.8, we'll first produce f and its inverse using both universal properties, and then we'll show that they compose to the identity both ways.

First, using the fact that Q is a product (with P in the role of Z) we get a unique map $f : P \to Q$ for which $\pi_1 = \pi_1' \circ f$ and $\pi_2 = \pi_2' \circ f$. (Note that this handles the uniqueness part of the claim about f already!) Switching P and Q and using the fact that P is a product gives us a unique map $g : Q \to P$ for which $\pi_1' = \pi_1 \circ g$ and $\pi_2' = \pi_2 \circ g$. We just need to show that these are inverses.

First, consider $g \circ f : P \to P$. The equations in the previous paragraph imply that $\pi_1 \circ g \circ f = \pi_1' \circ f = \pi_1$, and similarly for π_2. In other words, we can stick $g \circ f$ in as the dotted line in this diagram and it will commute:



But the identity map $\mathrm{Id}_P : P \to P$ *also* makes this diagram commute! Because P is a product (applying definition one more time with P itself in the role of Z) there can only be one such map, which means that in fact $g \circ f = \mathrm{Id}_P$. An exactly analogous argument shows that $f \circ g = \mathrm{Id}_Q$. □

Just as for terminal objects, as long as we are content to only pick out objects up to isomorphism, it makes sense to talk about "the" product of a pair of objects in a category rather than just "a" product, and this is in fact the usual convention. With this understanding, we'll use the notation $A \times B$ to refer to the product of A and B, even in categories where the objects aren't sets. The same warning about the *existence* of terminal objects applies to products as well: Theorem 2.2 doesn't guarantee that every pair of objects has a product, just that if two objects *do* have a product then that product is unique up to isomorphism.

2.2 Using the Universal Property

As alien as it might seem, the universal property of the product from Definition 2.1 can be used to prove some basic properties of products. If all we cared about were sets, then all of these properties could also be proved by looking directly at elements, but there's an advantage to proving them the way we do here: if you can manage to construct a proof that only uses maps and universal properties, then it applies to products in *all* categories, not just Set!

Proposition 2.3. *Products are commutative, that is, for any objects A and B in a category, $A \times B \cong B \times A$. (Recall that "$\cong$" means "is isomorphic to.")*

Proof. Because of Theorem 2.2, it's enough if we show that $A \times B$ satisfies the universal property characterizing a product of B and A. But this is immediate from the definition: if you switch the positions of A and B along with the projection maps π_1 and π_2, the entire definition is completely symmetrical, so anything that satisfies one version satisfies the other. □

Proposition 2.4. *Let A be an object in a category with products, and suppose that category has a terminal object T. Then $A \times T \cong A$*

Proof. To prove this, it's enough to show that A satisfies the definition of a product of A and T. To do this, we need product maps $\pi_1 : A \to A$ and $\pi_2 : A \to T$. We'll take π_1 to be the identity Id_A, and π_2 to be the unique map $A \to T$ that we get from the fact that T is terminal. We now need to check the universal property.

So, suppose we have some Z with maps $f_1 : Z \to A$ and $f_2 : Z \to T$. Can we find a map $f : Z \to A$ for which $f_1 = \pi_1 \circ f$ and $f_2 = \pi_2 \circ f$? Yes, we can just take $f = f_1$! The first equality is then true because π_1 is the identity, and the second is true because, by the fact that T is terminal, there is only one map $Z \to T$, and so f_2 and $\pi_2 \circ f$ must both be equal to that map. □

It's worth taking a moment to see why Proposition 2.4 is true in the category of sets: if T is a one-element set, why is there always a bijection from $A \times T$ to A?

Another nice perspective on this proof is to think about the way we originally summarized the universal property of products: giving a map into $A \times T$ is the same as giving two maps, one into A and one into T. If T is the terminal object, then specifying a map into T gives no information at all, since every object has exactly one. So the work of specifying both a map into A and a map into T is exactly the same work as just specifying a map into A.

Exercises

Exercises that are especially *important* are marked with ⚠. Exercises that are especially *challenging* are marked with (⋆).

2.1. ⚠*[This is one of the most important exercises to work through!]*

(a) Suppose A, B, A', B' are sets and $f : A \to A'$ and $g : B \to B'$ are functions. Come up with a natural way to produce a function $A \times B \to A' \times B'$. We'll call this function $f \times g$.

(b) Now suppose instead that A, B, A', B' are objects in an arbitrary category and $f : A \to A'$ and $g : B \to B'$ are maps. Use the universal property of products to produce a map $f \times g : A \times B \to A' \times B'$, and check that in the category of sets it coincides with the map you constructed in the previous part.

(c) Prove that, in an arbitrary category, $\mathrm{Id}_A \times \mathrm{Id}_B = \mathrm{Id}_{A\times B}$.

(d) Now, in addition to the objects and maps described earlier, suppose we also have objects A'' and B'' and maps $f' : A' \to A''$ and $g' : B' \to B''$. Prove that $(f' \times g') \circ (f \times g) = (f' \circ f) \times (g' \circ g)$.

2.2. Suppose A, B, C are objects in a category, $A \cong B$, and the products $A \times C$ and $B \times C$ exist. Prove that $A \times C \cong B \times C$.

2.3. (a) Come up with a definition for a product of *three* objects, and state and prove a version of Theorem 2.2 for these triple products.

(b) Use this to prove that products are associative (in the same sense in which we proved they are commutative in Proposition 2.3).

2.4. Does that category Nat (defined in Example 1.5) have products? If do, what do they look like?

2.5. *[Don't worry about this problem unless you've seen some group theory.]* Does the category of groups (defined in Exercise 1.5) have products? If so, what do they look like?

2.6. *[Don't worry about this problem unless you've seen some linear algebra.]* Does the category of vector spaces (defined in Example 1.4) have products? If so, what do they look like?

2.7. Recall that, in Exercise 1.10, you constructed a *slice category* $\mathscr{C}/A$ out of a category $\mathscr{C}$ and an object A. (If you haven't done that exercise yet, do it before this one!)

(a) Suppose instead you start with *two* objects A, B from $\mathscr{C}$. Come up with a definition of a category $\mathscr{C}/(A, B)$ whose objects are triples (C, π_C^A, π_C^B), where C is an object of $\mathscr{C}$ and $\pi_C^A : C \to A$ and $\pi_C^B : C \to B$ are maps.

(b) Show that a terminal object in $\mathscr{C}/(A, B)$ is the same as a product of A and B in $\mathscr{C}$, and that therefore you can think of Theorem 2.2 as a special case of Theorem 1.8.

2.8. ($\star$) Recall that, in Exercise 1.9, you built a function we called $f_* : \mathrm{Maps}(C, A) \to \mathrm{Maps}(C, B)$ for any map $f : A \to B$ and any object C. (If you haven't done that exercise yet, do it before this one!)

(a) We say that f is a **monomorphism** if $f_* : \mathrm{Maps}(C, A) \to \mathrm{Maps}(C, B)$ is an injective function for every object C. (More explicitly, f is a monomorphism if, for any maps $g, h : C \to A$, $f \circ g = f \circ h$ implies $g = h$.) Show that, in the category of sets, f is a monomorphism if and only if it is an injective function.

(b) In that problem you also constructed a map $f^* : \mathrm{Maps}(B, C) \to \mathrm{Maps}(A, C)$. We say that f is an **epimorphism** if f^* is injective for every C. Show that, in the category of sets, f is an epimorphism if and only if it is a surjective function.

(c) Show that, in an arbitrary category, f_* is surjective for every C if and only if f has a *right inverse*, that is, some map $g : B \to A$ such that $f \circ g = \mathrm{Id}_B$. If this happens, we say f is **right-invertible**.

(d) Similarly, show that f^* is surjective if and only if f is **left-invertible**.

(e) Show that, in the category of sets, f is a monomorphism if and only if it is left-invertible, and it is an epimorphism if and only if it is right-invertible. *[Note: One of these four implications requires the Axiom of Choice! Can you see which one?]*

(f) Find an example of a category where these equivalences do not hold. *[Hint: For one of them, you can use the category* Nat *from Example 1.5.]*

3 Initial Objects, Coproducts, and Duality

We've now succeeded in describing one-element sets and products in our new categorical language, and both descriptions come from focusing on what functions *into* the relevant set look like. One might wonder at this point how much of set theory can be captured in this way. For example, can we (to ask an extremely leading question) characterize the empty set in terms of what the functions into it look like?

3.1 The Empty Set as an Initial Object

A moment of thought might lead you to the conclusion that this is not going to work so well, for the simple reason that there aren't all that many functions into the empty set! It can be a bit confusing to think about functions when they involve the empty set, so it's worth thinking through carefully. In order to specify a function from set A to a set B, we need, for each element of A, to pick an element from B for it to map to. If A is nonempty and B is empty, then A has at least one element, say $a \in A$, but since B is empty there is nowhere a can be sent, and therefore there are no functions from A to B at all.

If A is empty, though, then no matter what B is, empty or not, it's trivial to specify a function from A to B: A has no elements, so our job is complete before we've even begun. There is, moreover, a *unique* such function, which you should take a moment to convince yourself of if it's not clear. In other words, while there are no functions from any nonempty set *into* the empty set, there is exactly one function *from* the empty set to any other set.

This inspires the following definition:

Definition 3.1. An object A in a category $\mathscr{C}$ is called **initial** if, for any other object B of $\mathscr{C}$, there is exactly one map $A \to B$.

The preceding discussion shows that, in the category of sets, the empty set is an initial object. Just as for terminal objects, initial objects in other categories might look different, and some categories might fail to have any initial objects at all.

Definition 3.1 is a universal property, but of a slightly different type than the two we've studied so far, since it's about maps out of the relevant object rather than maps into it. But we can still show that initial objects are unique using an argument very similar to the ones we've already seen.

Theorem 3.2. *Suppose A and A' are two initial objects in the same category. Then A and A' are isomorphic, and there is only one isomorphism between them.*

Proof. We can use the fact that A' is initial to get a map $f : A' \to A$. Notice that, by Definition 3.1, f is actually the *only* map from A' to A, so if A and A' are going to be isomorphic, this map we have here had better be the isomorphism! We can get a potential inverse for it by using the fact that A is initial to get a map $g : A \to A'$; we just have to show that $f \circ g = \mathrm{Id}_A$ and $g \circ f = \mathrm{Id}_{A'}$.

Notice that $f \circ g$ is a map from A to A. A is initial, so (applying the definition with $B = A$) there is only *one* map from A to A. But Id_A is *also* a map from A to A, which means that we must have $g \circ f = \mathrm{Id}_A$! A basically identical argument shows that $g \circ f = \mathrm{Id}_{A'}$. □

3.2 Duality

In fact, the proof of Theorem 3.2 is not just *similar* to the proof of our earlier theorem (Theorem 1.8) about terminal objects. I literally wrote it by copying and pasting the earlier proof, then reversing the order of all the arrows and compositions and changing the word “terminal” to “initial” everywhere.

I did this in order to illustrate an important general principle: whenever you have a theorem (like Theorem 3.2) that holds in an arbitrary category, you can produce another such theorem by reversing the order of all the arrows and all the compositions. One way to see that this ought to be true is just to look at our original definition of a category, Definition 1.1. Notice that everything about that definition is symmetrical: everything about maps from A to B is equally true about maps from B to A, and everything about function compositions in one order is equally true of compositions in the other order (provided you reverse the direction of the arrows at the same time). You’ll work out a more rigorous version of this argument in Exercise 3.5.

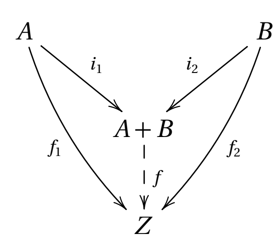
This phenomenon is called **duality**, and it’s a very powerful tool in category theory. It’s common to say two concepts are “duals” of each other if they are related by this arrow-reversal operation — for example, initial and terminal objects are duals. If one of the two concepts is more well-known than the other — or if it just got the name first — it’s common to name the lesser-known one by sticking the prefix “co” on the better-known one. So if initial objects didn’t have their own name already, we might call them “coterminal objects”. (And now you are equipped to understand the hilarious joke about how comathematicians are devices for turning cotheorems into ffee.)

It’s worthwhile to keep in mind one limitation of this philosophy though: while it’s true that, for example, you can dualize any statement about categories with initial objects to get a corresponding statement about categories with terminal objects, there is no guarantee that any *particular* category which has initial objects will also itself have terminal objects! Situations like this can and will break the symmetry we’re talking about here. You’ll see an example of this in Exercise 3.6.

3.3 Coproducts

What do we get when we dualize products, the only other construction we’ve seen so far? If we reverse the arrows and compositions in Definition 2.1, we get:

Definition 3.3. If A and B are two objects in a category, a **coproduct** of A and B is an object $A+B$, together with “inclusion maps” $i_1 : A \to A+B$ and $i_2 : B \to A+B$, satisfying the following property: for any other object Z with maps $f_1 : A \to Z$ and $f_2 : B \to Z$, there is a unique map $f : A+B \to Z$ such that $f_1 = f \circ i_1$ and $f_2 = f \circ i_2$.



What do coproducts look like in the category of sets? The answer involves a construction that might be somewhat unfamiliar (which is in fact the main reason I introduced products first):

Definition 3.4. If A and B are two sets, the **disjoint union** of A and B is the set

$$A+B=\{(a,0): a\in A\}\cup\{(b,1): b\in B\}.$$

The name "disjoint union" is meant to be suggestive of the point of the construction. It's like the union, except we take pains to guarantee that none of the elements from A can overlap with any of the elements from B — that is, to make the two sets disjoint. Because we only care about objects in a category up to isomorphism, the particular way we achieve this is much less important than the result. Our strategy here, of "tagging" the elements of A and B with 0 and 1 to indicate where they came from, is just one strategy of many that could work.

It's worthwhile to check that this definition actually produces a coproduct in the category of sets.

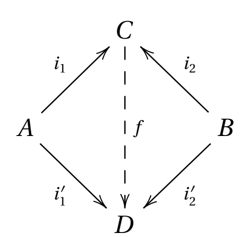
Proposition 3.5. *The category of sets has coproducts. The coproduct of two sets A and B is the disjoint union $A+B$, together with the inclusion maps given by $i_1(a)=(a,0)$ and $i_2(b)=(b,1)$.*

Proof. We just have to check that $A+B$, i_1, and i_2 satisfy the universal property. So suppose we're given another set Z and functions $f_1: A\to Z$ and $f_2: B\to Z$. How are we going to build an f that makes the diagram in Definition 3.3 commute? Because we need $f\circ i_1=f_1$, we're forced to set $f(a,0)=f_1(a)$ for every $a\in A$. Similarly, we need $f(b,1)=f_2(b)$ for every $b\in B$.

Notice, however, that every element of $A+B$ is either of the form $(a,0)$ or $(b,1)$, but never both, so these two requirements completely specify f! We therefore see that not only is there an f that makes the diagram commute, but also that there is only *one* possible f, which is exactly what we needed to prove. □

In the last section, after we showed that products were unique, we also used the universal property of products to prove some basic facts, like that products are commutative and that terminal objects act as the identity for products. At this point we might have started proving things about coproducts if not for a very fortunate fact about our situation: because of the duality principle, *we get all of those facts for free* just by dualizing the corresponding statements about products! For example, without doing any further work, we can now conclude:

Theorem 3.6. *Let A and B be two objects in a category, and suppose C and D are both coproducts of A and B, with inclusion maps $i_1: A\to C$, $i_2: B\to C$, $i_1': A\to D$, and $i_2': B\to D$. Then C and D are isomorphic, and there is a unique isomorphism $f: C\to D$ which makes the following diagram commute:*



Proposition 3.7. *Coproducts are commutative, that is, for any objects A and B in a category, $A+B\cong B+A$.*

Proposition 3.8. *Let A be an object in a category with coproducts, and suppose that category has an initial object I. Then $A+I\cong A$*

There is a useful perspective on these facts (and the corresponding facts about products) that comes from thinking about what they mean in the category of sets. Two sets are isomorphic if and only if they have the same number of elements, and so you can think about all of these statements in terms of what they tell you about the number of elements in a finite set. For example, if A and B are finite, it's easy to see that the size of the disjoint union $A+B$ is the sum of sizes of A and B. And since initial objects in Set are empty, they have zero elements. So, when applied to finite sets, Proposition 3.8 becomes the fact that, for any natural number a, $a+0=a$; and Proposition 3.7 says that addition of natural numbers is commutative. Similar statements, of course, hold about products and multiplication.

These statements are, of course, much more general than this, since they apply in an arbitrary category! Still, it can be useful to think about the "arithmetic" versions of statements about products and coproducts, if only to help you remember which statements you should expect to be true. (This is, in fact, the reason I've used the notation + for coproducts and disjoint unions, which is slightly nonstandard — the more common symbol is $\amalg$.) We'll explore this perspective more deeply in the next section.

Exercises

Exercises that are especially *important* are marked with ⚠. Exercises that are especially *challenging* are marked with $(\star)$.

3.1. Write down proofs of Theorem 3.6 and Propositions 3.7 and 3.8 by dualizing the proofs for products from the previous section *[Feel free to only do this until you are convinced you know how to dualize one of these proofs.]*

3.2. Suppose $\mathscr{C}$ is a category with both an initial object I and a terminal object T. Show that if there is a map $f : T \to I$, then f is an isomorphism, and therefore both I and T are both initial and terminal.

3.3. *[Don't worry about this problem unless you've seen some linear algebra.]* What do initial objects and coproducts look like in the category of vector spaces? Compare this to your answers to Exercises 1.6 and 2.6.

3.4. *[Don't worry about this problem unless you've seen some group theory.]*

(a) What do initial objects look like in the category of groups?

(b) We'll write Ab for the **category of abelian groups**, which is just like the category of groups except that we only include the abelian groups. Show that initial objects, terminal objects, and products in Ab are the same as the ones in Grp, but that the product of two groups is *also* a coproduct in Ab.

(c) $(\star)$ Prove that the product of two groups is *not* a coproduct in the category of groups. *[Hint: Show that $\mathbb{Z} \times \mathbb{Z}$ is not a coproduct of $\mathbb{Z}$ and $\mathbb{Z}$ (with inclusion maps like in the previous part) by finding a group G and two maps $\mathbb{Z} \to G$ for which there is no map $\mathbb{Z} \times \mathbb{Z} \to G$ making the coproduct diagram commute.]*

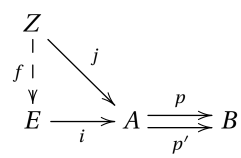
[Note: The category of groups does have coproducts, but the construction is very nonobvious, so I'm not asking about it here. I'm happy to talk about it more in person if you're interested though!]

3.5. ⚠Let $\mathscr{C}$ be a category. Construct a category $\mathscr{C}^{\text{op}}$ whose objects are the same as the objects of $\mathscr{C}$ but in which the direction of arrows and the order of composition are reversed. (This is called the **opposite category**.) How can you use this to make the duality principle more precise?

3.6. One of the categories we've been using as a running example has initial objects but not terminal objects. Which one?

3.7. ⚠Suppose A and B are two objects in some category, and $p, p' : A \to B$ are two maps. We say that an object E, together with a map $i : E \to A$, is an **equalizer** of p and p' if it satisfies the following two properties:

- $p \circ i = p' \circ i$
- Given any other object Z and any map $j : Z \to A$ for which $p \circ j = p' \circ j$, there is a unique map $f : Z \to E$ for which $j = i \circ f$:



(a) Describe what equalizers look like in the category of sets.

(b) *[Don't worry about this problem unless you've seen some group theory.]* Suppose $f : G \to H$ is a group homomorphism. Express the kernel $\ker f$ as an equalizer.

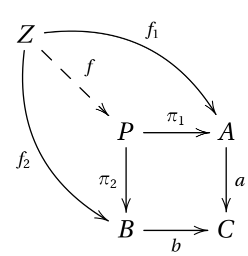
(c) The dual of an equalizer is called a **coequalizer**. Write down the definition of a coequalizer by dualizing the definition of equalizers given above.

(d) (⋆) What do coequalizers look like in the category of sets?

(e) (⋆) *[Don't worry about this problem unless you've seen enough group theory to know what a quotient group is.]* If you take your answer to the part about kernels above and dualize it, the resulting object is called a **cokernel**. How would you describe the cokernel of a group homomorphism in terms of the original map?

3.8. Let A, B, C be objects in some category, and suppose we're given two maps $a : A \to C$ and $b : B \to C$. We say that an object P, together with maps $\pi_1 : P \to A$ and $\pi_2 : P \to B$, is a **pullback** if it satisfies the following two properties:

- The diagram formed by the arrows a, b, π_1, π_2 commutes. (That is, $a \circ \pi_1 = b \circ \pi_2$.)
- If Z is any other object with maps $f_1 : Z \to A$ and $f_2 : Z \to B$ for which $a \circ f_1 = b \circ f_2$, then there is a unique map $f : Z \to P$ making the following diagram commute:



(The only additonal requirements this imposes beyond the ones appearing earlier are that $\pi_1 \circ f = f_1$ and $\pi_2 \circ f = f_2$.)

If this is all true, we write $P = A \times_C B$. Note that the notation $A \times_C B$ doesn't include the maps a and b, even though they are relevant! Sometimes, to pack everything into the terminology, we might call it "the pullback of the diagram $A \xrightarrow{a} C \xleftarrow{b} B$."

(a) Prove that pullbacks are unique in any category.

(b) Suppose S is a set, A and B are subsets of S, and $i_1 : A \to S$ and $i_2 : B \to S$ are the *inclusion functions*, that is, the functions defined by $i_1(a) = a$ and $i_2(b) = b$. Show that $A \cap B$ is a pullback of the diagram $A \xrightarrow{i_1} S \xleftarrow{i_2} B$.

(c) Now suppose $a : A \to C$ and $b : B \to C$ are any functions at all between sets A, B, C. Prove that the pullback still exists in the category of sets. What does it look like? *[Mild hint: the pullback in the category of sets sometimes goes by the name* ***fibered product****. If this phrase isn't familiar to you, then there's a good chance this isn't a set-theoretic construction you've ever seen before!]*

(d) The dual of a pullback is called a **pushout**. Write out the definition of a pushout by dualizing the definition of pullback given above.

(e) (⋆) What do pushouts look like in the category of sets?

4 Exponential Objects

At the end of the last section we saw how, when you apply them to finite sets, many of the facts that we proved about products and coproducts and initial and terminal objects amount to simple statements about arithmetic on natural numbers. If you did the exercises, you even proved a couple more. It makes sense, then, to ask how far we can take this. Is it possible to "categorify" everything about arithmetic in this way?

4.1 Categorifying Distributivity

Probably the next most complicated fact about arithmetic that we haven't established yet is the distributive law — that, for any natural numbers a, b, c, we have $a(b+c) = ab + ac$. The categorical version of this statement would be that for any objects A, B, C in a category with both products and coproducts, then

$$(A+B) \times C \cong (A \times C) + (B \times C).$$

Let's first see if we can build a bijection like this in the category of sets. Using our "tagged ordered pairs" construction of disjoint unions, an element of $(A \times C) + (B \times C)$ is either of the form $((a,c),0)$ for some $a \in A$ and $c \in C$, or it's $((b,c),1))$ for some $b \in A$ and $c \in C$. In the former case, we can send it to $((a,0),c) \in (A+B) \times C$, and in the latter case we can send it to $((b,1),c)$. The inverse function can be constructed very similarly, and it's a quick exercise to show that they are indeed inverses of each other.

So far so good then! Can we turn this construction into something that works in an arbitrary category with products and coproducts? To do this, we'll somehow need to use the universal property to construct maps in both directions. If you try to do this, though, you will find that it's perfectly possible to build the map $(A \times C) + (B \times C) \to (A+B) \times C$ and check that in Set it's the same as the map we described in the previous paragraph. Going the other way presents a problem though: we need a map $(A+B) \times C \to (A \times C) + (B \times C)$, but the universal properties of products and coproducts don't tell us anything about maps *out of* a product or *into* a coproduct!

This is a fatal problem with the plan — it's impossible to accomplish this goal with the tools we have on the table right now. One way to really drive the point home is to consider what happens when we dualize this statement. Suppose we had a proof of the distributive law as stated above. We could then reverse all the arrows and compositions to get a proof of the dual statement, that for any objects A, B, C category with both products and coproducts, we have

$$(A \times B) + C \cong (A+C) \times (B+C).$$

(Notice that the hypothesis that the category has products and coproducts is *self-dual* — dualizing simply interchanges products and coproducts.) This statement is definitely false for finite sets!

What's going on? Why were we able to "categorify" the distributive law as far as the category of sets, but not as far as an arbitrary category with products and coproducts? What we learn from the argument in the previous paragraph is that the answer is going to have to lie in some property of the category of sets that *isn't* self-dual, something that breaks the symmetry between products and coproducts. This extra piece of structure is what we're going to study today.

4.2 Exponential Objects

Just as products and coproducts of sets are analogous to products and sums of natural numbers, the new piece of structure that will solve our distributivity problem turns out to be analogous to *powers* of natural numbers.

Definition 4.1. For any two sets A, B, we will write B^A for the set of all functions from A to B.

Notice the order in which B and A appear in the notation. The reason it's written that way comes from thinking about the size of B^A when A and B are both finite: if A has size a and B has size b, then there are b^a functions from A to B, since you need to choose from b possible places to send each element of A, and you have to make this choice a times.

Describing B^A in terms of a universal property is more complicated than it was for the examples we've seen so far. We'll describe our goal informally first, and then see how to formalize it. The key idea is to think about what you can do with a function $g : Z \to B^A$. Such a function takes an element of Z and produces a way to take an element of A to an element of B. So if you're given an element of Z and an element of A, you can use g to produce an element of B: first you plug in z to get a function $g(z) : A \to B$, and the you plug a into *that* function. In other words, from our original function g, we've produced a *new* function $g^\sharp : Z \times A \to B$ by the somewhat-confusing-to-read rule $g^\sharp(z, a) = g(z)(a)$. It's a nice exercise to think about how you would explicitly go the other way, from a function $h : Z \times A \to B$ to a function $h^\flat : Z \to B^A$; if you do, you'll see that these two operations are inverses of each other.

This is the idea we want to formalize in a universal property: that specifying a function $Z \to B^A$ is the same as specifying a function $Z \times A \to B$. When we did this for products, the corresponding statement was that specifying a function $Z \to A \times B$ is the same as specifying a function $Z \to A$ and a function $Z \to B$. Recall, though, that for products we also needed the diagram formed from these three functions and the projection maps to commute. (Look back at Definition 2.1.) What should the analogue of that condition be here, and how do we write that condition referring only to functions and not to elements?

Here is one way to write it down. Consider the "evaluation function" $\mathrm{ev} : B^A \times A \to B$ which is defined by the rule $\mathrm{ev}(f, a) = f(a)$. You can think of this as serving a similar role to the projection maps in the product — the ability to evaluate them at an element of A is what makes the elements of B^A actual functions instead of just elements of some random set. Our rule from a couple paragraphs up can then be written $g^\sharp(z, a) = \mathrm{ev}(g(z), a)$.

This takes us halfway to our goal of writing $g^\sharp$ without referring to elements; we'll be done if we can find a way to write the function that takes (z, a) to $(g(z), a)$. To do this, recall from Exercise 2.1 that, given any maps $p : A \to A'$ and $q : B \to B'$, the universal property of products lets us produce a map $p \times q : A \times B \to A' \times B'$ which, in the category of sets, is given by $(p \times q)(a, b) = (p(a), q(b))$. The function we're after, then, is just $g \times \mathrm{Id}_A$! All together, then, we can express our construction for sets as $g^\sharp = \mathrm{ev} \circ (g \times \mathrm{Id}_A)$, and the thing that makes B^A an exponential object is that this function — the one sending g to $g^\sharp$ — is a bijection.

That was a lot of work, but the end result is that we now have a definition that we can use in any category:

Definition 4.2. Suppose A and B are objects in a category with products. An **exponential object** from A to B is an object B^A together with a map $\mathrm{ev} : B^A \times A \to B$ satisfying the following property: for any map $h : Z \times A \to B$ there exists a unique map $h^\flat : Z \to B^A$ such that $h = \mathrm{ev} \circ (h^\flat \times \mathrm{Id}_A)$.

Using our earlier notation, we could write this last condition as $h = (h^\flat)^\sharp$. The discussion above implies that Set has exponential objects. In the exercises you'll show that, just like every other object we've defined with a universal property, exponential objects are unique up to isomorphism.

Just like every other universal property definition we've introduced, nothing about Definition 4.2 implies that exponential objects have to exist. In fact, unlike all those other examples, the existence of exponential objects is relatively rare! They exist in Set, but in many common categories they don't.

4.3 Exponentials Imply Distributivity

We started this section by talking about how to get a categorical version of the distributive law, and we were led to ask what it was about the category of sets that meant we have a distributive law there but not in an arbitrary category with products and coproducts. The answer is, in my opinion, somewhat surprising — the mere *existence* of exponential objects in Set turns out to be the property we need, even though the statement of distributivity has nothing to do with them! Let's see how this works.

Theorem 4.3. *Suppose A, B, and C are objects in a category with products, coproducts, and exponential objects. Then* $(A+B)\times C \cong (A\times C)+(B\times C)$.

Proof. Our strategy will be to show that $(A+B)\times C$ satisfies the universal property that makes it a coproduct of $A\times C$ and $B\times C$. First we need to specify the inclusion maps. Letting $i_1 : A \to A+B$ and $i_2 : B \to A+B$ be the inclusion maps for $A+B$, we can build our inclusion maps using the "products of maps" construction from Exercise 2.1 we used earlier: they'll be $j_1 = i_1 \times \mathrm{Id}_C : A\times C \to (A+B)\times C$ and $j_2 = i_2 \times \mathrm{Id}_C : B\times C \to (A+B)\times C$.

Now, suppose we're given maps $f_1 : A\times C \to Z$ and $f_2 : B\times C \to Z$. We need to show that there is a unique map $f : (A+B)\times C \to Z$ such that $f_1 = f \circ j_1$ and $f_2 = f\circ j_2$. Notice that when we looked at this task earlier, we had to stop here, since nothing gave us any way to build maps out of a product. But now we have such a way, because maps $(A+B)\times C \to Z$ are in one-to-one correspondence with maps $A+B \to Z^C$! Building f will just be a matter of chasing through all the universal properties in front of us in order.

First, from $f_1 : A\times C \to Z$ we can construct $f_1^\flat : A \to Z^C$, and similarly $f_2^\flat : B \to Z^C$. From the universal property of the coproduct $A+B$, we get a map $g : A+B \to Z^C$ with $f_1^\flat = g\circ i_1$ and $f_2^\flat = g \circ i_2$. The map f we want is then $g^\sharp : (A+B)\times C \to Z$.

In order to be done, we need to check that $f_1 = g^\sharp \circ j_1$ and $f_2 = g^\sharp \circ j_2$, and that $g^\sharp$ is the only map with this property. Proving the first two properties requires unwrapping all the definitions one more time. We have

$$g^\sharp \circ j_1 = \mathrm{ev}\circ(g\times \mathrm{Id}_C)\circ(i_1\times\mathrm{Id}_C) = \mathrm{ev}\circ((g\circ i_1)\times \mathrm{Id}_C).$$

We said earlier that $g\circ i_1 = f_1^\flat$, so this last expression equals $\mathrm{ev}\circ(f_1^\flat \times \mathrm{Id}_C) = (f_1^\flat)^\sharp = f_1$, which is what we needed. The argument for j_2 is basically identical.

The only thing left to show is uniqueness, which you'll do in Exercise 4.2. □

Recall from the beginning of this section that we needed a way to break the symmetry that would have given us the "co-distributive law" $(A\times B)+C \cong (A+C)\times(B+C)$. The existence of exponential objects is exactly that — the reason we can't dualize the proof of Theorem 4.3 for the category of sets is this would require Set to have *coexponential objects*, that is, objects which satisfy Definition 4.2 but with all the arrows and compositions reversed and the products replaced with coproducts. The fact that Set has exponentials but not coexponentials is exactly what breaks the symmetry.

Exercises

Exercises that are especially *important* are marked with ⚠. Exercises that are especially *challenging* are marked with (⋆).

4.1. Prove that, as with all of the objects we've constructed using universal properties so far, exponential objects are unique up to a unique isomorphism. (In particular, formulate precisely what that means in this case!)

4.2. Fill in the missing step — the uniqueness of the map we constructed — in the proof of Theorem 4.3.

4.3. This is a list of facts about exponentiation of natural numbers. For each one, write down the corresponding fact about exponential objects in an arbitrary category, and then prove it using the relevant universal properties.

(a) $a^1 = a$

(b) $1^a = 1$

(c) $b^{ca} = (b^c)^a$

(d) $(bc)^a = b^a c^a$

(e) $a^0 = 1$

(f) $c^{a+b} = c^a c^b$

4.4. We can define **coexponential objects** in any category with coproducts by dualizing the definition of exponential objects: a coexponential object from A to B is any object B_A together with a map $\mathrm{coev} : B \to B_A + A$ with the property that, for any object Z and any map $f : B \to Z + A$, there is a unique map $f_\sharp : B_A \to Z$ such that $f = (f_\sharp + \mathrm{Id}_A) \circ \mathrm{coev}$. (In other words, maps $B \to Z + A$ are in bijection with maps $B_A \to Z$.) Prove directly that Set does not have coexponential objects. *[There is also an* indirect *proof which takes advantage of the fact that the codistributive law is false for sets!]*

4.5. Does the category Nat from Example 1.5 have exponential objects? What about coexponential objects? What do they look like?

4.6. Suppose A and B are objects in some category with a terminal object T. Let's say that a map $f : A \to B$ is *constant* if there exists a map $b : T \to B$ such that f is equal to the composition $A \to T \xrightarrow{b} B$, where the first arrow is the unique map arising from the definition of terminal objects.

(a) Prove that the constant maps in Set are exactly the functions that are constant in the ordinary sense.

(b) (⋆) Because $A \cong T \times A$, maps $f : A \to B$ are in bijection with maps $f^\flat : T \to B^A$. Construct a map $c : B \to B^A$ such that f is constant if and only if there exists a map $b : T \to B$ for which $f^\flat$ is equal to the composition $T \xrightarrow{b} B \xrightarrow{c} B^A$. *[Hint: First show that, in* Set*, you want $c(b)$ to be the constant function that always outputs b. If this is c, what is $c^\sharp$? The same formula for $c^\sharp$ will work in an arbitrary category. At one point, you may want to look back at the definition of $f \times g$ from Exercise 2.1.]*

5 Limits and Colimits

We've now seen a lot of examples of objects in categories that can be defined by universal properties — initial and terminal objects, products and coproducts, and, in the exercises, equalizers, coequalizers, pullbacks, and pushouts. If you spend some time looking at all of them at once, you'll find it's possible to divide them into two groups. Some of them (terminal objects, products, equalizers, and pullbacks) have a universal property that's about maps *into* the relevant object, and the rest (initial objects, coproducts, coequalizers, and pushouts) have a universal property that's about maps *out of* the relevant object.

The similarity in these definitions suggests that it might be possible with the right definition to view all of these constructions as special cases of one single concept. That's the task we'll take on in this section.

5.1 Diagrams and Cones

We'll start by describing the input data that these constructions will take in.

Definition 5.1. Let $\mathscr{C}$ be a category. A **diagram in** $\mathscr{C}$ is a set of objects of $\mathscr{C}$ together with some maps between them.

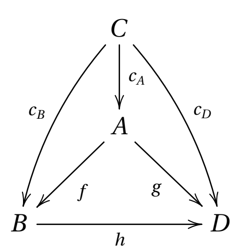
The objects and maps that appear in a diagram are completely unrestricted, so long as the source and target of every map is one of the chosen objects. There can be maps from an object to itself, and there can be more than one map between the same pair of objects. The set of objects could be empty, in which case there also can't be any maps. The diagram does not have to commute.

Definition 5.2. Let D be a diagram in a category $\mathscr{C}$. A **cone over** D consists of:

- An object C of $\mathscr{C}$, and
- for every object A appearing in the diagram D, a map $c_A : C \to A$,

with the property that, for every arrow $f : A \to B$ appearing in D, $f \circ c_A = c_B$.

The name "cone" comes from picturing C sitting "above" the diagram, with arrows going down from C to each of its objects. The requirement at the end of the definition amounts to requiring any triangles formed by two of these downward arrows and one arrow from the diagram to commute:



Note that this is the *only* commuting we require in this picture! There is, again, no requirement that the maps *within* the original diagram commute.

Recall that at the beginning of this section we split our universal properties into two buckets — the ones where the universal property is about maps *into* the object being defined and the one where it's about maps *out of* the object. Notice that all of the ones from the first bucket are cones! The product of A and B is a cone over the diagram consisting of just the objects A and B with no maps. Terminal objects are cones over the empty diagram. (It's a good exercise to work out how equalizers and pullbacks are cones. In particular, if you look at the definitions we gave earlier, you might notice that some of the maps required by the definition of cone seem to be missing; you should check in both cases that the missing arrows are uniquely determined by the definition of cone, and so nothing is lost if we leave them out.)

5.2 Limits and Colimits

Merely being a cone, though, is not the entirety of the definition of any of these concepts. The product of A and B, for example, is almost certainly not the only object that has maps to both A and B. We need some way to encode the universal property. To do this, it will be helpful to see that the cones over D in fact form a category in their own right. We just need to define what the maps are.

Definition 5.3. Let D be a diagram in a category $\mathscr{C}$. Suppose C and C' are cones over D, with maps $c_A : C \to A$ and $c'_A : C' \to A$ for every object of A. Then a **map of cones** from C to C' is a map $g : C \to C'$ for which $c'_A \circ g = c_A$ for every object A in the diagram. This definition of maps makes the cones over D into a category $\mathscr{C}/D$, which we call the **category of cones over** D.

You should convince yourself that this actually defines a category (that is, that it has identities and that composition works properly).

Now, with all this setup in place, the big unifying concept we've been alluding to actually has a fairly short definition.

Definition 5.4. Let D be a diagram in a category $\mathscr{C}$. A **limit** of D is a terminal object in the category of cones $\mathscr{C}/D$.

If you did Exercise 2.7, you already saw a way to express products as terminal objects. In fact, in a basically analogous way, all of the constructions in our first bucket are also limits! (It's worth working through exactly why this is true for equalizers and pullbacks.) Once we realize this, we get that the limit of D is unique up to isomorphism from just the fact that terminal objects are unique.

What about the second bucket, the ones that are defined by universal properties of maps *out of* them? It is probably not shocking that we can reach them by simply dualizing this definition!

Definition 5.5. Let D be a diagram in a category $\mathscr{C}$. A **cocone under** D consists of:

- An object K of $\mathscr{C}$, and
- for every object A appearing in the diagram D, a map $k_A : A \to K$,

with the property that, for every arrow $f : A \to B$ appearing in D, $k_B \circ f = k_A$.

We can form the **category of cocones under** D, written $\mathscr{C}\backslash D$ in a similar way, and a **colimit** of D is an initial object in this category.

This takes care of our second set of definitions — initial objects, coproducts, coequalizers, and pushouts are all colimits.

The one definition we've seen in this class that we haven't subsumed in this machinery is that of exponential objects. (In fact, the category of sets actually has *all* limits and colimits, but we've already seen that it doesn't have coexponential objects.) We haven't had time in this short class to talk about *functors* or *natural transformations*; if we had, we would see that Definition 4.2 is an example of something called an *adjunction*. I'm happy to talk to any of you about this if you're interested!