Lesson 1

Sam and Dennis

January 5, 2019

1 Introduction

Welcome to Math 98! We hope you'll want to stay in this class as much we want to facilitate it.

When people think of math, they usually think of numbers and calculations. One of the messages we hope to convey is that math majors (not us anyways) aren't training to become TI 84's. In one sense, its an attempt to capture some of the underlying structures of the world around us while abstracting away all the complicated details so that one can use logical analysis. Structures like sets, topological spaces, groups, manifolds, and categories can help us understand the world around us (as well as the world inside us: mathematics is often used in philosophy and also to study other mathematics!). In this course, we will focus on the last of the mentioned structures, which is quite a (relatively) recently discovered structure. Categories tend to get ignored in the undergraduate classes at Berkeley, although this may depend on whose 104, 110, or 113 section you take. We hope to shed some light on this fascinating subject and show how it links many areas of mathematics together.

1.1 Quick Logistics

Our course will hopefully be very light for all of you: we only require you to do the weekly readings and to submit a question on the readings you did. We're hoping that people will find something that they are personally interested in and can follow along from their own perspective. To that end, there will be a short final report where you can express your interests. The format doesn't *have to be* something like an essay. Just as a reminder, the course website is here.

2 Problem of Equality

Before we talk about categories, let us first briefly mention some of the reasons that it exists in the first place. You will hopefully read about some of these problems in the assigned readings.

Let's begin with equality. This is the simplest concept in math, right? We know it satisfies three important properties: reflexitivity ($x = x$), symmetry ($x = y$ implies $y = x$), and transitivity ($x = y$ and $y = z$ imply $x = z$).

We know equality of numbers extremely well. Everyone knows what $5 = 2 + 3$ means, after all. However, sometimes this notion is too strict. Think of dollar bills: there are millions of them around the world, and techinically, none of them are equal to anything other that itself! They all *look* identical to us, but they probably have slightly different molecular structures, different bendings, etc. However, for most purposes, we ignore all those differences, and we consider any dollar bill equal to another one! Just like in this case, there are many times where strict equality is too strong for mathematicians: the more useful notion of equality needs to be weakened in some way.

Let's move on to mathematical structures, then. Given two sets A and B, how should we define equality between them? For simplicity we will just think of sets as intuitive "collections" for now.

Indeed, many people define two sets to be equal whenever they contain the same things. This is called "extensionality". So, if $A = \{a, b, c\}$ and $B = \{b, c, a\}$, then we can say that $A = B$ since they both only contain a, b, c.

What about the two sets $A = \{a, b, c\}, C = \{x, y, z\}$? They are not equal, since $a \in A$ but $a \notin C$. However, there is a similarity here: they both have three elements! This is akin to saying that there is a similarity in having three cars versus having three pens: in both cases, you have three objects.

This similarity is not so trivial: if we take a new set D containing no elements in A and no elements in C, then we know the "union" (an operation on sets which takes two sets X and Y and creates a new set $X \cup Y$ containing elements from both X and Y) $A \cup D$ contains the same number of elements as $C \cup D$.

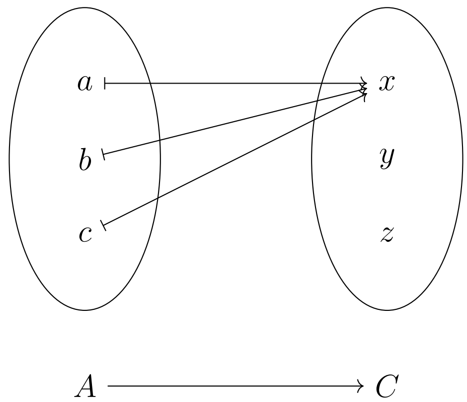
This "similarity," or having the same number of elements, is an important relation between sets. Can someone think of a rigorous way of defining this concept? How should one check whether A has the same number of elements of C? Can you also cover the infinite case?

The way mathematicians do it is by a 1-1 correspondence, or a "bijection". This notion removes the need to create a number system (cardinals), which can be done, but is much more difficult.

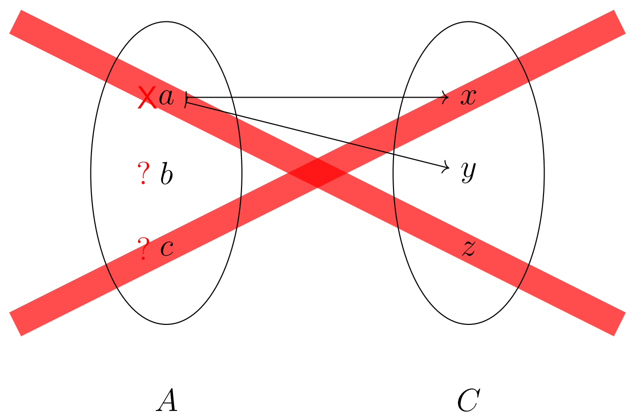
Let's get to it! There is a "bijection" between sets A and C if there is a "map", or association, of every element of A with a unique element of C, such that every element of C turns out to be paired with exactly one element of A as well. To make this idea precise, let us define a "map" of sets.

Definition 2.1 (Function)**.** Given two sets X and Y, a map, or function, $f : X \to Y$ is a pairing that assigns exactly one element of Y to every element of X. This means that we cannot have multiple elements of Y paired with the same element of X, but the other way around is allowed. We *can* have multiple elements of X paired with one element of Y. The unique element of Y assigned to an element $x \in X$ is called $f(x)$.

For example, given A and C above, we have a nice map that sends $a \mapsto x$, $b \mapsto x$, $c \mapsto x$. Clearly every element of A is now assigned exactly one element of C.



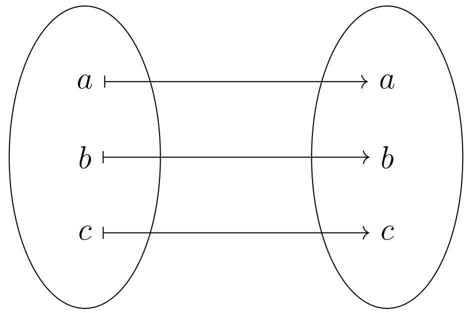
On the other hand, having $a \mapsto x$, $a \mapsto y$ is not allowed for two reasons: a is mapped to two different values, and not all elements of A are paired (you forgot about b and c)!



Here's one from geometry. Given the set of all points on the Cartesian plane, which is made up of pairs of real numbers (it is denoted by $\mathbf{R}^2$ for a reason we'll explain later), we have two important maps: one takes any point on the Cartesian plane to the x-coordinate, the other to the y-coordinate. These are both maps $\mathbf{R}^2 \to \mathbf{R}$, and they are called "coordinate maps".

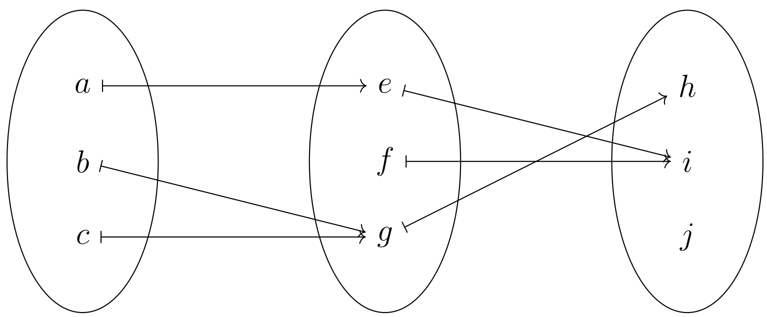
A small remark here: if you have a map $f : X \to Y$, then X is called the domain of f, and Y is called the codomain of f. In fact, there are intuitively two "maps" from the "set" of all maps between two sets to the "set" of all sets, given by taking in any map, and outputting the domain, or outputting the codomain. So, the dom function would take in f and output X, and the cod function takes in f and outputs Y.

What is the simplest map from A to A? The identity map! For any set X, there is always a map, called 1_X or the identity map, which sends $x \mapsto x$ for all $x \in X$.



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

Notice that if you have a map $f : X \to Y$ and another map $g : Y \to Z$, we can "compose them". This means, "first do f, then do g", and we get a perfectly good map, called $g \circ f : X \to Z$.



$$\begin{tikzcd} X \arrow[r, "f"] \arrow[rr, bend right, "g \circ f"'] & Y \arrow[r, "g"] & Z \end{tikzcd}$$

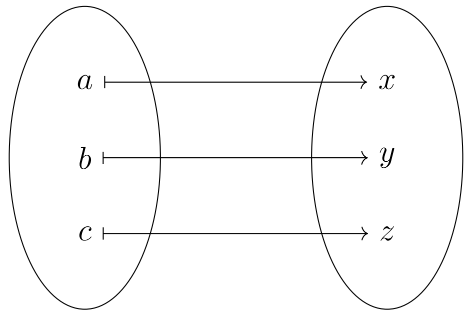
For example, say you are given a list of integers, and want to find its smallest element. For convenience, we denote the set of all lists of integers as L. Then, if you have a sorting function ($L \to L$, takes in a list and returns the same list, except sorted in order) and a function that returns the first element of a list ($L \to \mathbf{Z}$), you can create a new function, namely first sorting the list, then taking the first element. This gives you a new function $\mu : L \to \mathbf{Z}$, one that takes in lists of numbers and returns its smallest element. For the computer scientists, we're not saying this is the most efficient way to do things!

$$\begin{tikzcd} L \arrow[r, "\text{sort}"] \arrow[dr, "\mu"'] & L \arrow[d, "\text{first}"] \\ & \mathbf{Z} \end{tikzcd}$$

When we have the diagram above and $\mu = \text{first} \circ \text{sort}$, people say that *the diagram (above) commutes*. Now, we are finally ready to talk about a bijection between two sets.

Definition 2.2 (Bijection)**.** A map $f : X \to Y$ is called a "bijection" if there is another map $g : Y \to X$ such that $f \circ g = 1_Y$ and $g \circ f = 1_X$.

Finally we can arrive at our example. We've known intuitively that A and C are bijective, and here we can show why. If we take the function $f : A \to C$ that takes $a \mapsto x$, $b \mapsto y$, and $c \mapsto z$, we see that f has a natural inverse, namely the $g : C \to A$ that sends $x \mapsto a$, $y \mapsto b$, and $z \mapsto c$.



$$A \xrightarrow[f]{} C$$

This is not the only bijection. Can you count how many bijections there are between A and C? Can you extend this for any 2 sets with n elements, where $n \geq 0$?

In some cases, this "bijection" definition is a better judge of "equality" than the strict, on-the-nose equality. In fact, if you look at the structures of sets not using elements, but using only maps, you actually cannot tell the difference between two bijective sets!

In fact, it's usually "good enough" to have a notion when two objects are "close-enough" that it doesn't really matter which one we work with. As we'll see later on, definition 2.2 carried over to objects that aren't sets, like groups. In many of those cases, this relation will be called an "isomorphism" rather than a "bijection". So, one can say that a bijection is just an isomorphism of sets! (Don't do this though, you'll probably just confuse your reader). Furthermore, people tend to identify groups that are isomorphic. Most math majors are happy to embrace or at least condone this "abuse of notation".

3 Material versus Structural

A usual discussion about sets usually goes like this: we introduce a relation $\in$ where we write "$a \in A$" where a and A are both sets to mean a is an element of A. Equality is defined by some axiom of extensionality: $A = B$ when $x \in A$ if and only if $x \in B$. We then define what an ordered pair (a, b) is, where $a \in A$ and $b \in B$. Maps are then defined to be their graphs Γ, which are pairs (a, b) such that for each $a \in A$, there exists a unique $b \in B$ such that $(a, b) \in \Gamma$. A set theory that takes this membership relation as primitive is sometimes called a *material set theory*. This approach defines *everything* in terms of sets, even when it doesn't make much intuitive sense. This is opposed to a *structural set theory* which takes sets and maps between sets as primitives.

The usual approach does leave quite a bit to be desired. For example, we start with ordered pairs. Note the most important thing about an ordered pair is that there are the two coordinate maps, as mentioned above in the cartesian plane example. What this means is that given any sensible definition of an ordered pair, we should be able to extract its first coordinate and its second coordinate.

Following the material set theory approach, we must define an ordered pair as a set. We could have defined (a, b) to be $\{a, \{a, b\}\}$, $\{\{a\}, \{a, b\}\}$, or $\{\{a, c\}, \{b, d\}\}$ where $c \neq d$. Although these do satisfy what we want from ordered pairs, it is not at all obvious from these definitions that they have the necessary coordinate maps. Furthermore, a choice of any one of these constructions ends up introducing some distracting properties that aren't essential. For example, if we defined $(a, b) := \{a, \{a, b\}\}$. Then we have $a \in (a, b)$. If we defined $(a, b) := \{\{a\}, \{a, b\}\}$ then we have $\{a\} \in (a, b)$ and not $a \in (a, b)$. In a sense, our "ground up" construction, although it demonstrates the existence of ordered pairs, becomes distracting in the long run as we keep invoking the way we built or defined ordered pairs. Most people probably think of an ordered pair as consisting of an element from A and an element from B with the property that order matters rather than something like $\{a, \{a, b\}\}$.

To give another example, the usual construction of the natural numbers $0, 1, 2, \dots$ is usually to define 0 as $\varnothing$ and construct 1, 2, and so on from here. But there are at least two constructions: $\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \{\{\{\varnothing\}\}\}, \dots$ or $\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}, \dots$ (if you don't want to go crazy, this is a recursive definition. We define $0 := \varnothing$, then $n + 1 := n \cup \{n\}$. So, we get $0 = \varnothing$, $1 = \{0\}$, $2 = \{0, 1\}$, $3 = \{0, 1.2\}$, etc). Again, there are features that are simply by-products of our constructions, which have very little to do with the "essence" of a natural number. In fact, the most important thing about natural numbers is that you can use mathematical induction on them, something which is not obvious at all from either construction above. A usual 104 class will in fact proceed (usually choosing the latter method of building $\mathbf{N}$) in this manner, building $\mathbf{Z}$ from $\mathbf{N}$, then $\mathbf{Q}$ from $\mathbf{Z}$, and afterwards $\mathbf{R}$ from $\mathbf{Q}$. Just to drive the point further, people might give two different constructions of $\mathbf{R}$ via Dedekind cuts or via Cauchy completion.

The structural approach in one sense stipulates that what's more useful to us is to examine how *objects relate to other objects* instead of worrying so much about *what an object is*. How? Through morphisms! The idea is that if we are concerned with "what the constructions do", then it shouldn't bother us too much that two different constructions are isomorphic but not the same. Part of what "the constructions do" is captured by universal properties, which you'll have a sneak peek in 4.1. Later on, we'll see more of this expressed in the Yoneda lemma.

So coming back our discussion on sets, our structural set theory emphasizes studying sets by looking at how they interact with other sets, rather than "looking inside" using the membership relation.

Again, it's nice to see that we can give (material) set theoretic constuctions, but people usually don't think of real numbers as sets (and you probably shouldn't). We're not advocating that the structural approach is superior, rather that its closer to how a math major thinks on a daily basis, and is one approach to categorical styles of thinking.

One objection may be that this discussion seems to contradict the intuitive picture of a set as a bag of points, with functions taking points in a bag A to points in a bag B. One response is that the idea of sets being a bag of points becomes encoded in a special feature about the *collection of sets and their morphisms*, which we'll call **Set**. We'll use a very special set, call it T. This set is supposed to have one element in it. How we know that without looking inside T is discussed in 4.1. Note that a map $T \to S$ will take the sole element of T and take it to just one element of S. So a choice of $s : T \to S$ is a choice of $s \in S$. In fact, as a pun/abuse of notation people sometimes write $s \in S$ as shorthand for a morphism $s : T \to S$.

So what's the special feature? It should make sense that two maps $f, g : A \to B$ are equal* if their outputs on each element of A are equal. Reworded, given

$$A \underset{g}{\overset{f}{\rightrightarrows}} B$$

we say that $f = g$ if for all elements $a : T \to A$,

$$\left(T \xrightarrow{f \circ a} B \right) = \left(T \xrightarrow{g \circ a} B \right)$$

or that the diagram

$$T \xrightarrow{a} A \underset{g}{\overset{f}{\rightrightarrows}} B$$

commutes for all $a \in A$.

In more fancy terms, we say that 1 is a *separator* in **Set** or that **Set** is *well-pointed*. (More on this as we progress!)

4 Where we go from here

After this discussion, you probably have enough intuition for us to take a peek ahead!

4.1 A Preview of Universal Properties

We can characterize a construction *up to isomorphism* using universal properties. Let's look at an example of a universal property. If we know that a set T has just one element, how does it interact with other sets? We've already seen that $T \to S$ is equivalent to picking out an element from S. But what about morphisms *into* T? Since there's only one element of T, any morphism $S \to T$ can only send all elements of S to the sole element of T. The *universal property* of T is that for any other set S, *there exists a unique morphism* from S to T.

$$\begin{tikzcd} S \arrow[r, dashed, "f"] & T \end{tikzcd}$$

*You might object that we should only ask for maps up to isomorphism *of maps*, and you would absolutely have a point here. We might have a discussion on this later.

The usual definition given for the Cartesian product of two sets A and B defines $A \times B$ to be the set of all ordered pairs (a, b) where $a \in A$ and $b \in B$. This construction has a universal property, namely that the Cartesian product consists of an object $A \times B$ and two morphisms (the coordinate maps, as mentioned above!) $\pi_A : A \times B \to A$ and $\pi_B : A \times B \to B$ such that for any other object D with morphisms $j : D \to A$ and $k : D \to B$, there is a unique morphism ϕ such that

$$\begin{tikzcd}
D \arrow[drr, bend left, "k"] \arrow[dr, dashed, "\phi"] \arrow[ddr, bend right, "j"'] & & \\
& A \times B \arrow[r, "\pi_B"] \arrow[d, "\pi_A"'] & B \\
& A &
\end{tikzcd}$$

commutes.

If we know the univeral property of a construction $\mathcal{C}$, then if we can show that a different construction $\mathcal{C}'$ satisfies the same universal property then $\mathcal{C}$ and $\mathcal{C}'$ are actually isomorphic. This is where the idea of structural thinking starts to show since even we have differing constructions, if we know that the constructions *interact with other objects the same way*, then they are isomorphic.

Example 4.1. Recall that T is the terminal object in **Set**. If there is some T' that is also terminal in **Set**, then there exists a unique morphism $f : T \to T'$, and a unique morphism $g : T' \to T$ since both are terminal. This gives us the morphisms $f \circ g : T' \to T'$ and $g \circ f : T \to T$. Applying the fact that T is terminal, there can only be one morphism $T \to T$. But we already know id_T is a morphism from T to itself, hence $g \circ f = \mathrm{id}_T$. Similarly, $f \circ g = \mathrm{id}_{T'}$.

This makes sense without the formalism anyways since we'd expect the set $\{a\}$ to be have the same way as $\{b\}$ if we weren't allowed to look inside the brackets. Because of this, people usually just write 1 for the terminal object.

As for the Cartesian product, we could have come up with an alternate construction like $\{f : \{1, 2\} \to A \cup B : f(1) \in A \wedge f(2) \in B\}$. The point is that both constructions *have the same universal property and so are isomorphic*. Note that we don't consider the objects by themselves. Instead, we have the object and some morphisms alongside like $A \times B, \pi_A : A \times B \to B, \pi_B : A \times B \to B$. This fits in with the structuralism theme where we don't care so much for what $A \times B$ *is* and focus on how it relates to other objects.

Example 4.2. Fix A and B. The argument to show that $A \times B$ is unique up to isomorphism is the same: Suppose there was another such object C with morphisms $p_A : C \to A$ and $p_B : B \to C$ so that for any other object D that's equipped with morphisms $j : D \to A$ and $k : D \to B$, there is a unique morphism ϕ such that

$$\begin{tikzcd}
D \arrow[rrd, bend left, "k"] \arrow[rd, dashed, "\phi"] \arrow[rdd, bend right, "j"'] & & \\
& C \arrow[r, "p_B"] \arrow[d, "p_A"'] & B \\
& A &
\end{tikzcd}$$

Then we apply the universal properties of C and $A \times B$ to get arrows ψ and θ so that

$$\begin{tikzcd}
C \arrow[rrd, bend left, "p_B"] \arrow[rd, dashed, "\psi"] \arrow[rdd, bend right, "p_A"'] & & \\
& A \times B \arrow[r, "\pi_B"] \arrow[d, "\pi_A"'] & B \\
& A &
\end{tikzcd}
\qquad
\begin{tikzcd}
A \times B \arrow[rrd, bend left, "\pi_B"] \arrow[rd, dashed, "\theta"] \arrow[rdd, bend right, "\pi_A"'] & & \\
& C \arrow[r, "p_B"] \arrow[d, "p_A"'] & B \\
& A &
\end{tikzcd}$$

commute. This gives us arrows $\psi \circ \theta : A \times B \to A \times B$ and $\theta \circ \psi : C \to C$ which make the diagrams

$$\begin{tikzcd}
C \arrow[rrd, bend left, "p_B"] \arrow[rd, dashed, "\theta \circ \psi"] \arrow[rdd, bend right, "p_A"'] & & \\
& C \arrow[r, "p_B"] \arrow[d, "p_A"'] & B \\
& A &
\end{tikzcd}
\qquad
\begin{tikzcd}
A \times B \arrow[rrd, bend left, "\pi_B"] \arrow[rd, dashed, "\psi \circ \theta"] \arrow[rdd, bend right, "\pi_A"'] & & \\
& A \times B \arrow[r, "\pi_B"] \arrow[d, "\pi_A"'] & B \\
& A &
\end{tikzcd}$$

commute. This is less work than it might look like on first glance! But the universal properties also say that for each diagram directly above, there is a unique arrow which will make them commute. id_C and $\mathrm{id}_{A \times B}$ already make the diagrams commute, hence $\psi \circ \theta = \mathrm{id}_{A \times B}$ and $\theta \circ \psi = \mathrm{id}_C$. So C and $A \times B$ are isomorphic.

Some sources may give a formal definition of a universal property using comma categories, the Yoneda Lemma, adjunctions, or some other (related) machinery. We might get to these later, but the most important point for us is to have an intuitive idea of what universal properties are and what they're good for.

4.2 A Preview of Categorical Structures

We've seen that there's more than just sets and arrows. To start, we can actually compose them. This works by taking two arrows

$$A \xrightarrow{f} B \qquad , \qquad B \xrightarrow{g} C$$

and putting them together to give us an arrow

$$A \xrightarrow{\quad g \circ f \quad} C$$

So composition, without asking about its particular details, is a process where we take two arrows, provided the target of one arrow matches the source of the other, and put them together to obtain a new arrow!

Here's an especially important example:

Example 4.3. Let's fix a set S. If we consider the collection of arrows from S to itself, we have more than just a set $\mathrm{End}(S) := \{f : S \to S\}$. Given two $f, g \in \mathrm{End}(S)$, we can make a new $f \circ g \in \mathrm{End}(S)$. What we have is a *binary operation* $\circ$ on $\mathrm{End}(S)$. This binary operation has a neutral element, meaning there is a id_S so that $f \circ \mathrm{id}_S = \mathrm{id}_S \circ f = f$. This binary operation is also associative, meaning $f \circ (g \circ h) = (f \circ g) \circ h$. We say that $\mathrm{End}(S)$ is a *monoid*[†].

Later on, we'll see examples of objects (such as monoids, groups, rings, ...) acting on other objects (like sets, abelian groups, ...). The basic idea is that if we have, say, a monoid $\mathbf{B}M$ acting on a set S, then for each m of M, there should be a corresponding map $f_m : S \to S$.

$$\begin{tikzcd} S \arrow[loop above, "f_l"] \arrow[loop left, "f_m"] \arrow[loop right, "f_n"] \end{tikzcd}$$

This action should be subject to the condition that $f_{n \circ m} = f_n \circ f_m$ for n, m in $\mathbf{B}M$. This is so that the action respects the structure that already came with the monoid $\mathbf{B}M$. The more common notation for f_m is $m \cdot (-)$ where $(-)$ means that we're waiting to put an element of S in. So once we have some $s \in S$, we'd write $m \cdot s := f_m(s)$.

Right now we have some trivial examples like:

Example 4.4. Let $\mathbf{B}M$ be a monoid and S a set. We can let each $f_m = \mathrm{id}_S$. This is indeed an action since $f_m \circ f_n = f_{m \circ n}$ because both sides equal id_S.

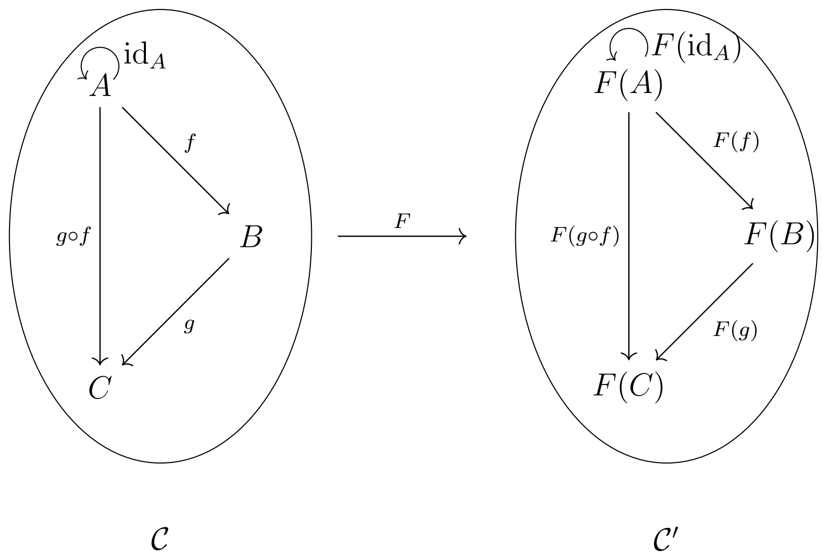
Example 4.5. If A is a set with n elements, there will be a monoid (actually a group) called S_n consisting of all permutations of the set $\{1, \ldots, n\}$, which acts on A by permuting the elements of A. For example, if $A = \{a, b\}$ then S_2 acts on A via $() \cdot a = a, () \cdot b = b$ and $(1\ 2) \cdot a = b, (1\ 2) \cdot b = a$.

We can also have a monoid act on a monoid, in particular itself.

We've had our preview of categories. So you might ask what goes between categories, as in what is a morphism of categories? A morphism of categories should first of all come with a map between the underlying objects, but it should also act on the arrows in

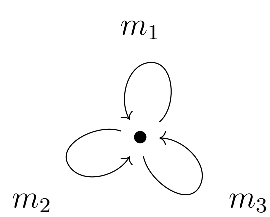
[†]This isn't quite right. It's really the delooping object $\mathbf{B}M$ of a monoid M, which falls into what people call the "periodic table of categories" which we might get to later

a category. If $f : A \to B$ is an arrow in $\mathcal{C}$, and $F : \mathcal{C} \to \mathcal{C}'$ is a morphism of categories, then there should be objects $F(A), F(B)$ in $\mathcal{C}'$ and a morphism $F(f) : F(A) \to F(B)$. Since we have identity arrows and composition in a category, F should respect those: $F(f \circ g) = F(f) \circ F(g)$ and $F(\mathrm{id}_A) = \mathrm{id}_{F(A)}$. People usually call a morphism of categories a *functor*.

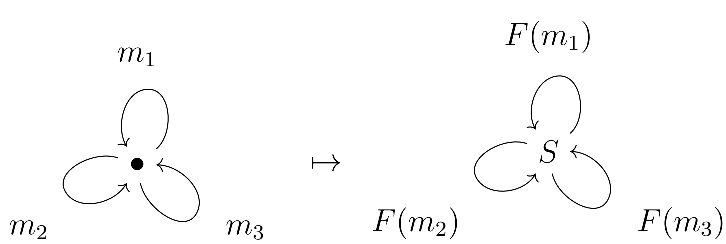


Exercise 4.6. Write down a description for a morphism of monoids. People usually call this a *monoid homomorphism*.

Example 4.7. We'll see that functors from $\mathbf{B}M$ to $\mathbf{Set}$ form the category of M sets, sets equipped with the action of a monoid. One way to think of this is to picture $\mathbf{B}M$ as a generic object, represented by $\bullet$, with arrows from $\bullet$ to itself:



Note how there's no specification for what the morphisms m_i act on, it's just some generic object $\bullet$. A functor F will send $\bullet$ to a set $S \coloneqq F(\bullet)$, and each $m_i \in M$ is sent to an endomorphism $F(m_i) : S \to S$.



What ended up happening is that we have a category which represents a theory, and we took set models of that theory by looking at functors into $\mathbf{Set}$. We could have taken models based on some other collection (like vector spaces) by taking functors into different categories: $F : \mathbf{B}M \to \mathcal{C}$. This is part of a much deeper story that tends to fall under the name *functorial semantics*.

We interpreted the domain category as some kind of theory or syntax for functorial semantics, but it could be that instead the domain category $\mathcal{C}$ encoded an object P (say, a group) that we wanted to study. To study this object P, we could try to "embed" the category $\mathcal{C}$ it inside a category that we like. For example, vector spaces. Then we study functors $\mathcal{C} \to {}_k\mathbf{Vect}$. Depending on your algebra class, you may end up studying *representations* of something like a group G which is usually defined as a group homomorphism from $G \to \mathrm{Aut}(V)$ where V is some vector space.

Actually, we can think of morphisms in a category with more than just one object as actions, with the difference being that the actions have "types", so we can only compose certain actions f and g when the source type of f matches with the target type of g i.e. $f : B \to C$ and $g : A \to B$. For example, the objects might represent states and the arrows are transformations between states. Some processes may only be possible during certain states.