Basic Category Theory
(OPLSS 2016)

Edward Morehouse
Carnegie Mellon University

June 2016
(revised June 30, 2016)

Contents

Introduction

Category theory can be thought of as a sort of generalized set theory, where the primitive concepts are those of *set* and *function*, rather than *set* and *membership*. This shift of perspective allows categories to more directly describe many structures, even those that are not particularly set-like. In category theory, the primitive concept of *set* generalizes to that of *object*, and *function* to *morphism*.

The only assumption that we make about these generalized functions is that they support a *composition structure*, whereby any configuration of compatible morphisms can be combined to yield a new morphism, and the details of how we go about combining the parts into a whole doesn't matter, only the configuration of those parts does.

This is reminiscent of many aspects of our physical world. When we build a castle out of Lego bricks, the order in which we assembled the bricks is not recorded anywhere in the finished product, only their configuration with respect to one another remains.

By beginning from very few assumptions, category theory permits a great deal of *axiomatic freedom*. Additional postulates (e.g. the axiom of choice) can then be selectively reintroduced in order to characterize a particular object theory of interest (e.g. set theory).

Because categorical characterizations are based on the concepts of object and morphism, they must describe their subjects *behaviorally* (or externally), rather than *structurally* (or internally): in category theory we can't pin down what the objects of our study actually *are*, only how they relate to one another via morphisms. In this sense, category theory is the sociology of formal systems.

For example, we will see how we can characterize the cartesian product once and for all using a *universal property*. This allows us to describe cartesian products of sets, of groups, of topological spaces, of types, of propositions, and of countless other things, all in one fell swoop, rather than on a tedious case-by-case basis.

Chapter 1

Basic Categories

1.1 Definition of a Category

Definition 1.1.0.1 (category) A **category** $\mathbb{C}$ consists of the following *data*:

- A collection of **objects**, $\mathbb{C}_0$ (comprising the 0-dimensional part of $\mathbb{C}$). We write "$A : \mathbb{C}$" to indicate that $A \in \mathbb{C}_0$.
- A collection of **morphisms** or "arrows", $\mathbb{C}_1$ (the 1-dimensional part). We write "$f :: \mathbb{C}$" to indicate that $f \in \mathbb{C}_1$.
- Two **morphism boundary** maps from arrows to objects: **domain**, "∂^-", and **codomain**, "∂^+". For $\mathbb{C}$-objects A and B, we indicate the collection of $\mathbb{C}$-arrows with domain A and codomain B by "$\mathbb{C}\,(A \to B)$", and call this collection "hom"[1]. When the category in question is obvious or irrelevant, we just write "$A \to B$". We indicate that an arrow f is a member of this collection by writing "$f : \mathbb{C}\,(A \to B)$" or "$f : A \to B$".
- An **identity morphism** map from objects to arrows, "id", such that both boundaries of an object's identity arrow are just that object itself:

$$\mathrm{id}(A) : A \to A$$

- A partial binary function on arrows, **morphism composition**, "$- \cdot -$", that is defined just in case the codomain of the first is equal to the domain of the second, in which case the composite arrow has the domain of the first and codomain of the second:

$$\text{if} \quad f : A \to B \quad \text{and} \quad g : B \to C \quad \text{then} \quad f \cdot g : A \to C$$

[1] presumably, short for "homomorphisms"

This data is required to respect the following *relations*.

- **composition left unit law**: for an arrow $f : \mathrm{A} \to \mathrm{B}$,

$$\mathrm{id}(\mathrm{A}) \cdot f \quad = \quad f$$

- **composition right unit law**: for an arrow $f : \mathrm{A} \to \mathrm{B}$,

$$f \cdot \mathrm{id}(\mathrm{B}) \quad = \quad f$$

- **composition associative law**: for arrows $f : \mathrm{A} \to \mathrm{B}$, $g : \mathrm{B} \to \mathrm{C}$ and $h : \mathrm{C} \to \mathrm{D}$,

$$(f \cdot g) \cdot h \quad = \quad f \cdot (g \cdot h)$$

By the associative law we may unambiguously write compositions without using brackets.

Definition 1.1.0.2 In order to avoid gratuitously naming the boundaries of arrows, we will call a pair of arrows $f\,, g :: \mathbb{C}$:

- **coinitial**, or a "span", if $\partial^-(f) = \partial^-(g)$,
- **coterminal**, or a "cospan", if $\partial^+(f) = \partial^+(g)$,
- **composable** if $\partial^+(f) = \partial^-(g)$,
- **parallel** if both coinitial and coterminal, and
- **anti-parallel** if composable in both orders.

Additionally, we will call an arrow an **endomorphism** if it is composable with itself, and a list of arrows a **path** if they are serially composable, that is, if $\partial^+(f_i) = \partial^-(f_{i+1})$ for the list $[f_0\,, \cdots, f_n]$.

Remark 1.1.0.3 (applicative order composition) It is common to see the composition $f \cdot g$ written as "$g \circ f$". This can be useful when we want to apply a composite morphism to an argument in a category where a morphism is some sort of function. Then $(g \circ f)(x) = g(f(x))$, which coincides with our custom to write function application with the argument on the right. It may help to read "$f \cdot g$" as "f then g", and to read "$g \circ f$" as "g after f".

Remark 1.1.0.4 (dimensional promotion) It is often convenient to call the identity arrow on an object by the same name as the object itself, e.g. to write "A" in place of id(A). This is called **dimensional promotion**, and will become useful as we introduce more complex arrow constructions and concision becomes more of an issue.

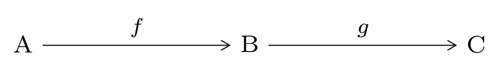
Remark 1.1.0.5 (unbiased presentation) There is an equivalent presentation of categories in terms of **unbiased composition**. There, instead of a single **binary composition** operation acting on a composable *pair* of arrows, we have a length-indexed composition operation for *paths* of arrows (still with unit and associative laws). In this presentation, an identity morphism is a **nullary composition**, a morphism itself is a **unary composition**, and in general, any length n path of arrows has a unique composite. Although more cumbersome to axiomatize, an unbiased presentation of categories makes it easier to appreciate the idea at the heart of the definition: every composable configuration of things should have a unique composite.

Exercise 1.1.0.6 (uniqueness of composition units) By definition, identity arrows act as (two-sided) units for composition. Prove that they are the only arrows with this property.

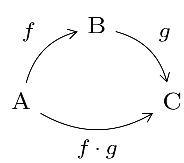
Hint (proof by *Fight Club*): suppose there were another arrow, $\mathrm{id}' : \mathrm{A} \to \mathrm{A}$, that acted as a unit for composition at A, then what would we know about the composite $\mathrm{id}(\mathrm{A}) \cdot \mathrm{id}'(\mathrm{A})$?

1.2 Diagrams

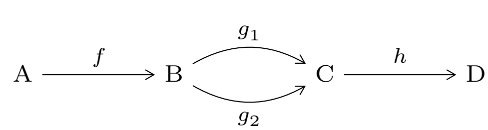
We think of an arrow as emanating *from* its domain and proceeding *to* its codomain. We can represent configurations of arrows of a category using a directed graph whose vertices are labeled by objects of the category and whose edges are labeled by arrows. Such a graph is known as a **diagram**, for example:



We can represent equations between arrows using diagrams as well. We say that a diagram is **commuting** or "commutes" if the composites of *parallel* paths depicted in the diagram are equal. For example, the fact that each pair of *composable* arrows has a unique composite gives us commuting composition triangles, such as:

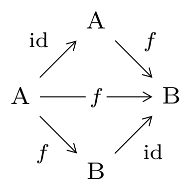


Commuting diagrams may be extended by pre- or post-composition of arrows. This is called **whiskering**, and depicts the fact that equality of morphisms is a *congruence* with respect to composition: if $g_1 = g_2$ then $f \cdot g_1 = f \cdot g_2$ and $g_1 \cdot h = g_2 \cdot h$ whenever the composites are defined. The name comes from the fact that the arrows pre- or post-composed to the diagram look like whiskers:

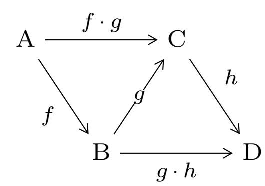


Pairs of commuting diagrams may also be combined along a common path. This is called **pasting**, and depicts the transitivity of equality: if $f_1 = f_2$ and $f_2 = f_3$ then $f_1 = f_3$.

We may express the *unit* and *associative* laws for composition succinctly using commuting diagrams:



and



In the diagram for unitality (left), the triangles representing the left and right unit laws are pasted together along the singleton path $[f]$. In the diagram for associativity (right), each of the two composition triangles is whiskered by an arrow (h and f, respectively), and the resulting diagrams are pasted together along the shared *path* $[f\ ,\ g\ ,\ h]$.

In the graphical language of diagrams, any vertex labeled by an object may be duplicated and the two copies joined by an edge labeled by the respective identity morphism. Conversely, any edge labeled by an identity morphism may be collapsed, identifying the two vertices at its boundary, which are necessarily labeled by the same object.

Except for the sake of emphasis, we generally omit composite arrows (including identitites, which are nullary composites) when drawing diagrams, because their existence may always be inferred. Notice that the associative law for composition is built into the graphical language of diagrams by the fact that there is no graphical representation for the bracketing of the arrows in a path.

In order to avoid gratuitously naming objects in diagrams, we will represent an anonymous object as a dot ("$\bullet$"). Two such dots occurring in a diagram need not represent the same object.

1.3 Structured Sets as Categories

1.3.1 Discrete Categories

The most trivial possible category has nothing in it. It is called the **empty category**, and written "$\mathbb{0}$". Despite having completely uninteresting *structure*, we will see that this category nevertheless has a very interesting *property*.

Only slightly less trivially, we can consider a category with just a single object, call it "$\star$", and no arrows other than the required identity. This describes a

singleton category, typically written "$\mathbb{1}$". This category will turn out to have a very interesting property as well.

Generalizing a bit, we can regard any *set* as a category. As a category, a set has its members as objects and no arrows other than the required identities. Categories in which all arrows are identities are called **discrete**.

1.3.2 Preorder Categories

A **preorder** is a reflexive and transitive binary relation on a set, typically written "$- \leq -$". We can interpret a preordered set $(\mathrm{P}, \leq)$ as a **preorder category** $\mathbb{P}$ in the following way.

objects: $\mathbb{P}_0 \;\coloneqq\; \mathrm{P}$

arrows: $\mathbb{P}\,(x \to y) \;\coloneqq\; \begin{cases} \{x \leq y\} & \text{if } x \leq y \\ \emptyset & \text{otherwise} \end{cases}$

identities: $\mathrm{id}(x) \;\coloneqq\; x \leq x$

composition: $x \leq y \;\cdot\; y \leq z \;\coloneqq\; x \leq z$

In other words, a preordered set is a category in which each hom collection is either empty, or else a singleton; and a hom is inhabited just in case its domain is less than or equal to its codomain according to the order relation.

Remark 1.3.2.1 A preorder need not have anything to do with our usual notion of order on a set. For example, the integers with the "divides" relation, $- \mid -$, is a perfectly good preordered set, in which $-2 \leq 2$, but also $2 \leq -2$ (and yet $-2 \neq 2$ – a preorder need not be *antisymmetric*).

In a preorder category the unit and associative laws of composition are trivially satisfied by the fact that all elements of a singleton or empty set are equal. In fact, every diagram in a preorder category must commute! Preorder categories are sometimes called "thin".

The simplest preorder category that is not discrete has two distinct objects and a single non-identity arrow from one to the other. It looks like this:

$$\bullet \longrightarrow \bullet$$

This category is called the **interval category**, and written "$\mathbb{I}$". It plays an important role in the study of higher-dimensional categorical structures.

1.3.3 Monoid Categories

A **monoid** is a set M together with an associative binary operation "$- * -$" with neutral element "ε". We can interpret a monoid $(\mathrm{M}, *, \varepsilon)$ as a **monoid category**

$\mathbb{M}$ in the following way.

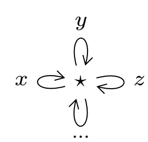
objects: $\mathbb{M}_0 \;\coloneqq\; \{\star\}$

arrows: $\mathbb{M}(\star \to \star) \;\coloneqq\; \mathrm{M}$

identities: $\mathrm{id}(\star) \;\coloneqq\; \varepsilon$

composition: $x \cdot y \;\coloneqq\; x * y$

Thus, a monoid becomes a category by "suspending" its elements into the hom of *endomorphisms* of an anonymous object, which I imagine looks something like this:



The unit and associative laws of composition are satisfied by the corresponding laws for the monoid operation.

If we wanted to make the simplest possible monoid category that is not discrete, we would have to think about what it means to be simple. We can begin by postulating a single non-identity arrow, $s : \star \to \star$. But because s is an endomorphism, we must say what $s \cdot s$, $s \cdot s \cdot s$, and in general, $s^{(n)}$ are. One possibility is to introduce no relations. This gives us the free monoid on one generator, better known as $(\mathbb{N}, +, 0)$.

1.4 Categories of Structured Sets

In addition to (structured) sets *as* categories, we also have categories *of* (structured) sets.

1.4.1 The Category of Sets

There is a **category of sets**, called "Set", whose objects are sets and whose arrows are functions between them. Not surprisingly, we take function composition for the composition of arrows and identity functions for the identity arrows. That is, given composable functions f and g,

$$f \cdot g \quad \coloneqq \quad \lambda\, x \,.\, g(f(x)) \qquad \text{and} \qquad \mathrm{id} \quad \coloneqq \quad \lambda\, x \,.\, x$$

Composition of Set-morphisms is associative and unital precisely because composition of functions is (check this!).

1.4.2 The Category of Preorders

There is a **category of preorders**, called "PreOrd", that has *preordered sets* as objects and monotone (i.e. order-preserving) functions as arrows. Arrow composition is again function composition and the identity arrows are the identity functions.

In order to conclude that this is a category we (i.e. you) must check that the composition of monotone functions is again monotone, and that identity functions are monotone. You just checked that function composition is associative and has identity functions as units, so since monotone functions are functions, you need not check associativity and unitality again for the special case.

1.4.3 The Category of Monoids

The **category of monoids**, Mon, has *monoids* as objects and monoid homomorphisms as arrows. A monoid homomorphism is a function between the underlying sets of the monoids that respects the operations and units:

$$\begin{array}{l} f : \textsc{Mon}\,((\mathrm{M}\,,*\,,\varepsilon) \to (\mathrm{N}\,,*'\,,\varepsilon')) \;:=\; f : \textsc{Set}\,(\mathrm{M} \to \mathrm{N}) \\ \text{such that} \quad f(x * y) \;=\; f(x) *' f(y) \quad \text{and} \quad f(\varepsilon) \;=\; \varepsilon' \end{array}$$

Abstract algebra provides a rich source of categories. These categories generally have sets with some form of algebraic structure as objects and structure-preserving functions as arrows. In addition to that of monoids, we have the category of groups (Grp), of rings (Rng), of modules over a ring, and so on.

1.5 Categories of Types and Terms

Although we are not yet in a position to give the details, we can begin to see how to use categories to interpret type theories. The objects of such categories will be interpretations of types – and more generally, of typing contexts. The arrows will be interpretations of terms-in-context, which we will usually abbreviate to "terms". We will interpret a term-in-context as a morphism from the interpretation of its context to that of its type:

$$[\![\Gamma \;\vdash\; \mathrm{M} : \mathrm{A}]\!] : [\![\Gamma]\!] \to [\![\mathrm{A}]\!]$$

when confusion is unlikely to result, we will abbreviate this to "$[\![\mathrm{M}]\!] : [\![\Gamma]\!] \to [\![\mathrm{A}]\!]$", since the context and type are recoverable from the arrow boundary.

We must postpone interpreting type and context formation until we have built up some more categorical machinery. So we temporarily restrict our attention to theories with only atomic types and where all contexts are singletons. We refer to this informally as "baby type theory".

There, we expect the following **variable rule** to be admissible:

$$\frac{}{x : \mathrm{A} \ \vdash \ x : \mathrm{A}} \ var$$

and we want to interpret a variable-in-singleton-context as an identity morphism:

$$[\![x : \mathrm{A} \ \vdash \ x : \mathrm{A}]\!] \ = \ \mathrm{id}([\![\mathrm{A}]\!]) : [\![\mathrm{A}]\!] \to [\![\mathrm{A}]\!]$$

Likewise, we expect the following **substitution rule** to be admissible:

$$\frac{x : \mathrm{A} \ \vdash \ \mathrm{M} : \mathrm{B} \qquad y : \mathrm{B} \ \vdash \ \mathrm{N} : \mathrm{C}}{x : \mathrm{A} \ \vdash \ \mathrm{N}[y {\mapsto} \mathrm{M}] : \mathrm{C}} \ sub$$

and we want to interpret the substitution of a term for a variable in a term as the composition of the respective terms:

$$[\![\mathrm{N}[y {\mapsto} \mathrm{M}]]\!] \ = \ [\![\mathrm{M}]\!] \cdot [\![\mathrm{N}]\!] : [\![\mathrm{A}]\!] \to [\![\mathrm{C}]\!]$$

In order to know that this is sound, we must check that the interpretation respects term equality.

unit laws There are two substitutions we can perform where one of the terms is a variable, namely, $\mathrm{M}[x {\mapsto} x]$ and $y[y {\mapsto} \mathrm{M}]$. By the definition of substitution, both terms are equal to M itself. Our interpretation is compatible with this fact by the respective composition unit laws:

$$[\![x : \mathrm{A} \ \vdash \ \underbrace{\mathrm{M}[x {\mapsto} x]}_{\mathrm{M}} : \mathrm{B}]\!] \ = \ \underbrace{[\![x : \mathrm{A} \ \vdash \ x : \mathrm{A}]\!]}_{\mathrm{id}([\![\mathrm{A}]\!])} \cdot [\![x : \mathrm{A} \vdash \mathrm{M} : \mathrm{B}]\!]$$

$$[\![x : \mathrm{A} \ \vdash \ \underbrace{y[y {\mapsto} \mathrm{M}]}_{\mathrm{M}} : \mathrm{B}]\!] \ = \ [\![x : \mathrm{A} \vdash \mathrm{M} : \mathrm{B}]\!] \cdot \underbrace{[\![y : \mathrm{B} \ \vdash \ y : \mathrm{B}]\!]}_{\mathrm{id}([\![\mathrm{B}]\!])}$$

associative law Likewise, there are two ways of using substitution to reduce the three terms

$$x : \mathrm{A} \ \vdash \ \mathrm{M} : \mathrm{B} \quad , \quad y : \mathrm{B} \ \vdash \ \mathrm{N} : \mathrm{C} \quad , \quad z : \mathrm{C} \ \vdash \ \mathrm{P} : \mathrm{D}$$

to a single term, namely, $\mathrm{P}[z {\mapsto} \mathrm{N}[y {\mapsto} \mathrm{M}]]$ and $\mathrm{P}[z {\mapsto} \mathrm{N}][y {\mapsto} \mathrm{M}]$. By the definition of substitution, these are the same term in baby type theory (why?). Our interpretation is compatible with this fact since:

$$\begin{array}{ll} & [\![\mathrm{P}[z {\mapsto} \mathrm{N}[y {\mapsto} \mathrm{M}]]]\!] \\ = & [\![\mathrm{N}[y {\mapsto} \mathrm{M}]]\!] \cdot [\![\mathrm{P}]\!] \\ = & ([\![\mathrm{M}]\!] \cdot [\![\mathrm{N}]\!]) \cdot [\![\mathrm{P}]\!] \\ = & [\![\mathrm{M}]\!] \cdot ([\![\mathrm{N}]\!] \cdot [\![\mathrm{P}]\!]) \\ = & [\![\mathrm{M}]\!] \cdot [\![\mathrm{P}[z {\mapsto} \mathrm{N}]]\!] \\ = & [\![\mathrm{P}[z {\mapsto} \mathrm{N}][y {\mapsto} \mathrm{M}]]\!] \end{array}$$

Indeed, this categorical semantics is sound for baby type theory.

1.6 Categories of Categories

We have interpreted some (hopefully) familiar mathematical structures (sets, preordered sets, monoids) *as* categories, but we have also described categories *of* these structures (Set, PreOrd, Mon). So these are in fact *categories of categories*! In each case, the objects comprise a sort of structured collection, and the arrows a mapping between these that respects the relevant structure.

Since categories themselves comprise a sort of structured collection, we may wonder whether we can identify a reasonable notion of arrow between categories, and thus define general categories of categories. Indeed we can, so long as we heed a broad foundational restriction and avoid allowing a category of categories to be an element of itself. Otherwise, we leave ourselves open to paradoxes.

1.6.1 Functors

Recall that a category has collections of objects and arrows, together with an (associative and unital) composition structure. It is precisely this composition structure that we want an *arrow between categories* to preserve.

Definition 1.6.1.1 (functor) Given two *categories* $\mathbb{C}$ and $\mathbb{D}$, a **functor** F with domain $\mathbb{C}$ and codomain $\mathbb{D}$ consists of:

- an object map, $F_0 : \mathbb{C}_0 \to \mathbb{D}_0$,
- an arrow map, $F_1 : \mathbb{C}_1 \to \mathbb{D}_1$,
  which respects the boundaries of arrows:

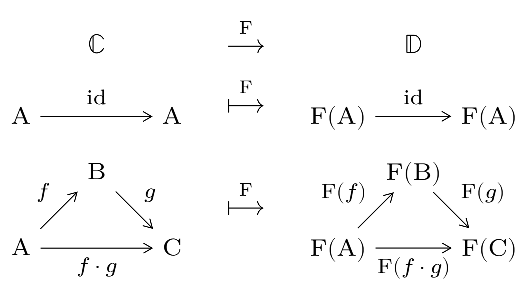
  $$f : \mathbb{C}\,(A \to B) \quad \longmapsto \quad F_1(f) : \mathbb{D}\,(F_0(A) \to F_0(B))$$

  and which furthermore respects the composition structure:

  nullary composition: $F_1(\mathrm{id}(A)) \;=\; \mathrm{id}(F_0(A))$

  binary composition: $F_1(f \cdot g) \;=\; F_1(f) \cdot F_1(g)$

It is customary to drop the dimension subscripts on the constituent maps of a functor. We can represent the composition structure-preserving aspect of a functor diagrammatically as follows:



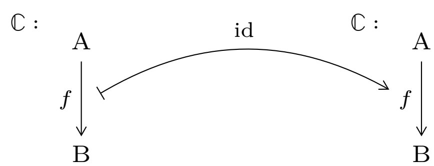
Equivalently, we could take the *unbiased* point of view and say that a functor respects the composition of arbitrary *paths* of arrows. As a consequence, functors must respect the commuting of diagrams: a functor image of a commuting diagram in its domain category is a commuting diagram in its codomain category.

Functors provide a notion of morphism of categories, so we can ask about *their* composition structure as well.

Definition 1.6.1.2 (identity functor) Given any category $\mathbb{C}$ we define the **identity functor** on $\mathbb{C}$, $\mathrm{id}(\mathbb{C}) : \mathbb{C} \to \mathbb{C}$, comprising identity maps on both objects and arrows:

$$(\mathrm{id}(\mathbb{C}))_0 \;:=\; \mathrm{id}(\mathbb{C}_0) \qquad \text{and} \qquad (\mathrm{id}(\mathbb{C}))_1 \;:=\; \mathrm{id}(\mathbb{C}_1)$$

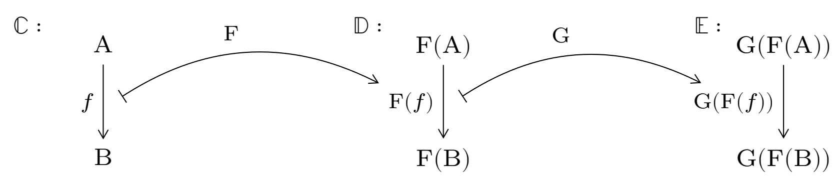
An identity functor takes an arrow, including its boundary, to itself:



Definition 1.6.1.3 (functor composition) Given functors F from $\mathbb{C}$ to $\mathbb{D}$ and G from $\mathbb{D}$ to $\mathbb{E}$, we define the **composition** $\mathrm{F} \cdot \mathrm{G}$ from $\mathbb{C}$ to $\mathbb{E}$, using the respective compositions on its object and arrow maps:

$$(\mathrm{F} \cdot \mathrm{G})_0 \;:=\; \mathrm{F}_0 \cdot \mathrm{G}_0 \qquad \text{and} \qquad (\mathrm{F} \cdot \mathrm{G})_1 \;:=\; \mathrm{F}_1 \cdot \mathrm{G}_1$$

That is:



Lemma 1.6.1.4 (categories of categories) Given a collection of categories and functors between them, we can form the category having:

- the categories as objects
- paths in the functors as arrows
- identity functors as identity arrows
- functor composition as arrow composition

It is easy to check that the associative and unit laws of composition are satisfied.

Functors are morphisms in categories whose objects are themselves categories. They are structure-preserving maps, where the structure in question is the composition structure of a category.

Exercise 1.6.1.5 What is a functor:

- between *discrete categories*?
- between *preorder categories*?
- between *monoid categories*?

Example 1.6.1.6 (forgetful functors) For a category of structured sets (e.g. monoids, groups, rings or topological spaces) there is a **forgetful functor** to the category of sets, which disregards the structure and retains just the underlying set.

For instance, there is a forgetful functor $\mathrm{U} : \textsc{Mon} \rightarrow \textsc{Set}$ that maps the monoid $(\mathbb{N}\,,+\,,0)$ to the set $\mathbb{N}$, and maps the monoid inclusion $(\mathbb{N}\,,+\,,0) \hookrightarrow (\mathbb{Z}\,,+\,,0)$ to the set inclusion $\mathbb{N} \hookrightarrow \mathbb{Z}$.

For any $\mathrm{A}\,,\mathrm{B} : \mathbb{C}$, we can consider the restriction of a functor $\mathrm{F} : \mathbb{C} \rightarrow \mathbb{D}$ to the hom $\mathbb{C}\,(\mathrm{A} \rightarrow \mathrm{B})$:

$$\mathrm{F}_1 \,|_{\mathbb{C}\,(\mathrm{A}\rightarrow\mathrm{B})} \;:\; \mathbb{C}\,(\mathrm{A} \rightarrow \mathrm{B}) \rightarrow \mathbb{D}\,(\mathrm{F}_0(\mathrm{A}) \rightarrow \mathrm{F}_0(\mathrm{B}))$$

The functor is called **full** if all such restrictions are surjective maps, that is, if

$$\forall\ \mathrm{A}\,,\mathrm{B} : \mathbb{C}\ .\ \mathrm{F}_1 \,|_{\mathbb{C}\,(\mathrm{A}\rightarrow\mathrm{B})} \quad \text{is surjective}$$

It is **faithful** if all such restrictions are injective maps, that is, if

$$\forall\ \mathrm{A}\,,\mathrm{B} : \mathbb{C}\ .\ \mathrm{F}_1 \,|_{\mathbb{C}\,(\mathrm{A}\rightarrow\mathrm{B})} \quad \text{is injective}$$

Exercise 1.6.1.7

- For a functor $\mathrm{F} : \mathbb{C} \rightarrow \mathbb{D}$, how is F being full different from $\mathrm{F}_1 : \mathbb{C}_1 \rightarrow \mathbb{D}_1$ being surjective?
- How is F being faithful different from $\mathrm{F}_1 : \mathbb{C}_1 \rightarrow \mathbb{D}_1$ being injective?
- Is the forgetful functor from example 1.6.1.6 full? Is it faithful?

1.6.2 The Special Role of Sets

A collection is called "small" if it is a set. A category is called **small** if its collection of arrows – and hence, of objects – is small. There is a **category of small categories** and functors between them, called "Cat". Observe that it is *not* the case that $\textsc{Cat} : \textsc{Cat}$, because $\textsc{Cat}_0$ contains all the small *discrete categories*, i.e. the sets, and this collection is already too large to be a set.

Often we don't care whether a category is (globally) small, but only that each of its hom collections is. A category is called **locally small** if for any pair of its objects, $\mathrm{A}, \mathrm{B} : \mathbb{C}$, the collection of arrows, $\mathbb{C}(\mathrm{A} \to \mathrm{B})$ is small. Many categories of interest are locally small. In particular, Set and Cat are locally small (if you know some basic set theory, try to work out why).

Unless otherwise specified, the categories that we encounter in this course will be locally small. Thus, we will stop being coy about what sort of "collection" a hom is, and refer instead to **hom sets**.

The fact that the collections of parallel arrows in a locally small category are sets puts the category Set in a privileged position. For example, if we fix an object $\mathrm{X} : \mathbb{C}$, then we can define a function that, given any object $\mathrm{A} : \mathbb{C}$, returns the set of arrows $\mathbb{C}(\mathrm{X} \to \mathrm{A})$. This function extends to a functor:

Lemma 1.6.2.1 (representable functors) For each object of a locally small category, $\mathrm{X} : \mathbb{C}$, there is a functor, known as a **representable functor**,

$$\begin{array}{rcl} & \underline{\mathbb{C}(\mathrm{X} \to -)} & \\ \mathbb{C} & \longrightarrow & \textsc{Set} \\ \mathrm{A} & \longmapsto & \mathbb{C}(\mathrm{X} \to \mathrm{A}) \\ f : \mathrm{A} \to \mathrm{B} & \longmapsto & \mathbb{C}(\mathrm{X} \to f) := - \cdot f : \mathbb{C}(\mathrm{X} \to \mathrm{A}) \to \mathbb{C}(\mathrm{X} \to \mathrm{B}) \end{array}$$

Unpacking this, it says that "$\mathbb{C}(\mathrm{X} \to -)$" is the name of a functor from $\mathbb{C}$ to Set, that maps an object $\mathrm{A} : \mathbb{C}$ to the *set* of arrows, $\mathbb{C}(\mathrm{X} \to \mathrm{A})$, and maps an arrow $f : \mathbb{C}(\mathrm{A} \to \mathrm{B})$ to the *function* that post-composes f to any arrow in $\mathbb{C}(\mathrm{X} \to \mathrm{A})$, yielding an arrow in $\mathbb{C}(\mathrm{X} \to \mathrm{B})$. The notation "$- \cdot f$" is just syntactic sugar for $\lambda\,(a : \mathrm{X} \to \mathrm{A})\;.\;a \cdot f$. The object $\mathrm{X} : \mathbb{C}$ is known as the "representing object" of this functor.

Proof. In order to show that $\mathbb{C}(\mathrm{X} \to -)$ is indeed a functor we must confirm that it preserves the composition structure:

nullary composition The idea is that post-composing an identity arrow does nothing, that is, it applies the identity function to the hom set. For $\mathrm{A} : \mathbb{C}$:

$$\begin{array}{cl} & \mathbb{C}(\mathrm{X} \to \mathrm{id}(\mathrm{A})) \\ = & [\text{definition of representable functor}] \\ & \lambda\, a\;.\;a \cdot \mathrm{id}(\mathrm{A}) \\ = & [\text{composition unit law}] \\ & \lambda\, a\;.\;a \\ = & [\text{definition of identity function}] \\ & \mathrm{id}(\mathbb{C}(\mathrm{X} \to \mathrm{A})) \end{array}$$

binary composition Here, the idea is that post-composing a composite of arrows post-composes the first, and then post-composes the second to the result, that is, it composes the post-compositions. For $f : \mathbb{C}\,(\mathrm{A} \to \mathrm{B})$ and $g : \mathbb{C}\,(\mathrm{B} \to \mathrm{C})$:

$$
\begin{array}{cl}
 & \mathbb{C}\,(\mathrm{X} \to f \cdot g) \\
= & \text{[definition of representable functor]} \\
 & \lambda\, a \;.\; a \cdot f \cdot g \\
= & [\beta\text{-expansion}] \\
 & \lambda\, a \;.\; (\lambda\, b \;.\; b \cdot g)(a \cdot f) \\
= & [\beta\text{-expansion}] \\
 & \lambda\, a \;.\; (\lambda\, b \;.\; b \cdot g)((\lambda\, a \;.\; a \cdot f)(a)) \\
= & \text{[definition of function composition]} \\
 & (\lambda\, a \;.\; a \cdot f) \cdot (\lambda\, b \;.\; b \cdot g) \\
= & \text{[definition of representable functor]} \\
 & \mathbb{C}\,(\mathrm{X} \to f) \cdot \mathbb{C}\,(\mathrm{X} \to g)
\end{array}
$$

□

Because of the special role of the category of sets, the study of representable functors provides one of several, ultimately equivalent, ways of understanding categories. Due to our choice of emphasis and time constraints, it is not the one we will pursue here, but it is worth being aware of.

1.7 New Categories from Old

Now that we have met a few categories, let's look at some ways to create new categories out of them.

1.7.1 Ordered Pair Categories

Recall that given any two sets, we can for their **set of ordered pairs**:

$$\mathrm{A} \times \mathrm{B} \quad \coloneqq \quad \{(a\,,b) \mid a \in \mathrm{A} \text{ and } b \in \mathrm{B}\}$$

Likewise, given any two categories, we can construct a new category whose constituent parts are just ordered pairs of the respective parts of the given categories.

Definition 1.7.1.1 (ordered pair category) For categories $\mathbb{C}$ and $\mathbb{D}$, the **ordered pair category** $\mathbb{C} \times \mathbb{D}$ has the following structure:

objects: $(\mathrm{A}\,,\mathrm{X}) \quad \text{for} \quad \mathrm{A} : \mathbb{C} \quad \text{and} \quad \mathrm{X} : \mathbb{D}$

arrows: $(f\,,p) \quad \text{for} \quad f :: \mathbb{C} \quad \text{and} \quad p :: \mathbb{D}$,
with $\partial^i((f\,,p)) \;\coloneqq\; (\partial^i(f)\,,\partial^i(p))$

identities: $\mathrm{id}((\mathrm{A}\,,\mathrm{X})) \;:=\; (\mathrm{id}(\mathrm{A})\,,\mathrm{id}(\mathrm{X}))$

composition: $(f\,,p)\cdot(g\,,q) \;:=\; (f\cdot g\,,p\cdot q)$

Soon we will be in a position to prove that ordered pair categories have the universal property of a categorical *product*, and we can use that property to define functors into them. But we will want to define functors out of them as well. Such a functor, whose domain is an ordered pair of categories, is called a **bifunctor**. For defining bifunctors, the following lemma is very helpful:

Lemma 1.7.1.2 (bifunctor lemma) Given categories $\mathbb{C}\,,\mathbb{D}\,,\mathbb{E}$, an object map, $\mathrm{F}_0 : \mathbb{C}_0 \times \mathbb{D}_0 \to \mathbb{E}_0$ and a boundary-respecting arrow map, $\mathrm{F}_1 : \mathbb{C}_1 \times \mathbb{D}_1 \to \mathbb{E}_1$, the pair $(\mathrm{F}_0\,,\mathrm{F}_1)$ constitutes a functor $\mathrm{F} : \mathbb{C} \times \mathbb{D} \to \mathbb{E}$ just in case:

(i) F is a functor in each argument separately:

$$\begin{array}{ll} \forall\ \mathrm{A} : \mathbb{C}\ .\ \mathrm{F}(\mathrm{A}\,,-) : \mathbb{D} \to \mathbb{E} & \text{is a functor} \\ \forall\ \mathrm{X} : \mathbb{D}\ .\ \mathrm{F}(-\,,\mathrm{X}) : \mathbb{C} \to \mathbb{E} & \text{is a functor} \end{array}$$

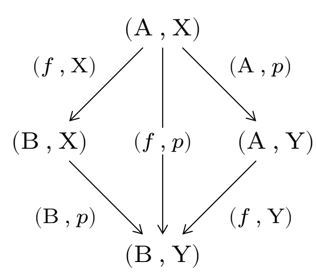
(ii) and for arrows $f : \mathbb{C}\,(\mathrm{A} \to \mathrm{B})\,,\, p : \mathbb{D}\,(\mathrm{X} \to \mathrm{Y})$ we have the *interchange* property:

$$\mathrm{F}(f\,,\mathrm{X})\cdot\mathrm{F}(\mathrm{B}\,,p) = \mathrm{F}(\mathrm{A}\,,p)\cdot\mathrm{F}(f\,,\mathrm{Y})$$

(Note the use of *dimensional promotion.*)

Proof. First we must show that if F is a bifunctor then the two conditions hold.

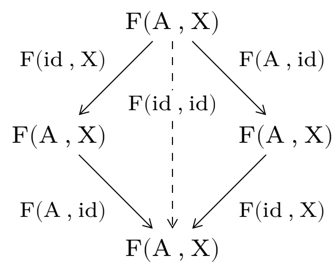
Condition (i) follows immediately by fixing the respective arrows to be identities. Next, observe that in the category $\mathbb{C} \times \mathbb{D}$, the following diagram commutes:



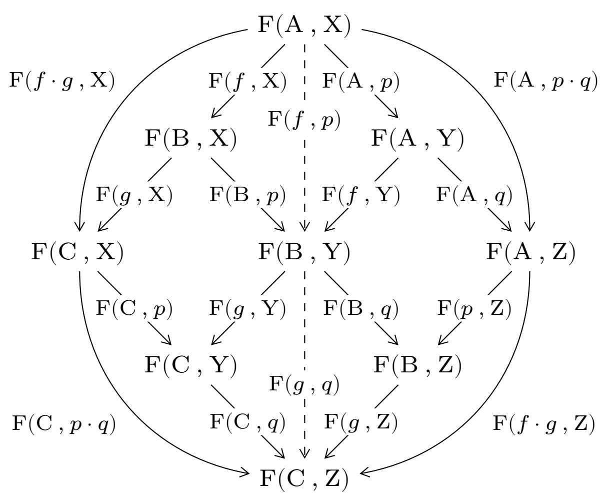
Condition (ii) then follows because functors preserve commuting diagrams.

Going the other way, using condition (ii) we may *define* $\mathrm{F}(f\,,p)$ to be this common arrow. Now we must show that F defined this way respects composition structure.

nullary composition In the diagram below, condition (i) ensures that each solid arrow is a functor image of an identity arrow, thus itself an identity arrow. So the diamond commutes to $\mathrm{id}(\mathrm{F}(\mathrm{A}\,,\mathrm{X}))$. Further, the diamond has the form described in condition (ii), so the defined map, F(id , id), is indeed the identity.



binary composition In the diagram below, each of the four diamonds commutes by condition (ii), giving us the defined maps $F(f, p)$ and $F(g, q)$ as shown. By condition (i) each of the four outer triangles commutes, so by condition (ii) we have the defined map $F(f \cdot g, p \cdot q) : F(A, X) \to F(C, Z)$. By pasting, the whole diagram commutes, so F preserves composites as desired.



□

1.7.2 Subcategories

Just as we may restrict our attention to a subset of a given set, we may single out a substructure of a category as well. However, since a category has more structure than a set, we must ensure that the substructure in question remains a category.

Definition 1.7.2.1 (subcategory) Given a category $\mathbb{C}$, we may take a **subcategory** $\mathbb{D}$ of $\mathbb{C}$, written "$\mathbb{D} \subseteq \mathbb{C}$" by taking for $\mathbb{D}_0$ a subcollection of $\mathbb{C}_0$ and for $\mathbb{D}_1$ a subcollection of $\mathbb{C}_1$, subject to the restrictions:

- if $f :: \mathbb{C}$ is in $\mathbb{D}_1$ then $\partial^-(f)$ and $\partial^+(f)$ are in $\mathbb{D}_0$.
- if $f\,, g :: \mathbb{C}$ are in $\mathbb{D}_1$ and are composable in $\mathbb{C}$ then $f \cdot g$ is in $\mathbb{D}_1$.
- if $\mathrm{A} : \mathbb{C}$ is in $\mathbb{D}_0$ then $\mathrm{id}(\mathrm{A})$ is in $\mathbb{D}_1$.

The composition structure of arrows when interpreted in $\mathbb{D}$ is the same as in $\mathbb{C}$.

The restrictions are necessary to ensure that the subcollections of $\mathbb{C}_0$ and $\mathbb{C}_1$ we choose do, in fact, form a category.

Whenever we have a subcategory $\mathbb{D} \subseteq \mathbb{C}$, we have also an **inclusion functor** $i : \mathbb{D} \to \mathbb{C}$, written "$\mathbb{D} \hookrightarrow \mathbb{C}$", sending each object and arrow of $\mathbb{D}$ to itself, but now viewed as an object or arrow of $\mathbb{C}$.

1.7.3 Opposite Categories

Recall that each arrow in a category has two boundary objects, its *domain* and *codomain*. Systematically swapping these gives rise to an involutive relation on categories.

Definition 1.7.3.1 (opposite category) To any category $\mathbb{C}$, there corresponds an **opposite category**, $\mathbb{C}^\circ$ (pronounced "$\mathbb{C}$-op"), having:

objects: ${\mathbb{C}^\circ}_0 \;:=\; \mathbb{C}_0$

arrows: $\mathbb{C}^\circ\,(\mathrm{A} \to \mathrm{B}) \;:=\; \mathbb{C}\,(\mathrm{B} \to \mathrm{A})$

identities: $\mathrm{id}(\mathrm{A}) :: \mathbb{C}^\circ \;:=\; \mathrm{id}(\mathrm{A}) :: \mathbb{C}$

composition: $f \cdot g :: \mathbb{C}^\circ \;:=\; g \cdot f :: \mathbb{C}$

Exercise 1.7.3.2 Check that an opposite category satisfies the unit and associative laws of composition, and that the opposite of an opposite category is just the original category.

Despite being simple and purely formal, the opposite category construction is very useful. Because it is an *involution* (for any category $\mathbb{C}$, we have that $(\mathbb{C}^\circ)^\circ = \mathbb{C}$), op is called a **duality**.

For any construction that we may perform in a given category, we can view it from the perspective of the opposite category instead. this determines a **dual construction**. In some cases a construction and its dual may arise within the same category and interact in interesting ways (as, for example with the *distributive law*).

Furthermore, for any proposition that we may state about a given category, there is a **dual proposition** about its opposite category that is true just in case the first proposition is true of the original category. This gives us **dual theorems** for free: in category theory theorems are always two for the proof of one!

Functors respect the op duality in the sense that whenever we have a functor $\mathrm{F} : \mathbb{C} \to \mathbb{D}$, we automatically also have the functor $\mathrm{F}^{\circ} : \mathbb{C}^{\circ} \to \mathbb{D}^{\circ}$. F° is really just the same functor as F, it merely lets the categories on its boundary imagine that their arrows are going the other way round.

A functor $\mathrm{F} : \mathbb{C}^{\circ} \to \mathbb{D}$ is called a **contravariant functor** from $\mathbb{C}$ to $\mathbb{D}$. Among the most important contravariant functors one encounters are the contravariant representable functors:

Lemma 1.7.3.3 (contravariant representable functors) For each object of a locally small category, $\mathrm{X} : \mathbb{C}$, there is a functor,

$$\begin{array}{rcl} & \underline{\mathbb{C}\,(- \to \mathrm{X})} & \\ \mathbb{C}^{\circ} & \longrightarrow & \textsc{Set} \\ \mathrm{A} & \longmapsto & \mathbb{C}\,(\mathrm{A} \to \mathrm{X}) \\ f : \mathrm{A} \to \mathrm{B} & \longmapsto & \mathbb{C}\,(f \to \mathrm{X}) \coloneqq f \cdot - : \mathbb{C}\,(\mathrm{B} \to \mathrm{X}) \to \mathbb{C}\,(\mathrm{A} \to \mathrm{X}) \end{array}$$

This is just an ordinary *representable functor* on the opposite category: $\mathbb{C}\,(- \to \mathrm{X}) = \mathbb{C}^{\circ}\,(\mathrm{X} \to -)$, because pre-composition in $\mathbb{C}$ is the same thing as post-composition in $\mathbb{C}^{\circ}$. For reasons that we won't dwell on, a contravariant representable functor is also known as a **representable presheaf**.

Exercise 1.7.3.4 (hom bifunctor) Use the *bifunctor lemma* and the definitions of *covariant* and *contravariant* representable functors to define a **hom bifunctor** for locally small categories:

$$\mathbb{C}\,(\tfrac{0}{-} \to \tfrac{1}{-}) \quad : \quad \mathbb{C}^{\circ} \times \mathbb{C} \to \textsc{Set}$$

1.7.4 Arrow Categories

Definition 1.7.4.1 (arrow category) Given a category $\mathbb{C}$, we may derive from it another category, "$\mathbb{C}^{\to}$", known as the **arrow category** of $\mathbb{C}$ with the following structure:

objects: ${\mathbb{C}^{\to}}_0 \;\coloneqq\; \mathbb{C}_1$

arrows: $\mathbb{C}^{\to}\,(f \to g) \;\coloneqq\; \{(i\,,j) \mid i : \mathbb{C}\,(\partial^-(f) \to \partial^-(g))$ and $j : \mathbb{C}\,(\partial^+(f) \to \partial^+(g))$ such that $i \cdot g = f \cdot j\}$

identities: $\mathrm{id}(f) \;\coloneqq\; (\mathrm{id}(\partial^-(f))\,,\mathrm{id}(\partial^+(f)))$

composition: $(i\,,j) \cdot (k\,,l) \;\coloneqq\; (i \cdot k\,,j \cdot l)$

In a bit more detail, the objects of $\mathbb{C}^{\to}$ are the arrows of $\mathbb{C}$. Given $\mathbb{C}^{\to}$-objects, $f : \mathbb{C}\,(\mathrm{A} \to \mathrm{B})$ and $g : \mathbb{C}\,(\mathrm{C} \to \mathrm{D})$, a $\mathbb{C}^{\to}$-arrow from f to g is a pair of $\mathbb{C}$-arrows,

$i : \mathbb{C}\,(\mathrm{A} \to \mathrm{C})$ and $j : \mathbb{C}\,(\mathrm{B} \to \mathrm{D})$ that form a commuting square with f and g in $\mathbb{C}$:

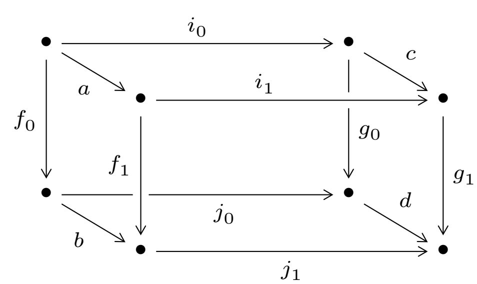
$$\begin{array}{ccc} \mathrm{A} & \xrightarrow{i} & \mathrm{C} \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle g} \\ \mathrm{B} & \xrightarrow[j]{} & \mathrm{D} \end{array} \tag{1.1}$$

Identity $\mathbb{C}^{\to}$-arrows are the commuting $\mathbb{C}$-squares with two opposite sides the same arrow and the other two opposite sides identity arrows. Composition in $\mathbb{C}^{\to}$ is the *pasting* of commuting squares in $\mathbb{C}$:

$$\begin{array}{ccc} \mathrm{A} & \xrightarrow{\mathrm{id}} & \mathrm{A} \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f} \\ \mathrm{B} & \xrightarrow[\mathrm{id}]{} & \mathrm{B} \end{array} \qquad\qquad \begin{array}{ccccc} \mathrm{A} & \xrightarrow{i} & \mathrm{C} & \xrightarrow{k} & \mathrm{E} \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle h} \\ \mathrm{B} & \xrightarrow[j]{} & \mathrm{D} & \xrightarrow[l]{} & \mathrm{F} \end{array}$$

The unit and associative laws of composition are satisfied in $\mathbb{C}^{\to}$ as a consequence of their holding in $\mathbb{C}$ (you should check this). So the arrows of $\mathbb{C}^{\to}$ are the commuting squares of $\mathbb{C}$ (with each commuting $\mathbb{C}$-square represented twice). This tells us something about the 2-dimensional structure of $\mathbb{C}$, namely, which of its squares commute.

We can iterate this construction to explore yet higher-dimensional structure of $\mathbb{C}$. One dimension up, the category $(\mathbb{C}^{\to})^{\to}$ has as objects $\mathbb{C}^{\to}$-arrows (i.e. $\mathbb{C}$-commuting squares) and as arrows $\mathbb{C}^{\to}$-commuting squares. Let's see what these ought to be. A nice way to think about it is to take diagram (1.1) and imagine that it's actually a 3-dimensional cube that we happen to be seeing orthographically along one face. If we shift our perspective slightly, we will see the following:



(1.2)

We begin with $(\mathbb{C}^{\to})^{\to}$-objects f and g, which are actually the $\mathbb{C}^{\to}$-arrows from a to b and from c to d, respectively. These, in turn, are the $\mathbb{C}$-commuting squares shown on the left and right of diagram (1.2). Now $(\mathbb{C}^{\to})^{\to}$-arrows between these will be $\mathbb{C}^{\to}$-arrows between their domains and codomains, i and j, which are the $\mathbb{C}$-commuting squares shown on the top and bottom. But there is also the condition that $i \cdot g = f \cdot j$ in $\mathbb{C}^{\to}$. Composition in $\mathbb{C}^{\to}$ is pasting in $\mathbb{C}$, and equality of arrows in $\mathbb{C}^{\to}$ is just pairwise equality in $\mathbb{C}$. So we need that $i_0 \cdot g_0 = f_0 \cdot j_0$ and $i_1 \cdot g_1 = f_1 \cdot j_1$ in $\mathbb{C}$, making the back and front faces commute. In other words, the top and bottom commuting $\mathbb{C}$-squares form a $(\mathbb{C}^{\rightarrow})^{\rightarrow}$-arrow between the left and right commuting $\mathbb{C}$-squares just in case the front and back $\mathbb{C}$-squares commute as well. Then all the paths shown in diagram (1.2) commute. So $(\mathbb{C}^{\rightarrow})^{\rightarrow}$-arrows are $\mathbb{C}$-commuting cubes.

The arrow category construction provides us with three important functors, that in a sense "mediate between dimensions". These are the **domain**, **reflexivity** and **codomain** functors:

$$\begin{array}{rcl} & \underline{dom} & \\ \mathbb{C}^{\rightarrow} & \longrightarrow & \mathbb{C} \\ f :: \mathbb{C} & \longmapsto & \partial^{-}(f) \\ (i\,,j) & \longmapsto & i \end{array} \qquad \begin{array}{rcl} & \underline{refl} & \\ \mathbb{C} & \longrightarrow & \mathbb{C}^{\rightarrow} \\ \mathrm{A} & \longmapsto & \mathrm{id}(\mathrm{A}) \\ f & \longmapsto & (f\,,f) \end{array} \qquad \begin{array}{rcl} & \underline{cod} & \\ \mathbb{C}^{\rightarrow} & \longrightarrow & \mathbb{C} \\ f :: \mathbb{C} & \longmapsto & \partial^{+}(f) \\ (i\,,j) & \longmapsto & j \end{array}$$

We will see later that these functors play an important role in the higher-dimensional structure of categories, but for now we will use the codomain functor to construct another important new category from old.

1.7.5 Slice Categories

Definition 1.7.5.1 (slice category) Given a category $\mathbb{C}$ and object $\mathrm{A} : \mathbb{C}$, there is a category, "$\mathbb{C}/\mathrm{A}$" called the **slice category** of $\mathbb{C}$ over A, with the following structure:

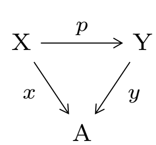
objects: $(\mathbb{C}/\mathrm{A})_0 \;:=\; \{x :: \mathbb{C} \mid \partial^{+}(x) = \mathrm{A}\}$

arrows: $\mathbb{C}/\mathrm{A}\,(x \to y) \;:=\; \{p : \mathbb{C}\,(\partial^{-}(x) \to \partial^{-}(y)) \mid p \cdot y = x\}$

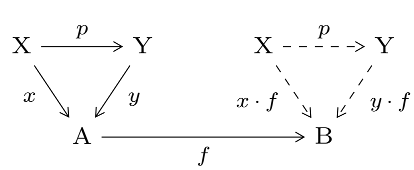
Identities and composition are inherited from $\mathbb{C}$.

The slice category $\mathbb{C}/\mathrm{A}$ is a subcategory of the arrow category $\mathbb{C}^{\rightarrow}$. It contains just those objects and arrows that the *cod* functor sends to A and to id(A), respectively.

I imagine the arrows of $\mathbb{C}/\mathrm{A}$ as the bases of inverted triangles with their vertices anchored at A:



Composing the sides of such triangles with an arrow $f : \mathrm{A} \to \mathrm{B}$ lets us move the anchor to B:



Lemma 1.7.5.2 (post-composition functor) Every arrow $f : \mathbb{C}\,(\mathrm{A} \to \mathrm{B})$ determines a functor:

$$\begin{array}{rcl} & \underline{f_!} & \\ \mathbb{C}/\mathrm{A} & \longrightarrow & \mathbb{C}/\mathrm{B} \\ x : \mathbb{C}\,(\mathrm{X} \to \mathrm{A}) & \longmapsto & x \cdot f \\ p : \mathbb{C}\,(\mathrm{X} \to \mathrm{Y}) & \longmapsto & p \end{array}$$

Chapter 2

Behavioral Reasoning

A fundamental question we must address when studying any kind of formal system is when two objects with distinct presentations should be considered to be essentially the same. We can ask this question about sets, groups, topological spaces, λ-terms, and even categories.

Certainly, whatever relation we choose should be an equivalence relation and should be a congruence for certain operations, but beyond that, general guidelines are hard to come by.

For example, we consider two sets to be essentially the same if there is a **bijection** between them, that is, if there is an injective and surjective function from one to the other. Recall that a function $p : \mathrm{X} \to \mathrm{Y}$ is **injective** if it "doesn't collapse any elements of its domain":

$$\forall\ x_0\,, x_1 \in \mathrm{X}\ .\ p(x_0) = p(x_1)\ \supset\ x_0 = x_1$$

and is **surjective** if it "doesn't miss any elements of its codomain":

$$\forall\ y \in \mathrm{Y}\ .\ \exists\ x \in \mathrm{X}\ .\ p(x) = y$$

We can't translate such element-wise definitions directly to the language of categories because the objects of a category need not be structured sets, so we must find equivalent *behavioral* characterizations.

2.1 Monic and Epic Morphisms

2.1.1 Monomorphisms

In the case of injections, we can do this by rephrasing the property so that rather than discussing the image under p of two points of X, we instead discuss the composition with p of two parallel functions into X. If p is injective then:

$$\forall\ f\ ,g : \mathrm{W} \to \mathrm{X}\ .\ \forall\ w \in \mathrm{W}\ .\ (p \circ f)(w) = (p \circ g)(w)\ \supset\ f(w) = g(w)$$

Universal quantification distributes over implication, i.e. if $\forall a : \mathrm{A}\ .\ \varphi\ a\ \supset\ \psi\ a$ then $(\forall a : \mathrm{A}\ .\ \varphi\ a)\ \supset\ (\forall a : \mathrm{A}\ .\ \psi\ a)$. Thus if p is injective then:

$$\forall\ f,g : \mathrm{W} \to \mathrm{X}\ .\ (\forall\ w \in \mathrm{W}\ .\ (p \circ f)(w) = (p \circ g)(w)) \supset (\forall\ w \in \mathrm{W}\ .\ f(w) = g(w))$$

It may seem that we've just made things worse by introducing two extraneous functions, but now we can use the fact that two functions are equal just in case they agree on all points to rephrase this again, doing away with the points entirely. So if p is injective then:

$$\forall\ f\ ,g : \mathrm{W} \to \mathrm{X}\ .\ f \cdot p = g \cdot p\ \supset\ f = g$$

This is a behavioral characterization that can be stated for any category.

Definition 2.1.1.1 (monomorphism) An arrow $m :: \mathbb{C}$ is a **monomorphism** (or "monic") if it is *post-cancelable*; that is, if for any arrows $f\ ,g :: \mathbb{C}$,

$$f \cdot m \quad = \quad g \cdot m \qquad \text{implies} \qquad f \quad = \quad g$$

Notice that we are being a bit economical here: in order for f and g to be composable with m, they must be *coterminal*, and in order for their composites with m to be equal, they must also be *coinitial*. So the implication is applicable only to *parallel* f and g composable with m, but all that can be inferred.

In diagrams, monomorphisms are conventionally drawn with a tailed arrow: "$\rightarrowtail$".

Lemma 2.1.1.2 (monics and composition)

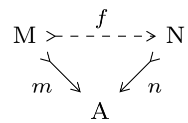
- Identity morphisms are monic.
- Composites of monics are monic.
- If the composite $m \cdot n$ is monic then so is m.

Proof.

$$\begin{array}{ll} & f \cdot \mathrm{id} \ = \ g \cdot \mathrm{id} \\ \Rightarrow & [\text{unit law}] \\ & f \ = \ g \end{array} \qquad \begin{array}{ll} & f \cdot m \cdot n \ = \ g \cdot m \cdot n \\ \Rightarrow & [n \text{ is monic}] \\ & f \cdot m \ = \ g \cdot m \\ \Rightarrow & [m \text{ is monic}] \\ & f \ = \ g \end{array} \qquad \begin{array}{ll} & f \cdot m \ = \ g \cdot m \\ \Rightarrow & [\text{whiskering}] \\ & f \cdot m \cdot n \ = \ g \cdot m \cdot n \\ \Rightarrow & [m \cdot n \text{ is monic}] \\ & f \ = \ g \end{array}$$

□

Remark 2.1.1.3 (subobjects) If we take the *subcategory* of a *slice category* containing just the monomorphisms then we get a *preorder*: given monics $m\,, n : \mathbb{C}/\mathrm{A}$, we say that $m \leq n$ just in case there is an $f : \mathbb{C}/\mathrm{A}\,(m \to n)$.



Such an f, if it exists, is unique because n is monic, and is itself monic by the preceding lemma.

The monics into an object behave very much like the partial order of subsets of a set, in fact, they are known as (representatives of) **subobjects**. This is the beginning of a branch of categorical logic known as **topos theory**, where subobjects are used to interpret predicates. However, using a preorder to interpret the entailment relation on propositions sacrifices *proof relevance*: it lets us say *that* one proposition entails another, but not *why* it does so.

2.1.2 Epimorphisms

Using the *op duality*, we can define the property dual to that of being monic. You should check that this amounts to the following:

Definition 2.1.2.1 (epimorphism) An arrow $e :: \mathbb{C}$ is an **epimorphism** (or "epic") if it is *pre-cancelable*; that is, if for any arrows $f\,, g :: \mathbb{C}$,

$$e \cdot f \quad = \quad e \cdot g \qquad \text{implies} \qquad f \quad = \quad g$$

This corresponds to the fact that a surjective function doesn't miss any points in its codomain, so if $p : \mathrm{X} \to \mathrm{Y}$ is surjective then for any parallel $f\,, g : \mathrm{Y} \to \mathrm{Z}$,

$$(\forall x \in \mathrm{X}\;.\;(f \circ p)(x) = (g \circ p)(x)) \quad \supset \quad (\forall y \in \mathrm{Y}\;.\;f(y) = g(y))$$

Eliminating the points gives us the definition of epimorphism.

In diagrams, epimorphisms are conventionally drawn with a double-headed arrow: "$\twoheadrightarrow$".

Exercise 2.1.2.2 State and prove the *dual theorems* to those in lemma 2.1.1.2.

In the category Set a function is injective just in case it is monic, and surjective just in case it is epic. In many categories of "structured sets" (e.g. Mon) the monomorphisms are exactly the injective homomorphisms. For instance, the inclusion $(\mathbb{N}\,, +\,, 0) \hookrightarrow (\mathbb{Z}\,, +\,, 0)$ in the category Mon is a monomorphism. It turns out to be an epimorphism as well, despite not being surjective on its underlying set. So, unlike the situation in Set, in an arbitrary category the existence of a monic and epic morphism between two objects does not suffice to ensure that they are *essentially the same*.

Monic and epic morphisms have other unsatisfactory properties, for example, they are not necessarily preserved by functors (the existence of a *forgetful functor* from monoids to sets, together with the last result implies this).

2.2 Split Monic and Epic Morphisms

Definition 2.2.0.1 (split monomorphism) An arrow s is a **split monomorphism** (or "split monic") if it is post-(semi-)invertible; that is, if there exists an arrow r such that $s \cdot r = \text{id}$.

The dual notion is that of:

Definition 2.2.0.2 (split epimorphism) An arrow r is a **split epimorphism** (or "split epic") if it is pre-(semi-)invertible; that is, if there exists an arrow s such that $s \cdot r = \text{id}$.

It would be perverse to name them this way unless split monics were monic and split epics were epic, which indeed they are.

Lemma 2.2.0.3 A split monomorphism is a monomorphism (and a split epimorphism is an epimorphism).

Proof. Suppose s is split-monic with $s \cdot r = \text{id}$,

$$\begin{array}{ll} & f \cdot s = g \cdot s \\ \Longrightarrow & [\text{whiskering}] \\ & f \cdot s \cdot r = g \cdot s \cdot r \\ \Longrightarrow & [\text{assumption}] \\ & f \cdot \text{id} = g \cdot \text{id} \\ \Longrightarrow & [\text{unit law}] \\ & f = g \end{array}$$

The other case is dual. □

Because *functors* must preserve the composition structure of categories, they must preserve split monics and epics as well.

Lemma 2.2.0.4 The functor-image of a split monic (split epic) is itself split monic (split epic).

Proof. Suppose s is split-monic with $s \cdot r = \mathrm{id}$,

$$\begin{array}{ll} & \mathrm{F}(s) \cdot \mathrm{F}(r) \\ = & [\text{functors preserve composition}] \\ & \mathrm{F}(s \cdot r) \\ = & [\text{assumption}] \\ & \mathrm{F}(\mathrm{id}) \\ = & [\text{functors preserve identities}] \\ & \mathrm{id} \end{array}$$

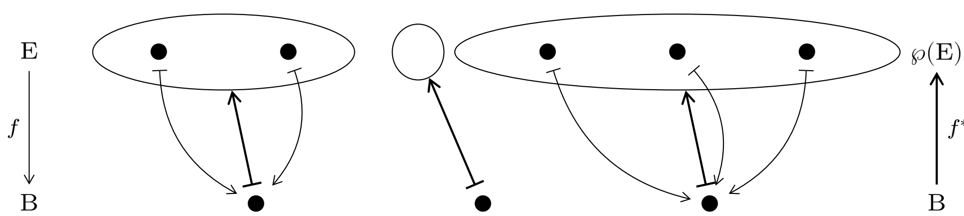
So $\mathrm{F}(s)$ is split-monic. □

Before moving on, let's consider a particularly pretty application of behavioral reasoning to the axiom of choice. This proposition states that given a family of non-empty sets, there is a function that chooses an element from each one.

We can represent any family of sets with an ordinary function in the following way. Given a function $f : \textsc{Set}\,(\mathrm{E} \to \mathrm{B})$, we can define a function,

$$\begin{array}{rcl} \mathrm{B} & \overset{f^*}{\longrightarrow} & \wp(\mathrm{E}) \hookrightarrow \textsc{Set} \\ b & \longmapsto & \{e \in \mathrm{E} \mid f(e) = b\} \end{array}$$

And given a family of sets, $\{\mathrm{E}_b\}_{b \in \mathrm{B}}$, which is just an map $\mathrm{B} \to \textsc{Set}$, we can define a *projection function* $\int_{b \in \mathrm{B}} \mathrm{E}_b \to \mathrm{B}$ mapping $e \in \mathrm{E}_b \longmapsto b$. These two constructions are inverse, both the function $f : \mathrm{E} \to \mathrm{B}$ and the family of sets $\{\mathrm{E}_b\}_{b \in \mathrm{B}}$ just sort the elements of E by those in B:



The **axiom of choice** states that if for each $b \in \mathrm{B}$ the set $f^*(b)$ is non-empty then there is a way to choose from E a family of elements $\{e_b\}_{b \in \mathrm{B}}$ such that $\forall\ b \in \mathrm{B}\ .\ f(e_b) = b$ – i.e. such that there is a function $s : \mathrm{B} \to \mathrm{E}$ with $s \cdot f = \mathrm{id}(\mathrm{B})$. Notice that the condition that the sets $f^*(b)$ be non-empty is equivalent to the requirement that f be a surjection. So the axiom of choice asserts that in the category Set, every epimorphism is split!

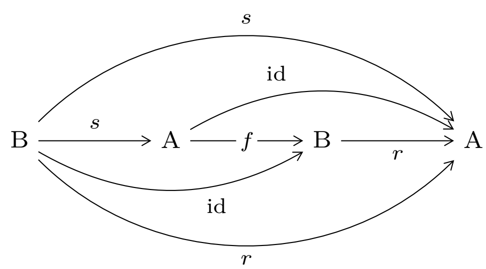
This is a behavioral characterization of a property that we may ask whether a given category satisfies. For example, because the inclusion $(\mathbb{N}, +, 0) \hookrightarrow (\mathbb{Z}, +, 0)$ is epic in the category Mon, it fails to hold there.

2.3 Isomorphisms

When we have arrows $s : \mathrm{A} \to \mathrm{B}$ and $r : \mathrm{B} \to \mathrm{A}$ such that $s \cdot r = \mathrm{id}(\mathrm{A})$ we say that s is a **section** of r and that r is a **retraction** of s. So *being* split monic means *having* a retraction, and *being* split epic means *having* a section.

Lemma 2.3.0.1 If a morphism has both a section and a retraction then the section and the retraction are equal.

Proof. Given arrow $f : \mathrm{A} \to \mathrm{B}$ with section s and retraction r, by pasting, $s = r$:



□

If $f : \mathrm{A} \to \mathrm{B}$ has section-and-retraction g, then $g : \mathrm{B} \to \mathrm{A}$ necessarily has retraction-and-section f. In other words, f and g are (two-sided) inverses for one another. This leads us to a good behavioral characterization of what it means for objects to be *essentially the same* in any category.

Definition 2.3.0.2 (isomorphism) An arrow $f : \mathrm{A} \to \mathrm{B}$ is an **isomorphism** if there exists an *anti-parallel* arrow $g : \mathrm{B} \to \mathrm{A}$, called an **inverse** of f, such that:

$$f \cdot g = \mathrm{id}(\mathrm{A}) \qquad \text{and} \qquad g \cdot f = \mathrm{id}(\mathrm{B})$$

It follows from lemma 2.3.0.1 that an inverse of f is unique, so we can write it unambiguously as "f^{-1}". To indicate the existence of an unspecified isomorphism between objects A and B, we write "$\mathrm{A} \cong \mathrm{B}$" and call the objects **isomorphic**.

Isomorphism is the right notion of "essentially the same" for objects of an arbitrary category because in categories we must characterize objects behaviorally, and there is generally no way to distinguish objects that behave identically.

Exercise 2.3.0.3 Show that isomorphism of objects is an equivalence relation; that is, for any objects A, B and C,

reflexivity: $\mathrm{A} \cong \mathrm{A}$

symmetry: $\mathrm{A} \cong \mathrm{B}$ implies $\mathrm{B} \cong \mathrm{A}$

transitivity: $\mathrm{A} \cong \mathrm{B}$ and $\mathrm{B} \cong \mathrm{C}$ implies $\mathrm{A} \cong \mathrm{C}$

Definition 2.3.0.4 (groupoid) A category in which every arrow is an isomorphism is called a **groupoid**. In particular, a single-object groupoid is a **group**.

Exercise 2.3.0.5 Check that this definition coincides with the usual definition of a group.

Chapter 3

Universal Constructions

A **universal construction** is a description of a construction within a category that determines it uniquely up to a canonical isomorphism. This is the best kind of description we can hope for in a behavioral setting, where we do not have direct access to the internal structure of the objects we are working with.

Universal constructions are defined using **universal properties**, which assert that the construction itself has some property, and that if any other construction in the category has the same property then there is a canonical relationship between the two.

In this chapter we introduce the universal constructions needed for the categorical interpretation of typing contexts and simple type formers. We do this in a deliberately methodical way, in order to emphasize the similarities in the constructions.

3.1 Terminal and Initial Objects

3.1.1 Terminal Objects

In the category Set, a singleton set S has the property that given any set X there is a unique function from X to S, namely, the constant function on the only element of S. This is a behavioral characterization that we may state in an arbitrary category.

Definition 3.1.1.1 (terminal object) In any category, a **terminal object** is an object T with the property that for any object X there is a unique morphism $x : \mathrm{X} \to \mathrm{T}$.

We write "!(X)" for the unique map from an object X to a terminal object and refer to it as a **bang map**.

Whenever some construction has a certain relationship to all constructions of the same kind within a category, it must, in particular, have this relationship to itself. Socrates' dictum to "know thyself" is as important in category theory as it is in life. So whenever we encounter a universal construction we will see what we can learn about it by "probing it with itself". In the case of a terminal object, this means choosing $X := T$ in the definition.

Lemma 3.1.1.2 (identity expansion for terminals) If T is a terminal object then $!(T) = id(T)$.

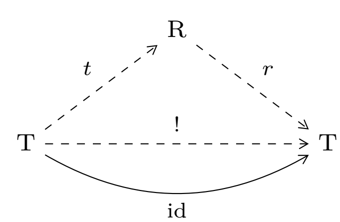
Proof. By assumption, $!(T)$ is the unique map $t : T \to T$, but $id(T)$ is an arrow in the same hom set. ☐

Universal constructions are each unique up to a unique structure-preserving isomorphism. In the case of a terminal object, the structure to be preserved is trivial: it's just a single object. Consequently, we obtain an especially strong uniqueness property.

Lemma 3.1.1.3 (uniqueness of terminals) When they exist, terminal objects are unique up to a unique isomorphism.

Proof. Suppose that T and R are two terminal objects in a category. By assumption, there are unique arrows $t : T \to R$ and $r : R \to T$ and:

$$\begin{aligned} & t \cdot r : T \to T \\ = \; & [\text{T is terminal}] \\ & !(T) : T \to T \\ = \; & [\text{identity expansion for terminals}] \\ & id(T) : T \to T \end{aligned}$$



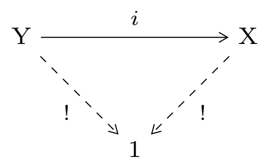
Symmetrically, we have that $r \cdot t = id(R)$. So t is an *isomorphism.* By the universal property of R, the hom set $T \to R$ is a singleton, so it must be the only one. ☐

Because terminal objects are unique up to unique isomorphism, we write "1" to refer to an arbitrary terminal object of a category.

Exercise 3.1.1.4 (pre-composing with a bang) Use the universal property of a terminal object to prove the following:

For a terminal object 1 and arrow $i : \mathrm{Y} \to \mathrm{X}$,

$$i \cdot !(\mathrm{X}) \quad = \quad !(\mathrm{Y}) \quad : \quad \mathrm{Y} \to 1$$



As mentioned, in SET, any singleton set is terminal. Likewise, in CAT, any *singleton category* is. In MON, the trivial monoid (having only the identity element) is terminal.

Exercise 3.1.1.5 Work out what a terminal object is in the category PREORD, and determine when a preordered set, as a category, has a terminal object.

3.1.2 Unit Type

The terminal object universal construction provides a categorical interpretation of the unit type, $\top$,

$$[\![\top]\!] \quad := \quad 1$$

The introduction rule for unit type,

$$\frac{}{\Gamma \ \vdash \ \star : \top} \ \top+$$

is interpreted by the bang map:

$$[\![\star]\!] \quad := \quad !([\![\Gamma]\!]) : [\![\Gamma]\!] \to 1$$

3.1.3 Global and Generalized Elements

In SET, there is a bijection between the elements of a set X and the functions from a singleton set to X: to each $x \in \mathrm{X}$ there corresponds the unique function $\ulcorner x \urcorner : 1 \to \mathrm{X}$ mapping $\star \longmapsto x$. We can use this behavioral characterization to define an analogue for set membership.

Definition 3.1.3.1 (global element) In a category with a terminal object, a **global element** (or "point") of an object X is an element of the hom set $1 \to \mathrm{X}$.

Definition 3.1.3.2 (generalized element) In contrast, a **generalized element** of an object X is just a morphism with codomain X; in other words, an object of the *slice category* over X.

In SET, we can determine whether or not two functions are the same by probing them with points because two parallel functions f , g : SET (X → Y) are equal just in case $\forall x \in \mathrm{X} \;.\; f(x) = g(x)$. This is known as the principle of **function extensionality**. Here is a categorical analogue:

Definition 3.1.3.3 (well-pointed category) A category with a terminal object is **well-pointed** if for every f , g : A → B, and global element $a : 1 \to \mathrm{A}$,

$$a \cdot f = a \cdot g \qquad \text{implies} \qquad f = g$$

Notice the similarity to the definition of an *epimorphism*. In fact, we say that a category is well-pointed if its points are *jointly epic*, that is, if points are collectively able to distinguish arrows.

Exercise 3.1.3.4 In contrast to the case with global elements, in any category we can determine whether or not two parallel arrows are the same by probing them with *generalized elements*. Prove this. (*Hint*: for any pair of arrows, a single "probe" suffices.)

3.1.4 Initial Objects

The concept *dual* to that of a terminal object is of an initial object.

Definition 3.1.4.1 (initial object) In any category, an **initial object** is an object S with the property that for any object X there is a unique morphism $x : \mathrm{S} \to \mathrm{X}$

We write "¡(X)" for the unique map in S → X and refer to it as a **cobang map**. By probing an initial object with itself we obtain a result dual to lemma 3.1.1.2:

Lemma 3.1.4.2 (identity expansion for initials) If S is an initial object then ¡(S) = id(S).

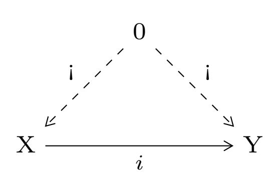
Exercise 3.1.4.3 (uniqueness of initials) Check that when they exist, initial objects are unique up to a unique isomorphism.

We write "0" to refer to an arbitrary initial object of a category.

Lemma 3.1.4.4 (post-composing with a cobang) dual to exercise 3.1.1.4:

For an initial object 0 and arrow $i : \mathrm{X} \to \mathrm{Y}$,

$$¡(\mathrm{X}) \cdot i \quad = \quad ¡(\mathrm{Y}) \quad : \quad 0 \to \mathrm{Y}$$



In SET, the empty set is initial. Likewise, in CAT, the *empty category* is. In MON, the trivial monoid is initial as well as terminal. (An object which is both terminal and initial is known as a **null object**.)

Exercise 3.1.4.5 Dualize exercise 3.1.1.5 by working out what an initial object is in the category PREORD, and determine when a preordered set, as a category, has an initial object.

3.1.5 Void Type

The initial object universal construction provides a possible categorical interpretation of the void type, $\bot$,

$$[\![\bot]\!] \quad := \quad 0$$

In the restricted setting where contexts are singletons, the elimination rule for void type,

$$\frac{}{z : \bot \ \vdash \ \mathrm{abort}^\dagger : \mathrm{A}} \ \bot{-}^\dagger$$

is then interpreted by the cobang map:

$$[\![\mathrm{abort}^\dagger]\!] \quad := \quad ¡([\![\mathrm{A}]\!]) : 0 \to [\![\mathrm{A}]\!]$$

Note that in this case, the interpretations of all terms of a given type in an inconsistent context coincide:

$$\forall (z : \bot \ \vdash \ \mathrm{M} \,, \mathrm{N} : \mathrm{A}) \ . \ [\![\mathrm{M}]\!] \cong [\![\mathrm{N}]\!]$$

Depending on how the type theory is set up, this equation may or may not be reflected there.

3.2 Products

3.2.1 Products of Objects

The *set of ordered pairs*:

$$\mathrm{A} \times \mathrm{B} \quad := \quad \{(a \,, b) \mid a \in \mathrm{A} \text{ and } b \in \mathrm{B}\}$$

comes equipped with two projection functions,

$$\begin{array}{ccc} \mathrm{A} \times \mathrm{B} & \xrightarrow{\pi_0} & \mathrm{A} \\ (a \,, b) & \longmapsto & a \end{array} \qquad \text{and} \qquad \begin{array}{ccc} \mathrm{A} \times \mathrm{B} & \xrightarrow{\pi_1} & \mathrm{B} \\ (a \,, b) & \longmapsto & b \end{array}$$

such that for ordered pair $c \in \mathrm{A} \times \mathrm{B}$,

$$c \;=\; (\pi_0\ c \,,\, \pi_1\ c)$$

So having a pair of elements, one from the set A and one from the set B, is the same thing as having a single element of the set $\mathrm{A} \times \mathrm{B}$: given an $a \in \mathrm{A}$ and $b \in \mathrm{B}$ we make an element of $\mathrm{A} \times \mathrm{B}$ by forming the tuple $(a \,,\, b)$, and given an element $c \in \mathrm{A} \times \mathrm{B}$ we recover elements of A and B by taking the projections.

Not every category is *well-pointed* like Set is, or for that matter, even has a terminal object. So to describe this situation behaviorally we must use *generalized elements*. This motivates the definition of products in an arbitrary category.

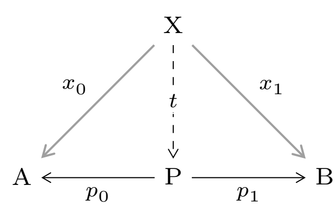
Definition 3.2.1.1 (product of objects) In any category, a cartesian[1] **product** of objects A and B is a *span* on A and B,

$$\mathrm{A} \xleftarrow{p_0} \mathrm{P} \xrightarrow{p_1} \mathrm{B}$$

with the property that for any span on A and B,

$$\mathrm{A} \xleftarrow{x_0} \mathrm{X} \xrightarrow{x_1} \mathrm{B}$$

there is a unique map $t : \mathrm{X} \to \mathrm{P}$ such that $t \cdot p_0 = x_0$ and $t \cdot p_1 = x_1$:



This says that there is a bijection between ordered pairs of maps $(x_0 \,,\, x_1)$ and single maps t such that the diagram commutes. We call A and B the **factors** of the product, p_0 and p_1 its (coordinate) **projections** and t the **tuple** of x_0 and x_1 and write it as "$\langle x_0 \,,\, x_1 \rangle$".

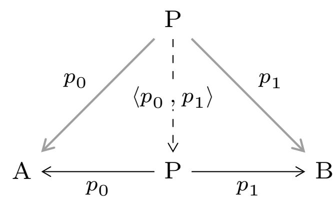
Let's see what we can learn from probing a product with itself by choosing $\mathrm{X} := \mathrm{P}$ and $(x_0 \,,\, x_1) := (p_0 \,,\, p_1)$.

Lemma 3.2.1.2 (identity expansion for products) If P is a product of A and B with projections p_0 and p_1, then $\langle p_0 \,,\, p_1 \rangle = \mathrm{id}(\mathrm{P})$.

Proof. By assumption, $\langle p_0 \,,\, p_1 \rangle$ is the unique map $t : \mathrm{P} \to \mathrm{P}$ with the property

[1] The word "cartesian" is sometimes used for emphasis to distinguish this construction from various other categorical constructions also known as "products" (e.g. the *monoidal* product, $- \otimes -$). But this is the only "product" that we will consider in this course.

that, $t \cdot p_0 = p_0$ and $t \cdot p_1 = p_1$:



but by the left unit law of composition, id(P) has this property. ☐

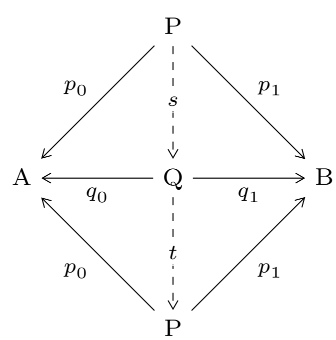
Because products are structures characterized by a universal property, we expect them to be uniquely determined up to a unique structure-preserving isomorphism. This is indeed the case:

Lemma 3.2.1.3 (uniqueness of products) When they exist, products of objects are unique up to a unique projection-preserving isomorphism.

Proof. Suppose that the spans:

$$\mathrm{A} \xleftarrow{p_0} \mathrm{P} \xrightarrow{p_1} \mathrm{B} \qquad \text{and} \qquad \mathrm{A} \xleftarrow{q_0} \mathrm{Q} \xrightarrow{q_1} \mathrm{B}$$

are both products of A and B. Because Q is a product there is a unique $s : \mathrm{P} \to \mathrm{Q}$ such that $s \cdot q_0 = p_0$ and $s \cdot q_1 = p_1$. Likewise, because P is a product there is a unique $t : \mathrm{Q} \to \mathrm{P}$ such that $t \cdot p_0 = q_0$ and $t \cdot p_1 = q_1$:



Then for $i \in \{0 , 1\}$:

$$\begin{array}{ll} & s \cdot t \cdot p_i \\ = & [\text{P is a product}] \\ & s \cdot q_i \\ = & [\text{Q is a product}] \\ & p_i \end{array}$$

Thus $s \cdot t = \langle p_0 , p_1 \rangle : \mathrm{P} \to \mathrm{P}$. By identity expansion for products, $s \cdot t = \mathrm{id}(\mathrm{P})$. Reversing the roles of P and Q, we get that $t \cdot s = \mathrm{id}(\mathrm{Q})$ as well. So s is an isomorphism. By the universal property of Q, it is the only one that respects the coordinate projections. □

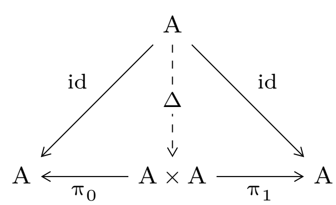
Because products are determined as uniquely as is possible by a behavioral characterization, we write "A × B" to refer to an arbitrary product of A and B. When the product in question is clear from context, we refer to the two coordinate projections generically as "π_0" and "π_1".

In the category SET, the *set of ordered pairs* is a cartesian product. Likewise, in CAT, the *ordered pair category* is. This justifies the notation − × − that we used in both cases.

Note that unlike the case with terminal objects, there is not necessarily a unique isomorphism between two products of the same factors. For example, in SET the identity function, $(x , y) \longmapsto (x , y)$, and swap map, $(x , y) \longmapsto (y , x)$, are both isomorphisms $\mathrm{A} \times \mathrm{A} \to \mathrm{A} \times \mathrm{A}$. But only the former respects the coordinate projections.

Definition 3.2.1.4 (diagonal map) For every object A, the universal property of the product gives a canonical **diagonal map**, which duplicates its argument:

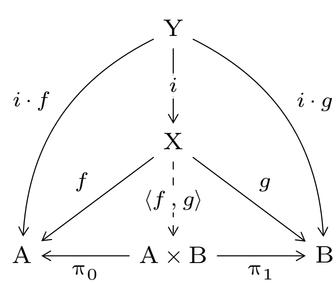
$$\Delta(\mathrm{A}) \quad := \quad \langle \mathrm{id}(\mathrm{A}) , \mathrm{id}(\mathrm{A}) \rangle \quad : \quad \mathrm{A} \to \mathrm{A} \times \mathrm{A}$$



Exercise 3.2.1.5 (pre-composing with a tuple) Use the diagram below and the universal property of a product of objects to prove the following:

For a product A × B, a tuple $\langle f , g \rangle : \mathrm{X} \to \mathrm{A} \times \mathrm{B}$ and an arrow $i : \mathrm{Y} \to \mathrm{X}$,

$$i \cdot \langle f , g \rangle \;=\; \langle i \cdot f , i \cdot g \rangle \;:\; \mathrm{Y} \to \mathrm{A} \times \mathrm{B}$$



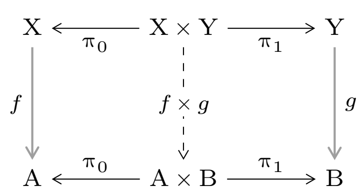
3.2.2 Product Functors

We can use the universal property of a product of objects to define a product of arrows as well:

Definition 3.2.2.1 (product of arrows) Given a pair of arrows $f : \mathrm{X} \to \mathrm{A}$ and $g : \mathrm{Y} \to \mathrm{B}$, and products $\mathrm{X} \times \mathrm{Y}$ and $\mathrm{A} \times \mathrm{B}$, we define the **product of arrows** by:

$$\begin{aligned} f \times g &: \mathrm{X} \times \mathrm{Y} \to \mathrm{A} \times \mathrm{B} \\ f \times g &:= \langle \pi_0 \cdot f \,, \pi_1 \cdot g \rangle \end{aligned}$$

By the universal property of the product $\mathrm{A} \times \mathrm{B}$, the arrow $f \times g$ is the unique morphism making the two squares commute:



This allows us to characterize the product as a functor – indeed, a *bifunctor*:

Lemma 3.2.2.2 (functoriality of products) If a category $\mathbb{C}$ has products for each pair of objects, then the given definition of products for arrows yields a functor,

$$- \times - \;:\; \mathbb{C} \times \mathbb{C} \to \mathbb{C}$$

called the **product functor**.

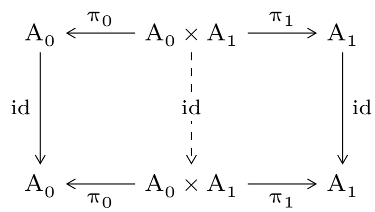
Before giving the proof, we pause to explain this statement, as it is easy to be confused about what is being asserted. In the theorem, "$\mathbb{C} \times \mathbb{C}$" is the *ordered pair category* (definition 1.7.1.1); i.e. the product of $\mathbb{C}$ with itself in Cat. In contrast, "$-\times-$" is the *name* of an alleged functor having as domain the category $\mathbb{C} \times \mathbb{C}$ and as codomain the category $\mathbb{C}$.

Proof. In order to prove that $- \times -$ is a functor, we must show that it preserves the composition structure.

nullary composition: We must show that

$$\mathrm{id}(\mathrm{A}_0) \times \mathrm{id}(\mathrm{A}_1) \;=\; \mathrm{id}(\mathrm{A}_0 \times \mathrm{A}_1)$$

In the diagram,

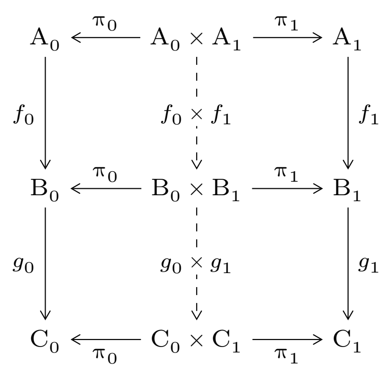


the arrow $\mathrm{id}(A_0 \times A_1)$ makes both squares commute, so the result follows by the definition of product of arrows.

binary composition: We must show that

$$(f_0 \cdot g_0) \times (f_1 \cdot g_1) \quad = \quad (f_0 \times f_1) \cdot (g_0 \times g_1)$$

In the diagram,



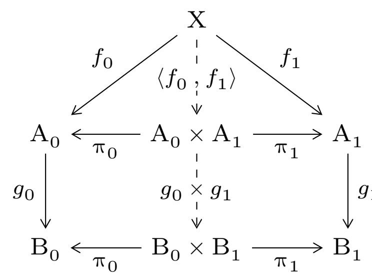
the top two squares commute by the definition of $f_0 \times f_1$ and the bottom two squares commute by the definition of $g_0 \times g_1$. By pasting, the rectangle comprising the two left squares commutes, and likewise the rectangle comprising the two right squares. By definition, $(f_0 \cdot g_0) \times (f_1 \cdot g_1)$ is the unique arrow from $A_0 \times A_1$ to $C_0 \times C_1$ making the outer square commute.

□

Exercise 3.2.2.3 (post-composing a product of arrows) Use the universal property of a product of objects to prove the following:

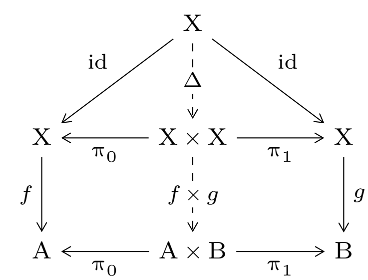
For arrows $\langle f_0\, , f_1 \rangle : X \to A_0 \times A_1$ and $g_0 \times g_1 : A_0 \times A_1 \to B_0 \times B_1$,

$$\langle f_0\, , f_1 \rangle \cdot (g_0 \times g_1) \quad = \quad \langle f_0 \cdot g_0\, , f_1 \cdot g_1 \rangle$$



Corollary 3.2.2.4 (tuple factorization) A tuple $\langle f , g \rangle : X \to A \times B$ factors through the diagonal as,

$$\langle f , g \rangle \;=\; \Delta(X) \cdot (f \times g)$$



3.2.3 Product Types

The product universal construction provides a categorical interpretation of product types,

$$[\![A \times B]\!] \quad := \quad [\![A]\!] \times [\![B]\!]$$

The introduction rule for products,

$$\frac{\Gamma \vdash M : A \qquad \Gamma \vdash N : B}{\Gamma \vdash \langle M , N \rangle : A \times B} \; \times +$$

is interpreted by the tuple construction:

$$[\![\langle M , N \rangle]\!] \quad := \quad \langle [\![M]\!] , [\![N]\!] \rangle$$

and the (negative) elimination rules for products,

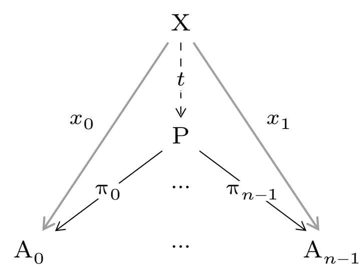
$$\frac{\Gamma \vdash P : A \times B}{\Gamma \vdash \mathrm{fst}(P) : A} \; \times -_0 \qquad \frac{\Gamma \vdash P : A \times B}{\Gamma \vdash \mathrm{snd}(P) : B} \; \times -_1$$

are interpreted by the coordinate projections:

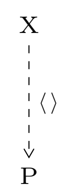
$$[\![\mathrm{fst}(P)]\!] \quad := \quad [\![P]\!] \cdot \pi_0 \qquad \text{and} \qquad [\![\mathrm{snd}(P)]\!] \quad := \quad [\![P]\!] \cdot \pi_1$$

3.2.4 Finite Products

Returning to the theme of unbiased presentations, we would like to define an n-ary product for each $n \in \mathbb{N}$. Let's think about what the universal property of such a construction would be. A product of n factors would consist of an object P, together with a coordinate projection, $\pi_i : P \to A_i$ for each factor such that for any n-ary span $x_i : X \to A_i$ over the same factors there is a unique n-tuple map $t : X \to P$ with $t \cdot \pi_i = x_i$.

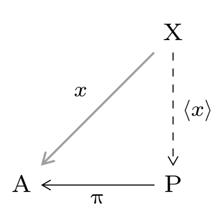


For $n := 0$, a **nullary product** is an object P (requiring no coordinate projections) such that for any object X (requiring no maps to the zero factors) there is a unique null-tuple $\langle\,\rangle : X \to P$ (satisfying no conditions):



But this is just a *terminal object*!

For $n := 1$, a **unary product** of an object A is an object P with a single coordinate projection, $\pi : P \to A$ such that for any arrow $x : X \to A$ there is a unique one-tuple $\langle x \rangle : X \to P$ with $\langle x \rangle \cdot \pi = x$:



A moment's thought confirms that the choice of $P := A$ and $\pi := \mathrm{id}(A)$ (and thus $\langle x \rangle := x$) satisfies this property. So any object is a unary product of itself.

Binary products have already been defined, so we have left to consider products of three or more factors. A ternary product is an object $A \times B \times C$, equipped with three coordinate projection maps such that for any 3-legged span over its factors there is a unique map from the vertex to $A \times B \times C$ commuting with the coordinate projections. But this is the same universal property enjoyed by $(A \times B) \times C$, which has projections $\pi_0 \cdot \pi_0$ to A, $\pi_0 \cdot \pi_1$ to B and π_1 to C. Any span over A, B and C contains a subspan over A and B, so by the universal property of $A \times B$, has a unique map from the vertex to this product, which together with the C leg of the span gives us a unique map from the vertex to $(A \times B) \times C$. The product of four or more factors is analogous.

Of course, there is nothing special about the choice of bracketing:

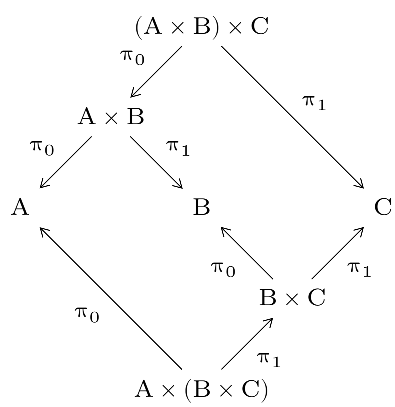
Lemma 3.2.4.1 (product associator) Products are associative, up to isomorphism:

$$A \times (B \times C) \;\cong\; (A \times B) \times C$$

Proof. The maps back and forth,

$$s : A \times (B \times C) \to (A \times B) \times C \quad \text{and} \quad t : (A \times B) \times C \to A \times (B \times C)$$

become clear when we draw the diagram showing how each compound product projects to the three factors, A, B and C:



From this we can simply read off:

$$s \;:=\; \langle\langle \pi_0 \,, \pi_1 \cdot \pi_0 \rangle \,, \pi_1 \cdot \pi_1 \rangle \;\;:\;\; A \times (B \times C) \to (A \times B) \times C$$

$$t \;:=\; \langle \pi_0 \cdot \pi_0 \,, \langle \pi_0 \cdot \pi_1 \,, \pi_1 \rangle\rangle \;\;:\;\; (A \times B) \times C \to A \times (B \times C)$$

And then we check:

$$
\begin{array}{rl}
& s \cdot t \\
= & \text{[definition } t\text{]} \\
& s \cdot \langle \pi_0 \cdot \pi_0 \, , \langle \pi_0 \cdot \pi_1 \, , \pi_1 \rangle\rangle \\
= & \text{[precomposing with a tuple]} \\
& \langle s \cdot \pi_0 \cdot \pi_0 \, , \langle s \cdot \pi_0 \cdot \pi_1 \, , s \cdot \pi_1 \rangle\rangle \\
= & \text{[definition } s\text{]} \\
& \langle \pi_0 \, , \langle \pi_1 \cdot \pi_0 \, , \pi_1 \cdot \pi_1 \rangle\rangle \\
= & \text{[precomposing with a tuple]} \\
& \langle \pi_0 \, , \pi_1 \cdot \langle \pi_0 \, , \pi_1 \rangle\rangle \\
= & \text{[identity expansion for products]} \\
& \langle \pi_0 \, , \pi_1 \cdot \mathrm{id} \rangle \\
= & \text{[composition unit law]} \\
& \langle \pi_0 \, , \pi_1 \rangle \\
= & \text{[identity expansion for products]} \\
& \mathrm{id}
\end{array}
$$

Similarly, $t \cdot s = \mathrm{id}$. □

Up to isomorphism, the cartesian product has the structure of a *monoid*:

Lemma 3.2.4.2 (product unitor) A terminal object is a unit for products, up to isomorphism:

$$\mathrm{A} \times 1 \;\cong\; \mathrm{A} \;\cong\; 1 \times \mathrm{A}$$

Proof. The projection $\pi_0 : \mathrm{A} \times 1 \to \mathrm{A}$ is an isomorphism, with inverse $\langle \mathrm{id}(\mathrm{A}) \, , !(\mathrm{A}) \rangle : \mathrm{A} \to \mathrm{A} \times 1$.

- By the universal property of the product,

$$\langle \mathrm{id}(\mathrm{A}) \, , !(\mathrm{A}) \rangle \cdot \pi_0 \;=\; \mathrm{id}(\mathrm{A}) \;:\; \mathrm{A} \to \mathrm{A}$$

- Going the other way,

$$
\begin{array}{rl}
& \pi_0 \cdot \langle \mathrm{id}(\mathrm{A}) \, , !(\mathrm{A}) \rangle \quad : \quad \mathrm{A} \times 1 \to \mathrm{A} \times 1 \\
= & \text{[pre-composing with a tuple]} \\
& \langle \pi_0 \cdot \mathrm{id}(\mathrm{A}) \, , \pi_0 \cdot !(\mathrm{A}) \rangle \\
= & \text{[composition unit law and pre-composing with a bang]} \\
& \langle \pi_0 \, , !(\mathrm{A} \times 1) \rangle \\
= & \text{[universal property of a terminal object]} \\
& \langle \pi_0 \, , \pi_1 \rangle \\
= & \text{[identity expansion for products]} \\
& \mathrm{id}(\mathrm{A} \times 1)
\end{array}
$$

□

And furthermore, this monoid is commutative:

Lemma 3.2.4.3 (product symmetry) Products are symmetric, up to isomorphism:

$$\mathrm{A} \times \mathrm{B} \;\cong\; \mathrm{B} \times \mathrm{A}$$

Proof. The **swap map**, $\sigma_{\mathrm{A,B}} := \langle \pi_1 , \pi_0 \rangle : \mathrm{A} \times \mathrm{B} \to \mathrm{B} \times \mathrm{A}$ is an isomorphism, with inverse the swap map, $\sigma_{\mathrm{B,A}} := \langle \pi_1 , \pi_0 \rangle : \mathrm{B} \times \mathrm{A} \to \mathrm{A} \times \mathrm{B}$:

$$\begin{array}{cl}
 & \sigma_{\mathrm{A,B}} \cdot \sigma_{\mathrm{B,A}} \\
= & [\text{definition}] \\
 & \langle \pi_1 , \pi_0 \rangle \cdot \langle \pi_1 , \pi_0 \rangle \\
= & [\text{pre-composing with a tuple}] \\
 & \langle \langle \pi_1 , \pi_0 \rangle \cdot \pi_1 , \langle \pi_1 , \pi_0 \rangle \cdot \pi_0 \rangle \\
= & [\text{universal property of a product}] \\
 & \langle \pi_0 , \pi_1 \rangle \\
= & [\text{identity expansion for products}] \\
 & \mathrm{id}(\mathrm{A} \times \mathrm{B})
\end{array}$$

and symmetrically $\sigma_{\mathrm{B,A}} \cdot \sigma_{\mathrm{A,B}} = \mathrm{id}(\mathrm{B} \times \mathrm{A})$ □

To have **finite products** – that is, n-ary products for all $n \in \mathbb{N}$, it suffices to have binary products and a terminal object. A category with all finite products is called a **cartesian category**.

3.2.5 Typing Contexts

Finite products provide the categorical structure needed to interpret (non-dependent, structural) typing contexts. We interpret the *empty context* with a terminal object:

$$[\![\emptyset]\!] \quad := \quad 1$$

And we interpret *context extension* with a cartesian product:

$$[\![\Gamma , x : \mathrm{A}]\!] \quad := \quad [\![\Gamma]\!] \times [\![\mathrm{A}]\!]$$

Or, from the unbiased perspective, for $\Gamma := \overrightarrow{x_i : \mathrm{A}_i}$,

$$[\![\Gamma]\!] \quad := \quad [\![\mathrm{A}_0]\!] \times \cdots \times [\![\mathrm{A}_n]\!]$$

Finite products give us just the right structure needed to implement what are commonly called "structural rules", which are rules that we expect to be admissible in type theories of a certain class, i.e. "structural type theories". The structural rules of weakening, contraction and exchange specify properties of contexts. Because contexts are interpreted as the domains of arrows, we expect their interpretations to behave contravariantly.

The principle of **context weakening** says that a well-typed term remains so in the presence of additional, unused assumptions:

$$\frac{\Gamma \vdash \mathrm{M} : \mathrm{B}}{\Gamma , x : \mathrm{A} \vdash \mathrm{M} : \mathrm{B}} \; cw$$

Its categorical interpretation must be some means of constructing a member of $[\![\Gamma]\!] \times [\![\mathrm{A}]\!] \to [\![\mathrm{B}]\!]$ from a member of $[\![\Gamma]\!] \to [\![\mathrm{B}]\!]$. We can do this by simply pre-composing a projection, or equivalently, up to the unit isomorphism for products, the product of an identity arrow and a *bang map*:

$$[\![\Gamma]\!] \times [\![\mathrm{A}]\!] \xrightarrow{\mathrm{id} \times !} [\![\Gamma]\!] \times 1 \xrightarrow{\cong} [\![\Gamma]\!] \xrightarrow{[\![\mathrm{M}]\!]} [\![\mathrm{B}]\!]$$

The principle of **context contraction** says that a well-typed term depending on two variables of the same type remains so when a single variable is substituted for both:

$$\frac{\Gamma , x : \mathrm{A} , y : \mathrm{A} \vdash \mathrm{M} : \mathrm{B}}{\Gamma , z : \mathrm{A} \vdash \mathrm{M}[(x , y) \mapsto (z , z)] : \mathrm{B}} \; cc$$

Its categorical interpretation must be some means of constructing a member of $[\![\Gamma]\!] \times [\![\mathrm{A}]\!] \to [\![\mathrm{B}]\!]$ from a member of $[\![\Gamma]\!] \times [\![\mathrm{A}]\!] \times [\![\mathrm{A}]\!] \to [\![\mathrm{B}]\!]$. We can do this by simply pre-composing the product of an identity arrow and a *diagonal map*:

$$[\![\Gamma]\!] \times [\![\mathrm{A}]\!] \xrightarrow{\mathrm{id} \times \Delta} [\![\Gamma]\!] \times [\![\mathrm{A}]\!] \times [\![\mathrm{A}]\!] \xrightarrow{[\![\mathrm{M}]\!]} [\![\mathrm{B}]\!]$$

The principle of **context exchange** says that a well-typed term remains so under a permution of its context:

$$\frac{\Gamma , y : \mathrm{B} , x : \mathrm{A} , \Gamma' \vdash \mathrm{M} : \mathrm{C}}{\Gamma , x : \mathrm{A} , y : \mathrm{B} , \Gamma' \vdash \mathrm{M} : \mathrm{C}} \; cx$$

Its categorical interpretation must be some means of constructing a member of $[\![\Gamma]\!] \times [\![\mathrm{A}]\!] \times [\![\mathrm{B}]\!] \times [\![\Gamma']\!] \to [\![\mathrm{C}]\!]$ from a member of $[\![\Gamma]\!] \times [\![\mathrm{B}]\!] \times [\![\mathrm{A}]\!] \times [\![\Gamma']\!] \to [\![\mathrm{C}]\!]$. We can do this by simply pre-composing the product of identity arrows and a *swap map*:

$$[\![\Gamma]\!] \times [\![\mathrm{A}]\!] \times [\![\mathrm{B}]\!] \times [\![\Gamma']\!] \xrightarrow{\mathrm{id} \times \sigma \times \mathrm{id}} [\![\Gamma]\!] \times [\![\mathrm{B}]\!] \times [\![\mathrm{A}]\!] \times [\![\Gamma']\!] \xrightarrow{[\![\mathrm{M}]\!]} [\![\mathrm{C}]\!]$$

So a system that a type theorist might call “structural”, a category theorist would call “cartesian”: it is one in which we may duplicate and discard (as well as reorder) elements of the context. Note that by convention, in type theory weakening and exchange are “silent”, in the sense that we don’t record them in the term itself.

Exercise 3.2.5.1 In structural type theories, the *variable rule* and *substitution rule* of baby type theory have the following genrealizations:

$$\frac{}{\Gamma\,, x : \mathrm{A} \;\vdash\; x : \mathrm{A}}\ \mathit{var} \qquad \text{and} \qquad \frac{\Gamma \;\vdash\; \mathrm{M} : \mathrm{B} \qquad \Gamma\,, y : \mathrm{B} \;\vdash\; \mathrm{N} : \mathrm{C}}{\Gamma \;\vdash\; \mathrm{N}[y{\mapsto}\mathrm{M}] : \mathrm{C}}\ \mathit{sub}$$

Why are these generalizations sound in a setting where contexts are interpreted as finite products?

3.3 Coproducts

A coproduct is the dual construction to a product. Categorically, that is all there is to say about the matter. But because of the asymmetry inherent in type theory – where inferences have a *collection* of assumptions, yet a *single* conclusion – we will have to say a bit more when it comes to our categorical semantics for type theory.

First, we record for convenience, but without further comment, the duals of our main results about products. If you're new to this, it would be an excellent exercise first to go back and see why these are the respective *dual theorems*, and then to prove each one explicitly – that is, by actually going through the argument, rather than by just saying, "by duality, Qed".

3.3.1 Coproducts of Objects

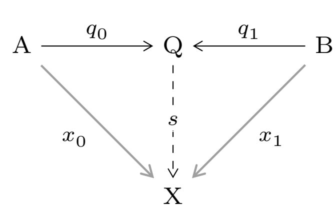
Definition 3.3.1.1 (coproduct of objects) In any category, a **coproduct** of objects A and B is a *cospan* on A and B,

$$\mathrm{A} \xrightarrow{q_0} \mathrm{Q} \xleftarrow{q_1} \mathrm{B}$$

with the property that for any cospan on A and B,

$$\mathrm{A} \xrightarrow{x_0} \mathrm{X} \xleftarrow{x_1} \mathrm{B}$$

there is a unique map $s : \mathrm{Q} \to \mathrm{X}$ such that $q_0 \cdot s = x_0$ and $q_1 \cdot s = x_1$:



We call A and B the **cases** of the coproduct, q_0 and q_1 its **insertions** and s the **cotuple** of x_0 and x_1, and write it as "$[x_0 \, , x_1]$".

Probing a coproduct with itself by choosing $\mathrm{X} \coloneqq \mathrm{Q}$ and $(x_0 \, , x_1) \coloneqq (q_0 \, , q_1)$, we learn:

Lemma 3.3.1.2 (identity expansion for coproducts) If Q is a coproduct of A and B with insertions q_0 and q_1, then $[q_0 \, , q_1] = \mathrm{id}(\mathrm{Q})$.

And being characterized by a universal property, we expect:

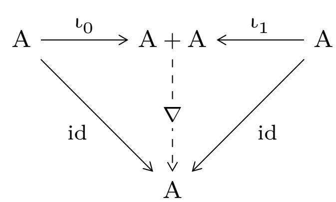
Lemma 3.3.1.3 (uniqueness of coproducts) When they exist, coproducts of objects are unique up to a unique insertion-preserving isomorphism.

We write "A + B" to refer to an arbitrary coproduct of A and B, When the coproduct in question is clear from context, we refer to the two case insertions generically as "ι_0" and "ι_1".

In the category SET, the *disjoint union* of two sets is their coproduct. In CAT, there is something similar: $\mathbb{C} + \mathbb{D}$ is the category whose collection of objects is the disjoint union of those of $\mathbb{C}$ and $\mathbb{D}$ and whose homs between pairs of $\mathbb{C}$-objects is the same as in $\mathbb{C}$, and likewise for $\mathbb{D}$, but where the "mixed" homs are empty.

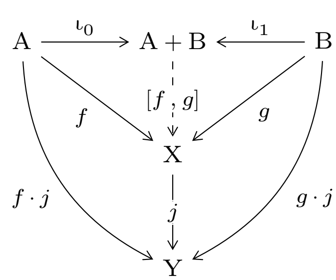
Definition 3.3.1.4 (codiagonal map) For every object A, the universal property of the coproduct gives a canonical **codiagonal map**, which forgets about case distinction:

$$\nabla(\mathrm{A}) \quad \coloneqq \quad [\mathrm{id}(\mathrm{A}) \, , \mathrm{id}(\mathrm{A})] \quad : \quad \mathrm{A} + \mathrm{A} \to \mathrm{A}$$



Lemma 3.3.1.5 (post-composing with a cotuple) For a coproduct A + B, a cotuple $[f \, , g] : \mathrm{A} + \mathrm{B} \to \mathrm{X}$ and an arrow $j : \mathrm{X} \to \mathrm{Y}$,

$$[f \, , g] \cdot j \;=\; [f \cdot j \, , g \cdot j] \;:\; \mathrm{A} + \mathrm{B} \to \mathrm{Y}$$

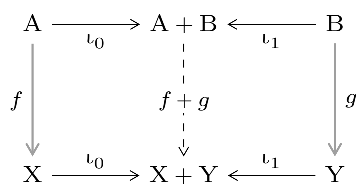


3.3.2 Coproduct Functors

Definition 3.3.2.1 (coproduct of arrows) Given a pair of arrows $f : \mathrm{A} \to \mathrm{X}$ and $g : \mathrm{B} \to \mathrm{Y}$, and coproducts $\mathrm{A}+\mathrm{B}$ and $\mathrm{X}+\mathrm{Y}$, we define the **coproduct of arrows** $f+g : \mathrm{A}+\mathrm{B} \to \mathrm{X}+\mathrm{Y}$ by:

$$\begin{array}{rcl} f+g & : & \mathrm{A}+\mathrm{B} \to \mathrm{X}+\mathrm{Y} \\ f+g & := & [f \cdot \iota_0 \,, g \cdot \iota_1] \end{array}$$

By the universal property of the coproduct $\mathrm{A}+\mathrm{B}$, the arrow $f+g$ is the unique morphism making the two squares commute:



This allows us to characterize the coproduct as a bifunctor:

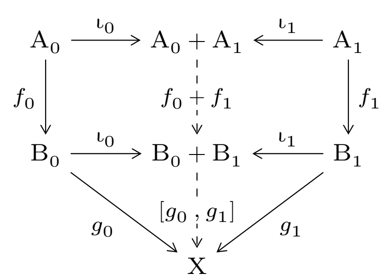
Lemma 3.3.2.2 (functoriality of coproducts) If a category $\mathbb{C}$ has coproducts for each pair of objects, then the given definition of coproducts for arrows yields a functor,

$$- + - \;:\; \mathbb{C} \times \mathbb{C} \to \mathbb{C}$$

called the **coproduct functor**.

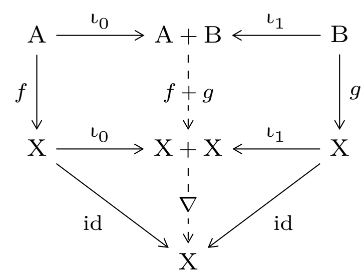
Lemma 3.3.2.3 (pre-composing a coproduct of arrows) For arrows $f_0 + f_1 : \mathrm{A}_0 + \mathrm{A}_1 \to \mathrm{B}_0 + \mathrm{B}_1$ and $[g_0 \,, g_1] : \mathrm{B}_0 + \mathrm{B}_1 \to \mathrm{X}$,

$$(f_0 + f_1) \cdot [g_0 \,, g_1] \;=\; [f_0 \cdot g_0 \,, f_1 \cdot g_1]$$



Corollary 3.3.2.4 (cotuple factorization) A cotuple $[f \,, g] : \mathrm{A}+\mathrm{B} \to \mathrm{X}$ factors through the codiagonal as,

$$[f \,, g] \;=\; (f+g) \cdot \nabla(\mathrm{X})$$



3.3.3 Sum Types

The coproduct universal construction provides a possible categorical interpretation of sum types,

$$[\![\mathrm{A}+\mathrm{B}]\!] \quad := \quad [\![\mathrm{A}]\!] + [\![\mathrm{B}]\!]$$

The introduction rules for sums,

$$\frac{\Gamma \ \vdash\ \mathrm{M} : \mathrm{A}}{\Gamma \ \vdash\ \mathrm{inl}(\mathrm{M}) : \mathrm{A}+\mathrm{B}}\ {+}{+}_0 \qquad \frac{\Gamma \ \vdash\ \mathrm{N} : \mathrm{B}}{\Gamma \ \vdash\ \mathrm{inr}(\mathrm{N}) : \mathrm{A}+\mathrm{B}}\ {+}{+}_1$$

are then interpreted by the case insertions:

$$[\![\mathrm{inl}(\mathrm{M})]\!] \quad := \quad [\![\mathrm{M}]\!] \cdot \iota_0 \qquad \text{and} \qquad [\![\mathrm{inr}(\mathrm{N})]\!] \quad := \quad [\![\mathrm{N}]\!] \cdot \iota_1$$

In the restricted setting where contexts are singletons, the elimination rule for sums,

$$\frac{x : \mathrm{A} \ \vdash\ \mathrm{P} : \mathrm{C} \qquad y : \mathrm{B} \ \vdash\ \mathrm{Q} : \mathrm{C}}{z : \mathrm{A}+\mathrm{B} \ \vdash\ \mathrm{case}^{\dagger}(x \,.\, \mathrm{P} \,;\, y \,.\, \mathrm{Q}) : \mathrm{C}}\ {+}{-}^{\dagger}$$

is then interpreted by the cotuple:

$$[\![\mathrm{case}^{\dagger}(x \,.\, \mathrm{P} \,;\, y \,.\, \mathrm{Q})]\!] := [[\![\mathrm{P}]\!] \,,\, [\![\mathrm{Q}]\!]]$$

3.3.4 Distributive Categories

In order to interpret the full rule for sum elimination, which allows for an arbitrary ambient context:

$$\frac{\Gamma \,,\, x : \mathrm{A} \ \vdash\ \mathrm{P} : \mathrm{C} \qquad \Gamma \,,\, y : \mathrm{B} \ \vdash\ \mathrm{Q} : \mathrm{C}}{\Gamma \,,\, z : \mathrm{A}+\mathrm{B} \ \vdash\ \mathrm{case}(x \,.\, \mathrm{P} \,;\, y \,.\, \mathrm{Q})(z) : \mathrm{C}}\ {+}{-}$$

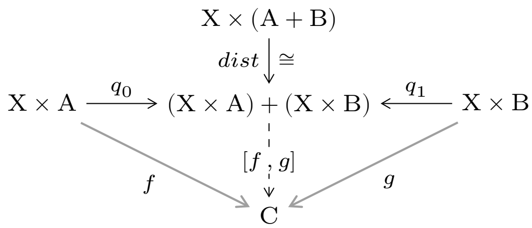
we need a categorical setting in which the **distributive law** holds:

$$dist \quad : \quad \mathrm{X} \times (\mathrm{A}+\mathrm{B}) \ \cong\ (\mathrm{X}\times\mathrm{A}) + (\mathrm{X}\times\mathrm{B})$$

since, in this setting we could define

$$[\![\text{case}]\!] \quad := \quad dist \cdot [- \,, -]$$

or



Likewise, in order to interpret the full rule for void elimination, which also allows for an arbitrary ambient context:

$$\frac{}{\Gamma \,, z : \bot \;\vdash\; \text{abort}(z) : \text{A}} \;\; \bot-$$

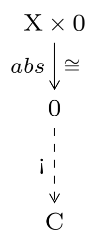
we need a categorical setting in which the **absorption law** holds:

$$abs \quad : \quad \text{X} \times 0 \;\cong\; 0$$

since, in this setting we could define

$$[\![\text{abort}]\!] \quad := \quad abs \cdot ¡$$

or



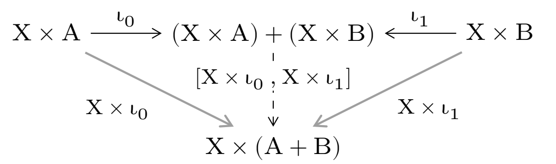
By duality with finite products, an initial object represents a **nullary coproduct**. So we can think of the absorption law as a nullary version on the distributive law in the sense that in the latter the product distributes over the two cases of the coproduct, while in the former it distributes over the zero cases of the initial object.

For the distributive law to hold we merely require the existence of *some* natural[2] isomorphism $\text{X} \times (\text{A} + \text{B}) \to (\text{X} \times \text{A}) + (\text{X} \times \text{B})$. It is a non-trivial theorem[3]

[2] We haven't met naturality yet, but we soon will. It requires that the isomorphism we seek should be "generic" in X, A and B.

[3] `http://arxiv.org/abs/0912.2126`

that the existence of any such natural isomorphism implies the existence of a certain canonical one. First, observe that whatever X, A and B are, there is always a canonical morphism $(X \times A) + (X \times B) \to X \times (A + B)$:



Call this map *undist*. If the distributive law holds then by the cited theorem *undist* is an isomorphism, and its inverse, *dist*, is the canonical distributor.

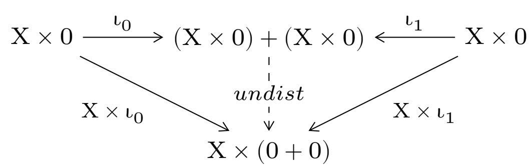
There is always a canonical – indeed unique – arrow $¡ : 0 \to X \times 0$. So if the absorption law is to hold, the absorber, *abs*, must be the inverse of the cobang arrow. Remarkably, the *distributive law* implies the *absorption law*:

Lemma 3.3.4.1 In a category, if products distribute over binary coproducts then they distribute over nullary coproducts.

Proof. We prove that $X \times 0 \cong 0$ by showing that the former has the universal property of an intitial object.

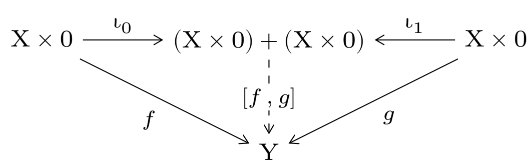
First, for any Y, the hom $X \times 0 \to Y$ is inhabited by $\pi_1 \cdot ¡$. So it remains to show uniqueness.

By the universal property of 0, the two insertions, $\iota_0 , \iota_1 : 0 \to 0 + 0$ are equal. Arrow equality is a congruence under products, so in the coproduct diagram:



we have, $\iota_0 \cdot \mathit{undist} = X \times \iota_0 = X \times \iota_1 = \iota_1 \cdot \mathit{undist}$. Post-composing *dist* shows that the two insertions, $\iota_0 , \iota_1 : X \times 0 \to (X \times 0) + (X \times 0)$ coincide as well.

Thus, for any $f , g : X \times 0 \to Y$ we have $f = \iota_0 \cdot [f , g] = \iota_1 \cdot [f , g] = g$ in:



□

A category with finite products and coproducts in which the distributive law holds is called a **distributive category**. Furthermore, the presence of the next universal construction we meet is sufficient to ensure that a category is distributive.

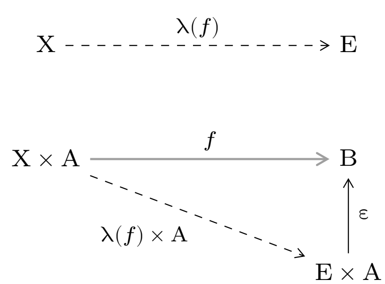
3.4 Exponentials

As functional programmers, we are familiar with the idea of function **currying**, that is, of viewing a function of two arguments as a *higher-order function* that takes the first argument and returns a new function, which, when provided the second argument, computes the same result as the original function when given both arguments at once. Once we get used to working with high-order functions, we wonder how we ever managed to program any other way.

3.4.1 Exponentials of Objects

Exponential objects are the categorical analogue of set-theoretic function space, allowing us to characterize function currying and λ-abstraction.

Definition 3.4.1.1 (exponential object) In a category with binary products, an **exponential** of objects A and B is an object E together with an arrow $\varepsilon : E \times A \to B$ with the property that for any object X and arrow $f : X \times A \to B$ there is a unique arrow $\lambda(f) : X \to E$ such that $\lambda(f) \times A \cdot \varepsilon = f$:



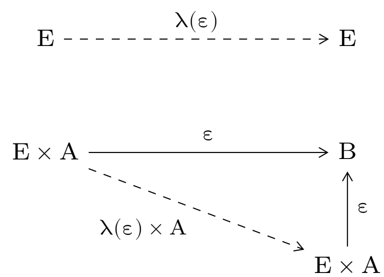
We call ε the **evaluation map** of the exponential, and λ(*f*) the **exponential transpose** or "curry" of *f*.

Notice that the "such that" clause of the definition lets us recover *f* from λ(*f*): just take the product with id(A) and compose with ε. This is just "uncurrying" to functional programmers. If this isn't clear to you, go back to the definitions of *product* and *coproduct* and see how the same principle allows us to recover *f* and *g* from $\langle f , g \rangle$ and from $[f , g]$, respectively.

Let's see what we learn from probing an exponential with itself by choosing $X := E$ and $f := \varepsilon$.

Lemma 3.4.1.2 (identity expansion for exponentials) If E is an exponential of A and B then $\lambda(\varepsilon) = \mathrm{id}(E)$.

Proof. By assumption, $\lambda(\varepsilon)$ is the unique map in the hom set $E \to E$ with the property that $\lambda(\varepsilon) \times A \cdot \varepsilon = \varepsilon$:



By the left unit law of composition, $\mathrm{id}(E \times A) \cdot \varepsilon = \varepsilon$, and by the definition of a product of arrows, $\mathrm{id}(E \times A) = \mathrm{id}(E) \times \mathrm{id}(A)$. Since id(E) has the desired property, the result follows from uniqueness. □

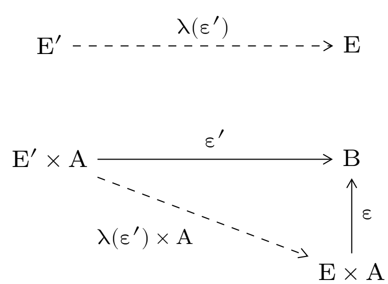
To summarize:

- currying the evaluation map yields the identity on the exponential, and
- uncurrying the identity on the exponential yields the evaluation map.

Because exponentials are structures characterized by a universal property, we expect them to be unique up to a unique structure-preserving isomorphism. This should be familiar by now.

Lemma 3.4.1.3 (uniqueness of expoentials) When they exist, exponentials are unique up to a unique evaluation-preserving isomorphism.

Proof. Suppose that $(E, \varepsilon, \lambda)$ and $(E', \varepsilon', \lambda')$ are both exponentials of A and B. By setting $X := E'$ and $f := \varepsilon'$ in the universal property of E, we have:

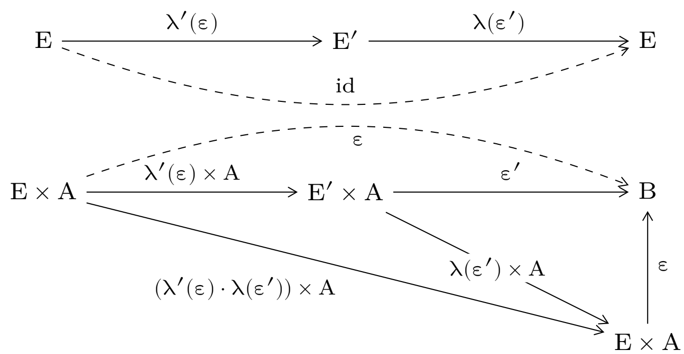


That is, $\lambda(\varepsilon') \times A \cdot \varepsilon = \varepsilon'$. Symmetrically, by the universal property of E', we have $\lambda'(\varepsilon) \times A \cdot \varepsilon' = \varepsilon$.

We want to show that $\lambda'(\varepsilon) \cdot \lambda(\varepsilon') : E \to E$ is the identity map. We do so by uncurrying it:

$$
\begin{aligned}
& (\lambda'(e) \cdot \lambda(\varepsilon')) \times \mathrm{id}(A) \cdot \varepsilon \\
= \; & [\text{product functor}] \\
& \lambda'(e) \times \mathrm{id}(A) \cdot \lambda(\varepsilon') \times \mathrm{id}(A) \cdot \varepsilon \\
= \; & [\text{universal property of E}] \\
& \lambda'(e) \times \mathrm{id}(A) \cdot \varepsilon' \\
= \; & [\text{universal property of E}'] \\
& \varepsilon
\end{aligned}
$$

So by identity expansion for exponentials, $\lambda'(\varepsilon) \cdot \lambda(\varepsilon') = \mathrm{id}(E)$.



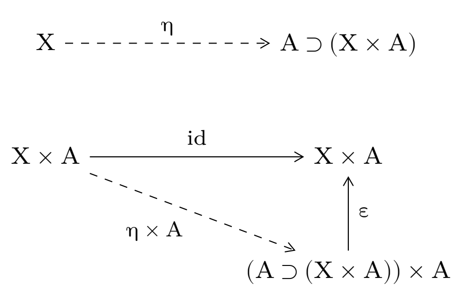
Similarly, we have $\lambda(\varepsilon') \cdot \lambda'(\varepsilon) = \mathrm{id}(E')$.

So $\lambda'(\varepsilon) : E \to E'$ is an isomorphism. By the universal property of E', it is the only one that respects the evaluation ε'. □

Because exponentials are determined as uniquely as is possible by a behavioral characterization, we write "$A \supset B$" to refer to an arbitrary exponential of A and B. The notation "B^A" is also common.

Definition 3.4.1.4 (pairing map) For every object A, the universal property of the exponential gives a canonical **pairing map**, which is a higher-order function that pairs an argument with a given parameter:

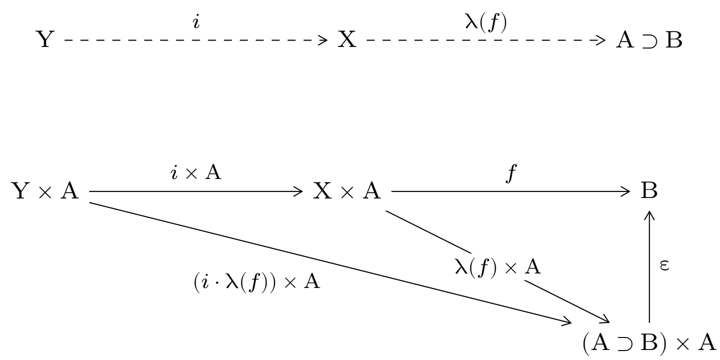
$$\eta(X) \quad := \quad \lambda(\mathrm{id}(X \times A)) \quad : \quad X \to A \supset (X \times A)$$



Exercise 3.4.1.5 (pre-composing with a curry) Use the diagram and the universal property of an exponential object to prove the following:

For an exponential $A \supset B$, an arrow $f : X \times A \to B$ and an arrow $i : Y \to X$,

$$i \cdot \lambda f \;=\; \lambda(i \times A \cdot f) \;:\; Y \to A \supset B$$



Having products and exponents lets a category talk about its own hom collections, indeed exponential objects are sometimes called "internal homs".

Given any arrow $f : A \to B$, we can precompose it with the isomorphism $1 \times A \to A$ from lemma 3.2.4.2 to obtain an arrow $f' : 1 \times A \to B$. We can then curry this arrow to obtain an arrow $\ulcorner f \urcorner := \lambda(f') : 1 \to A \supset B$, yielding a *global element* of the exponential called the **name** of f.

For any object A, we always have the identity arrow id(A), and hence a global element $\ulcorner \mathrm{id}(A) \urcorner : 1 \to A \supset A$, the **internal identity**.

Given any objects A , B , C, we can form the composite:

$$\begin{array}{c}
((A \supset B) \times (B \supset C)) \times A \\
\downarrow \sigma \times A \\
((B \supset C) \times (A \supset B)) \times A \\
\downarrow \alpha \\
(B \supset C) \times ((A \supset B) \times A) \\
\downarrow (B \supset C) \times \varepsilon \\
(B \supset C) \times B) \\
\downarrow \varepsilon \\
C
\end{array}$$

where σ is the product symmetry isomorphism (lemma 3.2.4.3), α is the product associativity isomorphism (lemma 3.2.4.1), and the εs are the respective evaluation maps. Currying this map gives a map $\kappa : (A \supset B) \times (B \supset C) \to A \supset C$, the **internal composition** operation. This has all the properties that we would expect from composition, in particular, for $f : \mathrm{A} \to \mathrm{B}$ and $g : \mathrm{B} \to \mathrm{C}$,

$$\langle \ulcorner f \urcorner , \ulcorner g \urcorner \rangle \cdot \kappa \quad = \quad \ulcorner f \cdot g \urcorner$$

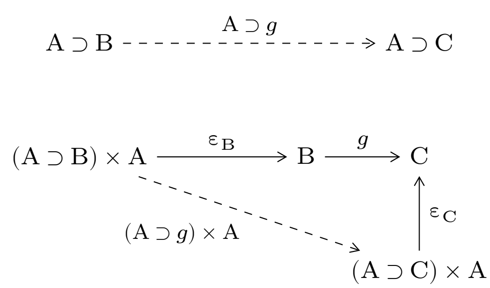
3.4.2 Exponential Functors

We can use the universal property of exponential objects to define a covariant exponential functor.

Definition 3.4.2.1 (covariant exponential of an arrow) For a fixed object A, we define the **exponential of an arrow** $g : \mathrm{B} \to \mathrm{C}$ to be:

$$\mathrm{A} \supset g \quad := \quad \lambda(\varepsilon_{\mathrm{B}} \cdot g) \quad : \quad \mathrm{A} \supset \mathrm{B} \to \mathrm{A} \supset \mathrm{C}$$

(where we subscript the evaluation maps to match their exponentials). By the universal property of exponentials, $\mathrm{A} \supset g$ is the unique arrow making the triangle commute:



If this definition seems rather unmotivated, it may help to keep in mind that the idea behind an exponential of an arrow, $\mathrm{A} \supset g$, is to somehow "post-compose g to the B in $\mathrm{A} \supset \mathrm{B}$". This will make more sense shortly, when we will be in a position to see that this definition makes exponential evaluation maps into a *natural transformation*.

Lemma 3.4.2.2 (functoriality of exponentials) In a category $\mathbb{C}$ with finite products and a fixed object A, the given definition of exponential of arrows yields a functor,

$$\mathrm{A} \supset - \;:\; \mathbb{C} \to \mathbb{C}$$

Proof. In order to prove that $\mathrm{A} \supset -$ is a functor, we must show that it preserves the composition structure.

nullary composition: We must show that

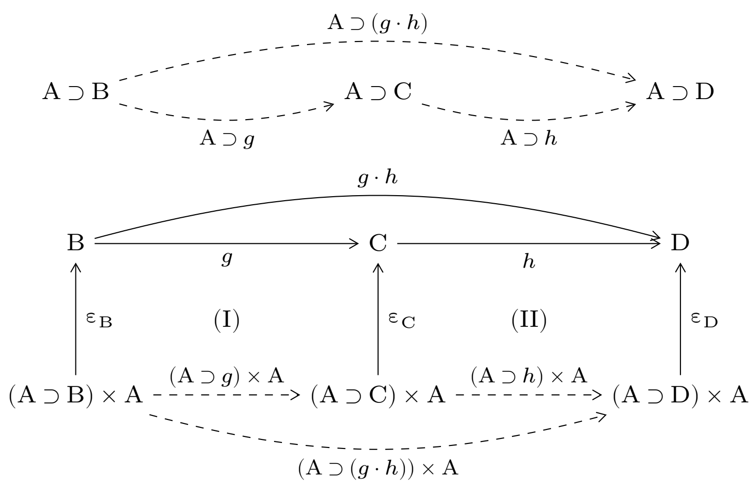
$$\mathrm{A} \supset \mathrm{id}(\mathrm{B}) \;=\; \mathrm{id}(\mathrm{A} \supset \mathrm{B})$$

$$
\begin{array}{ll}
& A \supset \mathrm{id}(B) \\
= & [\text{definition of } A \supset - \text{ on arrows}] \\
& \lambda(\varepsilon_B \cdot \mathrm{id}(B)) \\
= & [\text{composition unit law}] \\
& \lambda(\varepsilon_B) \\
= & [\text{identity expansion for exponentials}] \\
& \mathrm{id}(A \supset B)
\end{array}
$$

binary composition: We must show that

$$A \supset (g \cdot h) \;=\; A \supset g \cdot A \supset h$$

In the diagram,



the outer square commutes by the definition of $A \supset (g \cdot h)$ and the inner squares (I) and (II) commute by the definitions of $A \supset g$ and $A \supset h$, respectively. By the functoriality of the cartesian product,

$$(A \supset g) \times A \cdot (A \supset h) \times A \quad = \quad (A \supset g \cdot A \supset h) \times A$$

By pasting squares (I) and (II), we see that:

$$\varepsilon_B \cdot (g \cdot h) \quad = \quad (A \supset g) \times A \cdot (A \supset h) \times A \cdot \varepsilon_D$$

The result follows by the uniqueness clause of the universal property of exponentials.

□

3.4.3 Function Types

The exponential universal construction provides the categorical interpretation of function types,

$$[\![A \to B]\!] \quad := \quad [\![A]\!] \supset [\![B]\!]$$

The introduction rule,

$$\frac{\Gamma\,,\, x : A \;\vdash\; M : B}{\Gamma \;\vdash\; \lambda x\,.\, M : A \to B} \;\to +$$

is interpreted by currying:

$$[\![\lambda x\,.\, M]\!] \quad := \quad \lambda([\![M]\!])$$

and the elimination rule,

$$\frac{\Gamma \;\vdash\; P : A \supset B \qquad \Gamma \;\vdash\; N : A}{\Gamma \;\vdash\; P\ N : B} \;\to -$$

is interpreted by the evaluation map:

$$[\![P\ N]\!] \quad := \quad \langle [\![P]\!]\,,\, [\![N]\!] \rangle \cdot \varepsilon$$

3.5 Cartesian Closed Categories

A category having all finite products (i.e. a terminal object and binary products), as well as all exponentials, is known as a **cartesian closed category**. A category that additionally has all finite coproducts (i.e. an initial object and binary coproducts) is called **bicartesian closed**.

It turns out that every bicartesian closed category is *distributive*. We don't have time to give a full explanation of why this is the case, but the short version is that it can be seen using the **Yoneda principle**, which says essentially that for given objects $X\,,\,Y : \mathbb{C}$, if for all objects Z, we have $\mathbb{C}\,(X \to Z) \cong \mathbb{C}\,(Y \to Z)$ in Set then $X \cong Y$ in $\mathbb{C}$. You should think of this as generalizing the following familiar fact about preordered sets:

$$(\forall Z\,.\, X \leq Z \Longleftrightarrow Y \leq Z) \qquad \text{implies} \qquad X \quad \cong \quad Y$$

Using this principle, the *distributive law* follow from:

$$
\begin{array}{ll}
 & \mathbb{C}\,(\mathrm{X} \times (\mathrm{A} + \mathrm{B}) \to \mathrm{Z}) \\
\cong & [\text{product symmetry}] \\
 & \mathbb{C}\,((\mathrm{A} + \mathrm{B}) \times \mathrm{X} \to \mathrm{Z}) \\
\cong & [\text{currying}] \\
 & \mathbb{C}\,(\mathrm{A} + \mathrm{B} \to \mathrm{X} \supset \mathrm{Z}) \\
\cong & [\text{uncotupling}] \\
 & \mathbb{C}\,(\mathrm{A} \to \mathrm{X} \supset \mathrm{Z}) \times \mathbb{C}\,(\mathrm{B} \to \mathrm{X} \supset \mathrm{Z}) \\
\cong & [\text{uncurrying}] \\
 & \mathbb{C}\,(\mathrm{A} \times \mathrm{X} \to \mathrm{Z}) \times \mathbb{C}\,(\mathrm{B} \times \mathrm{X} \to \mathrm{Z}) \\
\cong & [\text{product symmetry}] \\
 & \mathbb{C}\,(\mathrm{X} \times \mathrm{A} \to \mathrm{Z}) \times \mathbb{C}\,(\mathrm{X} \times \mathrm{B} \to \mathrm{Z}) \\
\cong & [\text{cotupling}] \\
 & \mathbb{C}\,((\mathrm{X} \times \mathrm{A}) + (\mathrm{X} \times \mathrm{B}) \to \mathrm{Z})
\end{array}
$$

Chapter 4

Two Dimensional Structure

We have deliberately presented the universal constructions of a bicartesian closed category in such a way as to highlight the parallels between them. The reader may be wondering whether there is some more general construction lurking in the wings, in terms of which terminal and initial objects, products, coproducts and exponentials can all be described. This is indeed the case: the mystery construction is called an "adjunction". But in order to describe it we will need to first understand some 2-dimensional category theory.

4.1 Naturality

Naturality is the carrier of the two-dimensional structure of categories of categories. It imposes a sort of *comprehension* or *uniformity* principle that allows us to interpret a family of features within a particular category as a single feature in the ambient category of categories.

4.1.1 Natural Transformations

Definition 4.1.1.1 (natural transformation) For parallel functors, $\mathrm{F}\,,\mathrm{G} : \mathbb{C} \to \mathbb{D}$, a **natural transformation** φ from F to G is a functor $\varphi : \mathbb{C} \to \mathbb{D}^{\to}$ (where $\mathbb{D}^{\to}$ is the *arrow category* of $\mathbb{D}$) such that:

$$\varphi \cdot \mathit{dom} = \mathrm{F} \qquad \text{and} \qquad \varphi \cdot \mathit{cod} = \mathrm{G}$$

Explicitly, this means that,

- for each object $\mathrm{A} : \mathbb{C}$ there is an arrow,

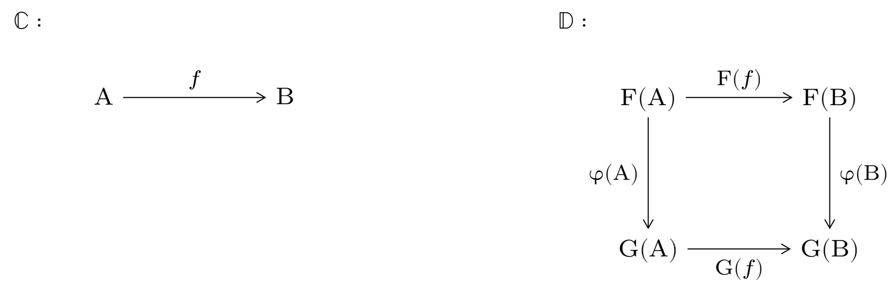
$$\varphi(\mathrm{A}) : \mathbb{D}\,(\mathrm{F}(\mathrm{A}) \to \mathrm{G}(\mathrm{A}))$$

called the **component** of φ at A, and

- for each arrow $f : \mathbb{C}\,(\mathrm{A} \to \mathrm{B})$ there is a commuting square,

$$\mathrm{F}(f) \cdot \varphi(\mathrm{B}) \;=\; \varphi(\mathrm{A}) \cdot \mathrm{G}(f)$$

called the **naturality square** for φ at f:



One way to think about this is that a functor "projects" an image of its domain category into its codomain category. In this sense, a functor acts as a lens, which may "distort" the structure of the source category by identifying distinct objects or arrows. Under this interpretation, a component of a natural transformation acts as a "homotopy" between the images of an object cast by two parallel functors, and a naturality square ensures that these object homotopies are consistent with the images of arrows.

You may also think of the naturality square above as a "2-dimensional arrow" from the 1-dimensional arrow F(f) to the 1-dimensional arrow G(f), acting as the "component" of φ at f. But in an ordinary (1-dimensional) category, the only "2-dimensional arrows" available are the trivial ones – i.e. equalities.

4.1.2 Functor Categories

Natural transformations provide a notion of arrows between parallel functors, turning a hom *set* into a hom *category*:

Definition 4.1.2.1 (functor category) For categories ℂ and 𝔻, define the **functor category**, "Fun(ℂ , 𝔻)", to have the following structure:

objects $\mathrm{Fun}(\mathbb{C}, \mathbb{D})_0 \;\coloneqq\;$ functors from ℂ to 𝔻

arrows $\mathrm{Fun}(\mathbb{C}, \mathbb{D})\,(\mathrm{F} \to \mathrm{G}) \;\coloneqq\;$ natural transformations from F to G

identities $\mathrm{id}(\mathrm{F})(\mathrm{A}) \;\coloneqq\; \mathrm{id}(\mathrm{F}(\mathrm{A}))$
(A component of an *identity natural transformation* is an identity arrow.)

composition $(\varphi \cdot \psi)(\mathrm{A}) \;\coloneqq\; \varphi(\mathrm{A}) \cdot \psi(\mathrm{A})$
(A component of a *composite natural transformation* is the composition of the constituent components.)

Composition of natural transformations in the functor category $\textsc{Fun}(\mathbb{C}, \mathbb{D})$ is associative and unital just because composition of morphisms in $\mathbb{D}$ is.

An important class of natural transformations is the natural isomorphisms. A **natural isomorphism** is simply an *isomorphism* in a functor category. Unpacking this a bit, for parallel functors, $F, G : \mathbb{C} \to \mathbb{D}$, a natural transformation $\varphi : \textsc{Fun}(\mathbb{C}, \mathbb{D})\,(F \to G)$ is a natural isomorphism just in case it has an *inverse* $\varphi^{-1} : \textsc{Fun}(\mathbb{C}, \mathbb{D})\,(G \to F)$.

Exercise 4.1.2.2 Show that a natural transformation is a (natural) isomorphism in the functor category $\textsc{Fun}(\mathbb{C}, \mathbb{D})$ just in case each of its components is an isomorphism in $\mathbb{D}$.

An *exponential* object $A \supset B$ in a category $\mathbb{C}$ is an object representing the collection of morphisms $\mathbb{C}\,(A \to B)$. If $\mathbb{C}$ is a category of categories, then an object in $\mathbb{C}$ is a category and the morphisms between any two such are functors. So it is natural to wonder whether functor categories are exponential objects in categories of categories. This is generally the case when such exponential categories exist. In particular, it is true in the category of small categories:

Fact 4.1.2.3 (exponential categories) The category $\textsc{Cat}$ has functor categories as exponential objects. That is, for $\mathbb{A}, \mathbb{B} : \textsc{Cat}$,

$$\mathbb{A} \supset \mathbb{B} \quad = \quad \textsc{Fun}(\mathbb{A}, \mathbb{B})$$

From now on we will use exponential notation for functor categories.

4.2 2-Categories

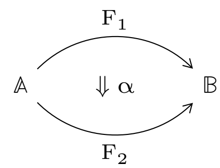
4.2.1 2-Dimensional Categorical Structure

Something subtle and profound has just happened, so let's go through it carefully. Recall that when we introduced categories, we gave them structure at two different dimensions:

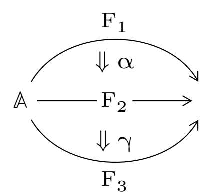
- at dimension 0, we have "points", in the form of objects,
- and at dimension 1, we have "lines", in the form of arrows.

But in introducing natural transformations, we just said that for any two fixed objects, we have a whole *category* of functors and natural transformations between them, so the hom collections in $\textsc{Cat}$ are not 0-dimensional sets, but rather 1-dimensional categories. This gives the category $\textsc{Cat}$ structure at dimension 2 as well!

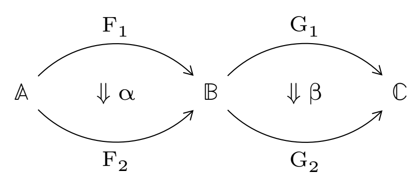
For fixed categories $\mathbb{A}$ and $\mathbb{B}$, and parallel functors $F_1, F_2 : \mathbb{A} \to \mathbb{B}$, we can draw a natural transformation $\alpha : F_1 \to F_2$ as a directed "surface" in a 2-dimensional diagram in Cat:



And if we have another parallel functor $F_3 : \mathbb{A} \to \mathbb{B}$ and natural transformation $\gamma : F_2 \to F_3$, we can paste the diagrams together along F_2 and draw the composite natural transformation $\alpha \cdot \gamma : F_1 \to F_3$ as:



Now we come to the question of what happens if we don't require the pair of objects under consideration to remain fixed. Consider the following pair of "adjacent" natural transformations:



We have the parallel composite functors,

$$(F_1 \cdot G_1)\,,(F_2 \cdot G_2) : \mathbb{A} \to \mathbb{C}$$

Is there some way to form a composite natural transformation, $\alpha \cdot\cdot \beta : F_1 \cdot G_1 \to F_2 \cdot G_2$?[1] Well, given an object $A : \mathbb{A}$, we know that the component of such a composite natural transformation at A would be an arrow,

$$(\alpha \cdot\cdot \beta)(A) \;\;:\;\; \mathbb{C}\,((G_1 \circ F_1)(A) \to (G_2 \circ F_2)(A))$$

Here is how we can define it:

Lemma 4.2.1.1 (horizontal composition of natural transformations) In the situation just described, there is a natural transformation

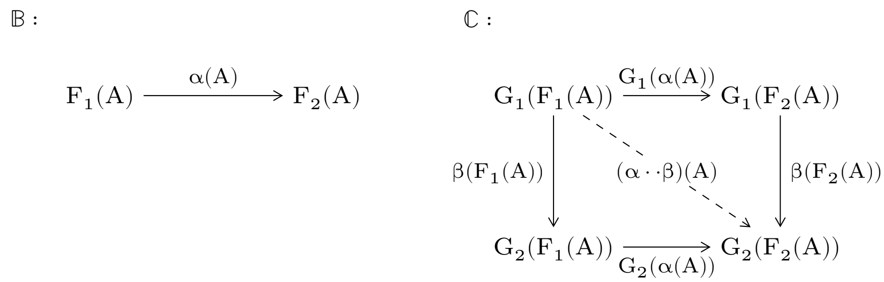
$$\alpha \cdot\cdot \beta \;\;:\;\; F_1 \cdot G_1 \to F_2 \cdot G_2$$

[1] Read "$\alpha \cdot\cdot \beta$" as "α beside β" – that is, composed along their common boundary two dimensions down, rather than the usual one dimension down with "– · –". The reader may amuse herself thinking about how to extend this pattern to still higher dimensions.

called the **horizontal composition** of α and β, with components,

$$(\alpha\cdot\cdot\beta)(A) \ := \ G_1(\alpha(A))\cdot\beta(F_2(A)) \ = \ \beta(F_1(A))\cdot G_2(\alpha(A))$$

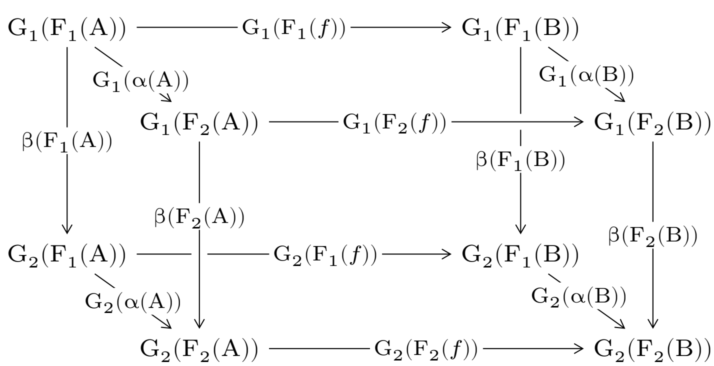
Proof. To see that the two composites are indeed equal consider the component of α at A in $\mathbb{B}$. This determines a naturality square for β at α(A) in $\mathbb{C}$:



(4.1)

establishing that the two expressions for the putative components of α · ·β are equal.

Now we must show that this definition of components respects arrows in $\mathbb{A}$. Let $f : \mathbb{A}(A \to B)$. Then in the diagram in $\mathbb{C}$,



- the left and right squares are the naturality squares for β at α(A) and α(B),
- the top and bottom squares are the G_1 and G_2 functor-images of the naturality squares for α at f,
- and the back and front squares are the naturality squares for β at $F_1(f)$ and $F_2(f)$.

Pasting the top and front – or equivalently, the back and bottom – squares establishes the naturality of α · ·β at f. □

Remark 4.2.1.2 (connection to arrow categories) The construction in the preceding proof should remind you of the double arrow category construction. This is because the *arrow category* $\mathbb{C}^{\rightarrow}$ is equivalent to the functor category $\mathbb{I} \supset \mathbb{C}$, where $\mathbb{I}$ is the *interval category*.

Exercise 4.2.1.3 Check that horizontal composition respects the composition structure of each functor category by pasting squares onto diagram 4.1.

Mercifully, we do not need to think about natural transformations in this cumbersome, component-wise manner. After all, the whole point of the categorical approach is to allow us to reason behaviorally rather than structurally. The component-wise definition of natural transformations is reminiscent of Plato's *Allegory of the Cave*: the component arrows and naturality squares in the codomain category are mere one-dimensional shadows cast by the flesh-and-blood natural transformations, which are two-dimensional morphisms between parallel one-dimensional morphisms living in a 2-dimensional category of categories. Let us turn our heads and stumble into the light.

Definition 4.2.1.4 (2-category) A strict 2-dimensional globular category, or **2-category**, $\mathbb{C}$ consists of the following data:

- A collection of 0**-cells** of $\mathbb{C}$.

- For each $A , B : \mathbb{C}$, a **hom category** $\mathbb{C}\,(A \rightarrow B)$, whose objects are 1**-cells** of $\mathbb{C}$ and whose arrows are 2**-cells** of $\mathbb{C}$.

  We write the composition of these arrows as "$-\cdot-$" and call it the "vertical composition of 2-cells".

- For each $A , B , C : \mathbb{C}$, a composition bifunctor:

$$\kappa_{A,B,C} : \mathbb{C}\,(A \rightarrow B) \times \mathbb{C}\,(B \rightarrow C) \rightarrow \mathbb{C}\,(A \rightarrow C)$$

  We write its action on objects as "$- \cdot -$" and call it the "composition of 1-cells". We write its action on arrows as "$-\cdot\cdot-$" and call it the "horizontal composition of 2-cells".

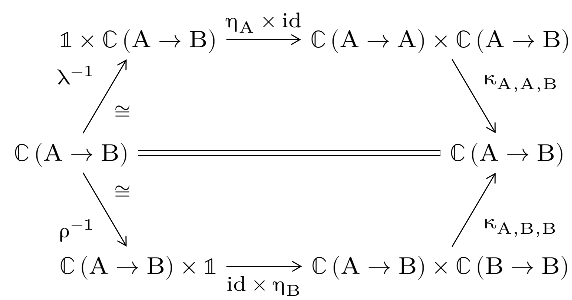
- For each $A : \mathbb{C}$, an identity cell functor:

$$\eta_A : \mathbb{1} \rightarrow C\,(A \rightarrow A)$$

  We write the image of $\star$ as "id(A)" and call it the "identity 1-cell". We write the image if id($\star$) as "$\mathrm{id}^2(A)$" and call it the "double-identity 2-cell".

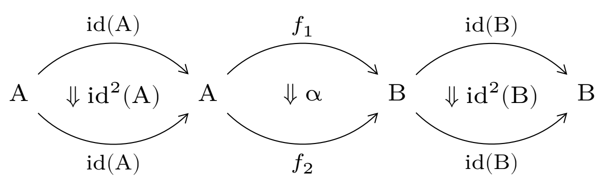
This data is required to respect the following relations:

composition unitality



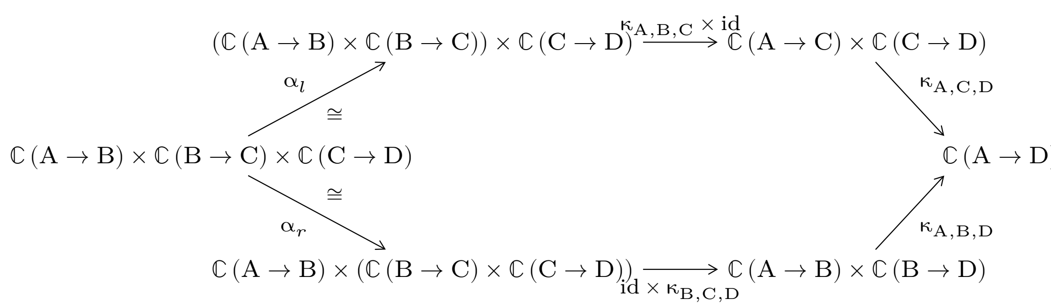
where λ and ρ are the product unit isomorphisms.

Explicitly, this means that in the situation:



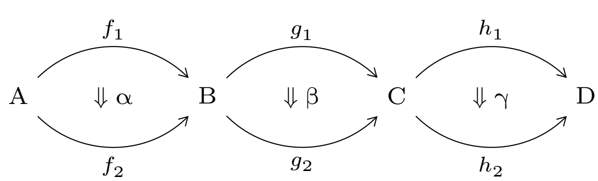
$$\text{id}(A) \cdot f_i = f_i = f_i \cdot \text{id}(B) \qquad \text{and} \qquad \text{id}^2(A) \cdot\cdot \alpha = \alpha = \alpha \cdot\cdot \text{id}^2(B)$$

composition associativity



where the αs are the product associativity isomorphisms.

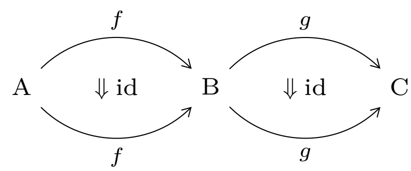
Explicitly, this means that in the situation:



$$(f_i \cdot g_i) \cdot h_i = f_i \cdot (g_i \cdot h_i) \qquad \text{and} \qquad (\alpha \cdot\cdot \beta) \cdot\cdot \gamma = \alpha \cdot\cdot (\beta \cdot\cdot \gamma)$$

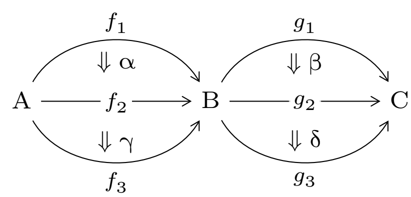
The definition may seem daunting, but the intuition is straightforward. We want to be able to compose 1-cells along their common 0-cell boundaries and 2-cells along both their common 1-cell boundaries and along their common 0-cell boundaries. Furthermore, we want every composable configuration of 2-cells to have a unique composite. The last goal is accomplished by the composition unitality and associativity laws, together with the requirement that composition be a bifunctor, which implies the following:

nullary composition In the situation,



$$\mathrm{id}(f) \cdot \cdot \mathrm{id}(g) \quad = \quad \mathrm{id}(f \cdot g)$$

binary composition In the situation,



$$(\alpha \cdot \gamma) \cdot \cdot (\beta \cdot \cdot \delta) \quad = \quad (\alpha \cdot \cdot \beta) \cdot (\gamma \cdot \cdot \delta)$$

This relation is known as the **interchange law**.

The category Cat with its categories, functors and natural transformations, is a 2-category. Its hom categories are the *functor categories*, and its horizontal composition of 2-cells is the *horizontal composition* of natural transformations.

4.2.2 String Diagrams

If we take the Poincaré (or graph) dual of the 2-dimensional diagrams we have been drawing, we obtain a very useful graphical language for 2-categories, called **string diagrams**. Specifically, we will use "planar progressive" string diagrams to represent configurations of 0-, 1-, and 2-cells.

In the graphical language of string diagrams,

0-**cells** are represented by regions in the plane,

1-**cells** are represented by lines or "strings" or "wires", (in our convention) progressing from top to bottom,

2-cells are represented by points, fattened up into "nodes" or "beads" with the wires representing their domain 1-cells entering from above, and those representing their codomain 1-cells exiting from below.

identity 1-cells are represented by a dashed wire, or usually, not at all,

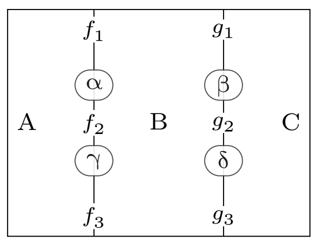
1-cell composition is represented by juxtaposing wires side-by-side, separated by the region representing their common boundary,

identity 2-cells are represented by a dashed node, or usually, not at all,

2-cell vertical composition is represented by wiring the output sockets of the first node to the input sockets of the second,

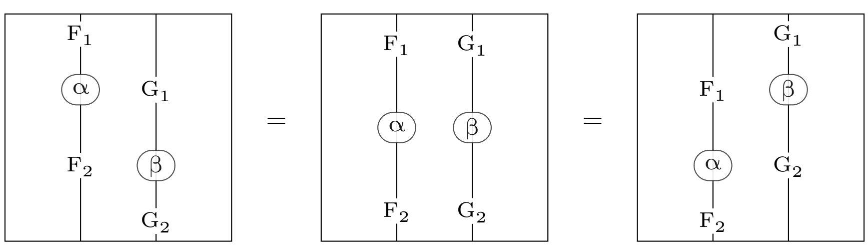
2-cell horizontal composition is represented by juxtaposing nodes, together with their respective wires, side-by-side,

As an example, the pasting diagram for the interchange law becomes the string diagram:



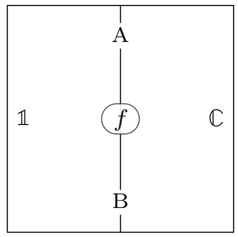
Typically, we omit labeling the regions as their identities can always be inferred.

The commuting of diagram 4.1 in the definition of horizontal composition of natural transformations represents the notion of **naturality as independence**, depicted in string diagrams by the property that two beads without an output-to-input connection between them may freely "slide" past one another along their wires, and it makes no difference which is above or below the other.



The idea of independence is the heart of naturality; the rest of its properties can be recovered from this.

There is a handy "trick" for transforming diagrams within a category into string diagrams using *global elements*. Notice that for any object $A : \mathbb{C}$ there is a functor (which we typically overload with the same name) $A : \mathbb{1} \to \mathbb{C}$ picking out that object, and for any arrow $f : \mathbb{C}\,(\mathrm{A} \to \mathrm{B})$ there is a natural transformation between the respective functors, which we can represent as the string diagram:



Exercise 4.2.2.1 Show how the commuting of *naturality squares* is a consequence of naturality as independence.

4.3 Adjunctions

Adjunctions are constructions that may exist in the context of a 2-dimensional category. In any 2-category adjunctions have a behavioral or "external" characterization. In the 2-category Cat they also have structural or "internal" characterizations.

4.3.1 Behavioral Characterization

For our primary definition of adjunction we adopt the following behavioral one, which we call the "zigzag characterization":

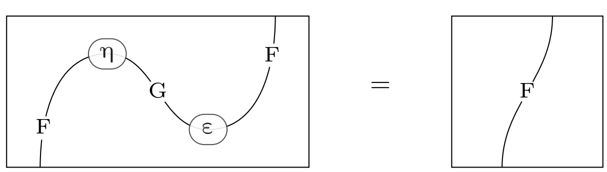
Definition 4.3.1.1 (adjunction – zigzag characterization) Anti-parallel functors $\mathrm{F} : \mathbb{C} \to \mathbb{D}$ and $\mathrm{G} : \mathbb{D} \to \mathbb{C}$ form an **adjunction**, written "$\mathrm{F} \dashv \mathrm{G}$", if there exist natural transformations:

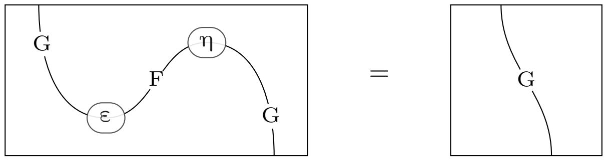
- $\eta \;:\; \mathrm{id}(\mathbb{C}) \to \mathrm{F} \cdot \mathrm{G},$ called the adjunction's **unit**, and
- $\varepsilon \;:\; \mathrm{G} \cdot \mathrm{F} \to \mathrm{id}(\mathbb{D}),$ called the adjunction's **counit**,

satisfying the relations:

- **left zigzag law**: $(\eta \cdot \cdot \mathrm{F}) \cdot (\mathrm{F} \cdot \cdot \varepsilon) \;=\; \mathrm{id}(\mathrm{F})$
- **right zigzag law**: $(\mathrm{G} \cdot \cdot \eta) \cdot (\varepsilon \cdot \cdot \mathrm{G}) \;=\; \mathrm{id}(\mathrm{G})$

The reason for the name "zigzag" becomes apparent when the laws are drawn as string diagrams:





The chirality of the zigzag laws comes from the fact that when $\mathrm{F} \dashv \mathrm{G}$, F is called **left adjoint** to G, and G is called **right adjoint** to F.

4.3.2 Structural Characterizations

Instead of thinking of an adjunction as a single structure living in the 2-category Cat, we can think of it as a correlation between families of structures living in two particular categories. This is like the component-wise presentation of a *natural transformation*. One such characterization of an adjunction is the following:

Definition 4.3.2.1 (adjunction – natural bijection of hom sets charaterization) Anti-parallel functors $\mathrm{F} : \mathbb{C} \to \mathbb{D}$ and $\mathrm{G} : \mathbb{D} \to \mathbb{C}$ form an adjunction $\mathrm{F} \dashv \mathrm{G}$ if for any $\mathrm{A} : \mathbb{C}$ and $\mathrm{B} : \mathbb{D}$ there is a natural bijection of hom sets:

$$\frac{\mathbb{C}\left(\mathrm{A} \to \mathrm{G}(\mathrm{B})\right)}{\mathbb{D}\left(\mathrm{F}(\mathrm{A}) \to \mathrm{B}\right)} \ \theta$$

This characterization is internal or structural because, unlike the zigzag characterization, we look inside the categories $\mathbb{C}$ and $\mathbb{D}$. We will call the downward direction of such a bijection "$-^\sharp$" and the upward direction "$-^\flat$", so:

$$\begin{array}{ccc} f : \mathbb{C}\left(\mathrm{A} \to \mathrm{G}(\mathrm{B})\right) & \overset{-^\sharp}{\longmapsto} & f^\sharp : \mathbb{D}\left(\mathrm{F}(\mathrm{A}) \to \mathrm{B}\right) \\ & \text{and} & \\ g : \mathbb{D}\left(\mathrm{F}(\mathrm{A}) \to \mathrm{B}\right) & \overset{-^\flat}{\longmapsto} & g^\flat : \mathbb{C}\left(\mathrm{A} \to \mathrm{G}(\mathrm{B})\right) \\ & \text{and} & \\ (f^\sharp)^\flat = f & \text{and} & (g^\flat)^\sharp = g \end{array}$$

We call the image of an arrow under this bijection its **adjoint complement**.

A bijection of hom sets is *natural* if it extends along its boundary by the relevant functors. In this case, that means that for any $a : \mathbb{C}\left(\mathrm{A}' \to \mathrm{A}\right)$ and $b : \mathbb{D}\left(\mathrm{B} \to \mathrm{B}'\right)$ we have,

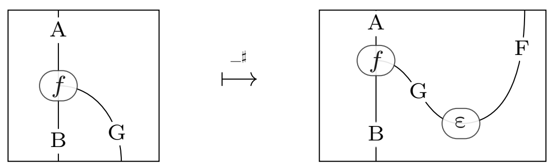
$$\begin{array}{ll} \mathbb{C}: & \mathrm{A}' \xrightarrow{\;a\;} \mathrm{A} \xrightarrow{\;f = g^\flat\;} \mathrm{G}(\mathrm{B}) \xrightarrow{\;\mathrm{G}(b)\;} \mathrm{G}(\mathrm{B}') \\ \hline\hline \mathbb{D}: & \mathrm{F}(\mathrm{A}') \xrightarrow[\mathrm{F}(a)]{} \mathrm{F}(\mathrm{A}) \xrightarrow[g = f^\sharp]{} \mathrm{B} \xrightarrow[b]{} \mathrm{B}' \end{array}$$

Technically, this is a *natural isomorphism* in the functor category $(\mathbb{C}^\circ \times \mathbb{D}) \supset \textsc{Set}$ between $\mathbb{C}\,(\overset{1}{-} \to \mathrm{G}(\overset{2}{-}))$ and $\mathbb{D}\,(\mathrm{F}(\overset{1}{-}) \to \overset{2}{-})$.

We won't prove the equivalence of the various characterizations of adjunctions in this course, but the fact that the zigzag characterization implies the natural bijection of hom sets characterization is easy to see using string diagrams.

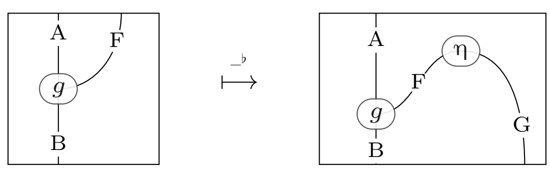
Given an arrow $f : \mathbb{C}\,(\mathrm{A} \to \mathrm{G}(\mathrm{B}))$, the obvious way to construct an arrow $f^\sharp : \mathbb{D}\,(\mathrm{F}(\mathrm{A}) \to \mathrm{B})$ out of the parts at hand is by defining,

$$f^\sharp \quad := \quad \mathrm{F}(f) \cdot \varepsilon(\mathrm{B})$$



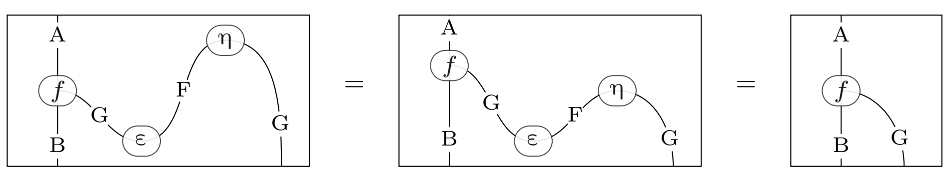
Likewise, given an arrow $g : \mathbb{D}\,(\mathrm{F}(\mathrm{A}) \to \mathrm{B})$, the obvious way to construct an arrow $g^\flat : \mathbb{C}\,(\mathrm{A} \to \mathrm{G}(\mathrm{B}))$ is by defining,

$$g^\flat \quad := \quad \eta(\mathrm{A}) \cdot \mathrm{G}(g)$$

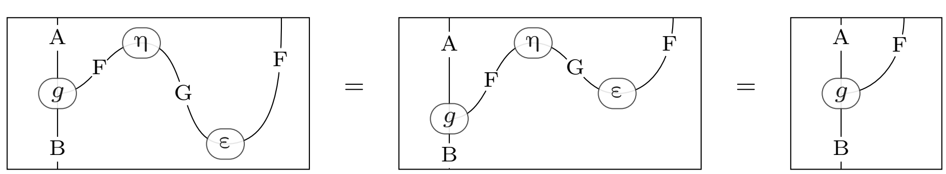


We can use the zigzag laws to show that $-^\sharp$ and $-^\flat$ are inverse operations:

$(f^\sharp)^\flat \quad =$

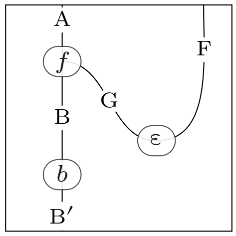


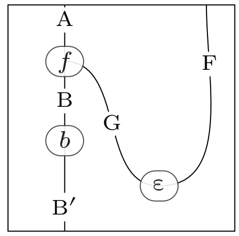
$(g^\flat)^\sharp \quad =$



The naturality of the bijection in the domain and codomain coordinates is also obvious in the graphical language.

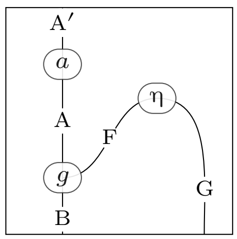
- For naturality in the codomain coordinate, given $b : \mathbb{D}\,(\mathrm{B} \to \mathrm{B}')$ we have:

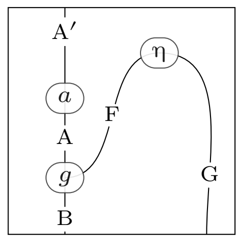
$f^\sharp \cdot b$ =


=


= $(f \cdot \mathrm{G}(b))^\sharp$

- For naturality in the domain coordinate, given $a : \mathbb{C}\,(\mathrm{A}' \to a)$ we have:

$a \cdot g^\flat$ =


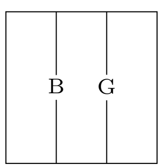
=


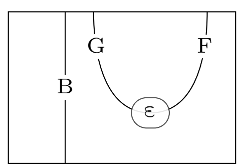
= $(\mathrm{F}(a) \cdot g)^\flat$

Exercise 4.3.2.2 With f, g, a and b as in the proof above, verify the following:

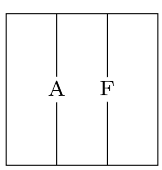
- $(g \cdot b)^\flat \quad = \quad g^\flat \cdot \mathrm{G}(b)$
- $(a \cdot f)^\sharp \quad = \quad \mathrm{F}(a) \cdot f^\sharp$

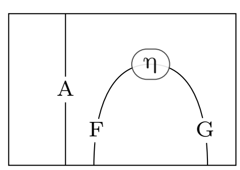
We can work out that the adjoint complements of components of the unit and counit of an adjunction are identities, which are easily seen in string diagrams:

id(G(B)) =


$\overset{\_^\sharp}{\longmapsto}$


= ε(B)

id(F(A)) =


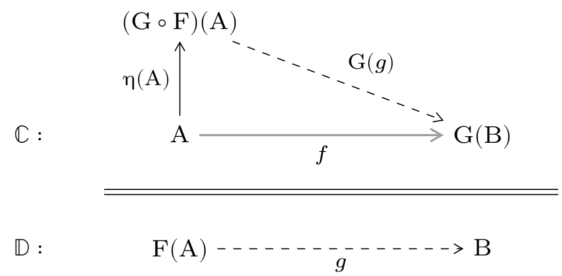
$\overset{\_^\flat}{\longmapsto}$


= η(A)

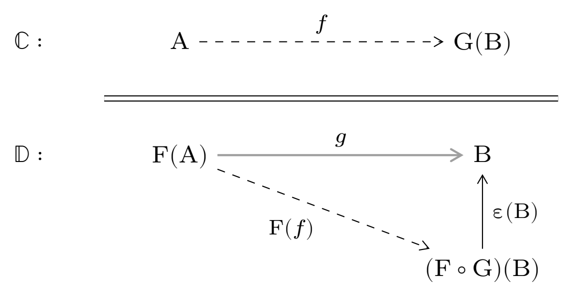
Two more structural characterizations of an adjunction – which are dual to one anther – are given by the universal properties of its unit and counit.

Definition 4.3.2.3 (adjunction – universal property of unit characterization) Anti-parallel functors $\mathrm{F} : \mathbb{C} \to \mathbb{D}$ and $\mathrm{G} : \mathbb{D} \to \mathbb{C}$ form and adjunction $\mathrm{F} \dashv \mathrm{G}$ if there is a natural transformation $\eta : \mathrm{id}(\mathbb{C}) \to \mathrm{F} \cdot \mathrm{G}$ such that for any $f : \mathbb{C}\,(\mathrm{A} \to$

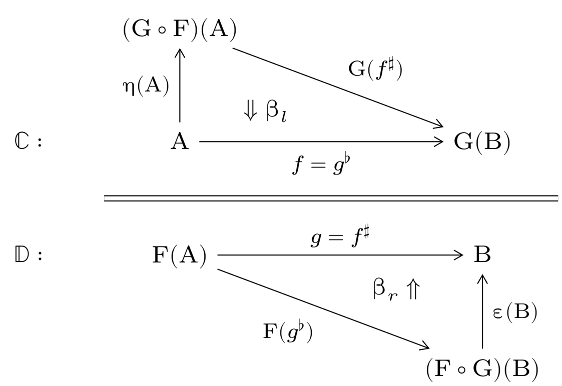
G(B)) there is a unique $g : \mathbb{D}\,(\mathrm{F}(\mathrm{A}) \to \mathrm{B})$ such that $\eta(\mathrm{A}) \cdot \mathrm{G}(g) = f$:



Definition 4.3.2.4 (adjunction – universal property of counit characterization) Anti-parallel functors $\mathrm{F} : \mathbb{C} \to \mathbb{D}$ and $\mathrm{G} : \mathbb{D} \to \mathbb{C}$ form and adjunction $\mathrm{F} \dashv \mathrm{G}$ if there is a natural transformation $\varepsilon : \mathrm{G} \cdot \mathrm{F} \to \mathrm{id}(\mathbb{D})$ such that for any $g : \mathbb{D}\,(\mathrm{F}(\mathrm{A}) \to \mathrm{B})$ there is a unique $f : \mathbb{C}\,(\mathrm{A} \to \mathrm{G}(\mathrm{B}))$ such that $\mathrm{F}(f) \cdot \varepsilon(\mathrm{B}) = g$:



Of course, $g = f^\sharp$ and $f = g^\flat$. So the "internal picture" of an adjunction looks like this:



The 2-cells labeled "β_l" and "β_r" are both equalities because $\mathbb{C}$ and $\mathbb{D}$ are just ordinary (1-dimensional) categories so equality is the only possible kind of 2-cell. But it is convenient to give them a suggestive name and orientation.

4.3.3 Conversion Relations

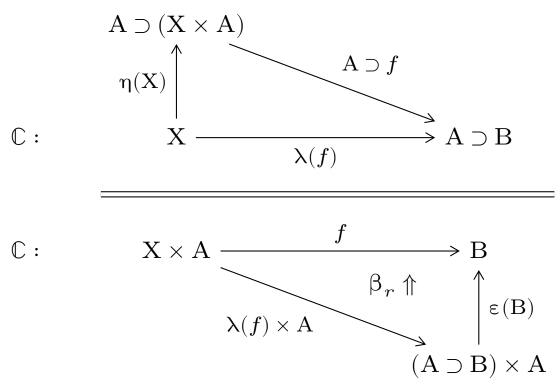
The universal property of the counit should immediately remind you of the definition of an *exponential* object. Indeed, for a fixed object A, we have endofunctors

$$- \times \mathrm{A} \quad : \quad \mathbb{C} \to \mathbb{C} \qquad \text{and} \qquad \mathrm{A} \supset - \quad : \quad \mathbb{C} \to \mathbb{C}$$

and an adjunction

$$- \times \mathrm{A} \quad \dashv \quad \mathrm{A} \supset -$$

The counit of this adjunction is the *evaluation map* and the unit is the *pairing map*. The internal picture of this adjunction looks like this:



Recall our interpretations for the introduction and elimination rules for function types interpreted as exponentials (section 3.4.3):

$$[\![\lambda x \,.\, \mathrm{M}]\!] \quad := \quad \lambda([\![\mathrm{M}]\!]) \qquad \text{and} \qquad [\![\mathrm{P}\ \mathrm{N}]\!] \quad := \quad \langle [\![\mathrm{P}]\!] \,,\, [\![\mathrm{N}]\!] \rangle \cdot \varepsilon$$

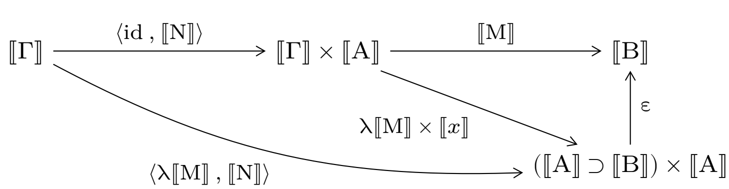
where $\Gamma \,,\, x : \mathrm{A} \ \vdash \ \mathrm{M} : \mathrm{B}$, $\Gamma \ \vdash \ \mathrm{N} : \mathrm{A}$, and $\Gamma \ \vdash \ \mathrm{P} : \mathrm{A} \to \mathrm{B}$. The universal property of exponentials is just the universal property of the counit of this adjunction. Its commuting triangle β_r,

$$\lambda(f) \times \mathrm{A} \cdot \varepsilon(\mathrm{B}) \quad = \quad f$$

expresses the β-equivalence relation for function type:

$$(\lambda x \,.\, \mathrm{M})\mathrm{N} \quad \overset{\beta}{\simeq} \quad \mathrm{M}[x {\mapsto} \mathrm{N}]$$

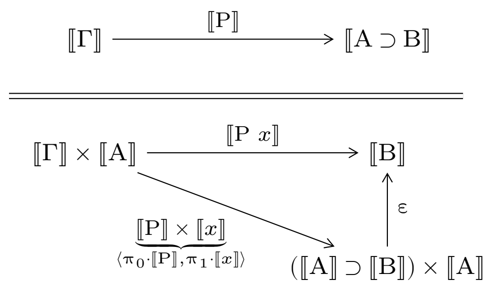
as



Furthermore, the bijection between the arrows f and $\lambda(f)$ expresses the η-equivalence relation for function type:

$$\mathrm{P} \quad \overset{\eta}{\simeq} \quad \lambda x \,.\, \mathrm{P}x$$

as



where $[\![\mathrm{P}]\!] \times [\![x]\!] = \langle \pi_0 \cdot [\![\mathrm{P}]\!] \,,\, \pi_1 \cdot [\![x]\!] \rangle$ by lemma 3.2.1.2 and exercise 3.2.2.3, and the projections interpret *context weakening*, which is silent in the syntax of type theory.

Indeed, the $\beta\eta$-conversion relations of all the simple types we have considered here can be characterized in this way due to the existence of adjunctions

$$- + - \quad \dashv \quad \Delta \quad \dashv \quad - \times - \qquad \text{and} \qquad 0 \quad \dashv \quad ! \quad \dashv \quad 1$$

where Δ and ! are respectively the *diagonal* and *bang* functors:

$$\begin{array}{ccc} & \underline{\Delta} & \\ \mathbb{C} & \longrightarrow & \mathbb{C} \times \mathbb{C} \\ \mathrm{A} & \longmapsto & (\mathrm{A} \,,\, \mathrm{A}) \\ f & \longmapsto & (f \,,\, f) \end{array} \qquad\qquad \begin{array}{ccc} & \underline{!} & \\ \mathbb{C} & \longrightarrow & \mathbb{1} \\ \mathrm{A} & \longmapsto & \star \\ f & \longmapsto & \mathrm{id}(\star) \end{array}$$

4.3.4 Context Distributivity Revisited

The context distributivity laws for the positively-presented types, the *distributive law* and *absorption law*, are actually both instances of a more general result about cartesian closed categories and adjoint functors called **Frobenius reciprocity**.

Proposition 4.3.4.1 (Frobenius reciprocity) For anti-parallel functors $\mathrm{F} : \mathbb{C} \to \mathbb{D}$ and $\mathrm{G} : \mathbb{D} \to \mathbb{C}$ between cartesian closed categories, if there is an adjunction $\mathrm{F} \dashv \mathrm{G}$ and the right adjoint G preserves exponentials then for objects $\mathrm{A} : \mathbb{C}$ and $\mathrm{X} : \mathbb{D}$,

$$\mathrm{X} \times \mathrm{F}(\mathrm{A}) \quad \cong \quad \mathrm{F}(\mathrm{G}(\mathrm{X}) \times \mathrm{A})$$

The idea is that X is the interpretation of some ambient context and F is the functor determining some positively-presented type. Admittedly, the condition that G preserves exponentials seems unmotivated, but when you try to prove the result, you see that it is exactly what is needed to make it go through.

Proof. By the *Yoneda principle*, for an arbitrary $Z : \mathbb{D}$,

$$\begin{array}{ll} & \mathbb{D}\,(X \times F(A) \to Z) \\ \cong & [\text{product symmetry}] \\ & \mathbb{D}\,(F(A) \times X \to Z) \\ \cong & [\text{currying}] \\ & \mathbb{D}\,(F(A) \to X \supset Z) \\ \cong & [\text{adjoint complement } -^{\flat}] \\ & \mathbb{C}\,(A \to G(X \supset Z)) \\ \cong & [\text{assumption that G preserves exponentials}] \\ & \mathbb{C}\,(A \to G(X) \supset G(Z)) \\ \cong & [\text{uncurrying}] \\ & \mathbb{C}\,(A \times G(X) \to G(Z)) \\ \cong & [\text{adjoint complement } -^{\sharp}] \\ & \mathbb{D}\,(F(A \times G(X)) \to Z) \\ \cong & [\text{product symmetry}] \\ & \mathbb{D}\,(F(G(X) \times A) \to Z) \end{array}$$

□

For example, in the case of the adjunction $- + - \dashv \Delta$, Frobenius reciprocity tells us:

$$X \times (+(A\,,B)) \quad \cong \quad +(\Delta(X) \times (A\,,B))$$

or, in other words:

$$X \times (A + B) \quad \cong \quad (X \times A) + (X \times B)$$

which is the distributive law!

What is really going on here is that there is a *natural isomorphism* in the *functor category* $(\mathbb{D} \times \mathbb{C}) \supset \mathbb{D}$,

$$\underline{\ ^1\ } \times F(\underline{\ ^2\ }) \quad \cong \quad F(G(\underline{\ ^1\ }) \times \underline{\ ^2\ })$$