# Introduction to Categorical Logic

[DRAFT: October 7, 2019]

Steve Awodey Andrej Bauer

October 7, 2019

# Contents

# Chapter 2

# First-Order Logic

Having considered equational theories, we now move on to first-order logic. This is the usual predicate logic with propositional connectives like $\wedge$ and $\Rightarrow$, and quantifiers $\forall$ and $\exists$. The general approach to studying logic via category theory is to determine categorical structures that model the first-order logical operations, or a suitable fragment of it, and then consider categories with these structures and functors that preserve them. Here adjoint functors play an imporant role, as the basic logical operations are recognized as adjoints. We again show that the semantics is "functorial", meaning that models of a theory are functors that preserve suitable categorical structure. We again construct classifying categories representing theories, which are the counterparts of the algebraic theories that we have already met.

Let us demonstrate our approach informally with an example. In section **??** we considered models of algebraic theories in categories with finite products. Recall that e.g. a group is a structure of the form:

$$e : 1 \to G \,, \qquad m : G \times G \to G \,, \qquad i : G \to G \,.$$

for which, moreover, certain diagrams built from these basic arrows must commute. We can express some properties of groups in terms of further equations, for example commutativity

$$x \cdot y = y \cdot x \,.$$

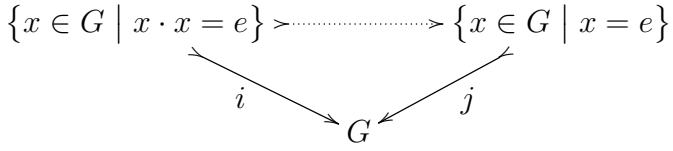
As we saw, such equations can be interpreted in any category with finite products. This provides a large scope for categorical semantics of algebraic theories.

However, there are also many significant properties of algebraic structures which cannot be expressed with equations. Consider the statement that a group $(G, e, m, i)$ has no non-trivial square roots of unity,

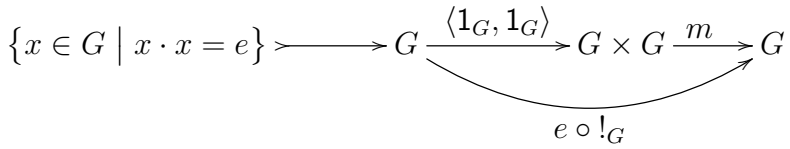
$$\forall\, x : G \,.\, (x \cdot x = e \Rightarrow x = e) \;. \tag{2.1}$$

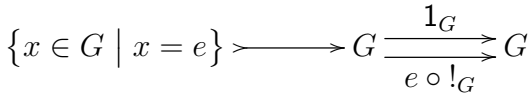
This is a first-order logical statement which cannot be rewritten as a system of equations (proof!). To see what its categorical interpretation ought to be, we look at its usual set-theoretic interpretation. Each of subformulas, $x \cdot x = e$ and $x = e$, determines a subset

of $G$:




The implication $x \cdot x = e \Rightarrow x = e$ holds when $\{x \in G \mid x \cdot x = e\}$ is contained in $\{x \in G \mid x = e\}$. In categorical language we can say that the inclusion $i$ factors through the inclusion $j$. Observe also that such a factorization is necessarily a mono and is unique, if it exists. The defining formulas of the subsets $\{x \in G \mid x \cdot x = e\}$ and $\{x \in G \mid x = e\}$ are equations, and so the subsets themselves can be constructed as equalizers (as above, interpreting $\cdot$ as $m$):




In sum, we can interpret condition (2.1) in any category with products and equalizers, i.e. in any category with finite limits.[1] This allows us to define the notion of a group without square roots of unity in any category $\mathcal{C}$ with finite limits as an object $G$ with morphisms $e : 1 \to G$, $m : G \times G \to G$ and $i : G \to G$ such that $(G, e, m, i)$ is a group in $\mathcal{C}$, and the equalizer of $m \circ \langle 1_G, 1_G \rangle$ and $e \circ !_G$ factors through $e : 1 \to G$.

The aim of this chapter is to analyze how such examples can be treated in general. We want to relate first-order logic and fragments of it to categorical structures that are suitable for the interpretation of the logic. The general outline will be as follows:

1. A language $\mathcal{L}$ for a first-order theory consists, as usual, of some basic relation, function, and constant symbols, say $\mathcal{L} = (R, f, c)$.

2. An $\mathcal{L}$-structure in a category $\mathcal{C}$ with finite limits is an interpretation of $\mathcal{L}$ in $\mathcal{C}$ as an object $M$ equipped with corresponding relations and operations (of appropriate arities), e.g.

$$\begin{aligned} R^M &\rightarrowtail M \times \cdots \times M \\ f^M &: M \times \cdots \times M \to M \\ c^M &: 1 \to M. \end{aligned}$$

[1] We are *not* claiming that finite limits suffice for an interpretation of arbitrary formulas built from universal quantifiers and implications. The formula at hand has a very special form $\forall x \,.\, (\varphi(x) \Longrightarrow \psi(x))$, where $\varphi(x)$ and $\psi(x)$ do not contain further $\forall$ or $\Longrightarrow$.

3. Formulas $\varphi(x_1, \ldots, x_n)$ in (some fragment of) first-order logic will be interpreted as "generalized subsets", i.e. subobjects,

$$[\![\varphi(x_1, \ldots, x_n)]\!] \rightarrowtail M \times \cdots \times M.$$

   The interpretation makes use of categorical operations in $\mathcal{C}$ corresponding to the logical ones appearing in the formula $\varphi(x_1, \ldots, x_n)$.

4. A theory $\mathbb{T}$ in (a fragment of) first-order logic will consist of a set of (binary) sequents,

$$\varphi(x_1, \ldots, x_n) \vdash \psi(x_1, \ldots, x_n).$$

5. A model of the theory is then an interpretation $M$ in which the corresponding subobjects satisfy all the sequents of $\mathbb{T}$, in the sense that

$$[\![\varphi(x_1, \ldots, x_n)]\!] \leq [\![\psi(x_1, \ldots, x_n)]\!] \qquad \text{in } \mathsf{Sub}(M^n).$$

6. We shall give a deductive calculus for such sequents, prove that it is sound with respect to categorical models, and then use it to construct a classifying category $\mathcal{C}_\mathbb{T}$, with the expected universal property: models of $\mathbb{T}$ correspond to (structure-preserving) functors on $\mathcal{C}_\mathbb{T}$.

7. Completeness of the calculus in general follows from classification, and more specialized completeness results from embedding theorems applied to the classifying category.

## 2.1 Theories

A *first-order theory* $\mathbb{T}$ consists of an underlying *type theory* and a set of formulas in a *fragment of first-order logic*. Anticipating Chapter **??**, the type theory is given by a set of basic types, a set of basic constants together with their types, rules for forming types, and rules and axioms for deriving typing judgments

$$x_1 : A_1, \ldots, x_n : A_n \mid t : B\ ,$$

expressing that term $t$ has type $B$ in typing context $x_1 : A_1$, …, $x_n : A_n$, and a set of axioms and rules of inference which tell us which equations between terms

$$x_1 : A_1, \ldots, x_n : A_n \mid t = u : B\ ,$$

are valid. This part of the theory $\mathbb{T}$ may be regarded as providing the underlying structure, on top of which the logical formulas are defined. For first-order logic, the underlying type theory will be essentially the same as the equational logic that we already met in Chapter **??**.

A fragment of first-order logic is then given by a set of *basic relation symbols* together with a specification of which first-order operations are being considered in building formulas. Each basic relation symbol has a *signature* $(A_1, \ldots, A_n)$, which specifies the types of its arguments. The *arity* of a relation symbol is the number of arguments it takes. The judgment[2]

$$x_1 : A_1, \ldots, x_n : A_n \mid \phi \ \texttt{pred}$$

states that $\phi$ is a well-formed formula in typing context $x_1 : A_1, \ldots, x_n : A_n$. For each basic relation symbol $R$ with signature $(A_1, \ldots, A_n)$ there is an inference rule

$$\frac{\Gamma \mid t_1 : A_1 \qquad \cdots \qquad \Gamma \mid t_n : A_n}{\Gamma \mid R(t_1, \ldots, t_n) \ \texttt{pred}}$$

Depending on what fragment of first-order logic is involved, there may be other rules for forming logical formulas. For example, if equality is present, then for each type $A$ there is a rule

$$\frac{\Gamma \mid t : A \qquad \Gamma \mid u : A}{\Gamma \mid t =_A u \ \texttt{pred}}$$

and if conjunction is present, then there is a rule

$$\frac{\Gamma \mid \varphi \ \texttt{pred} \qquad \Gamma \mid \psi \ \texttt{pred}}{\Gamma \mid \varphi \wedge \psi \ \texttt{pred}}$$

Other such rules will be given when we come to the study of particular logical operations.

The basic logical judgment of a first-order theory is *logical entailment* between formulas,

$$x_1 : A_1, \ldots, x_n : A_n \mid \varphi_1, \ldots, \varphi_m \vdash \psi$$

which states that in the typing context $x_1 : A_1, \ldots, x_n : A_n$, the hypotheses $\varphi_1$, $\ldots$, $\varphi_m$ entail $\psi$. It is understood that the terms appearing in the formulas are well-typed in the typing context, and that formulas $\varphi_1, \ldots, \varphi_m, \psi$ are part of the fragment of the logic of $\mathbb{T}$. When the fragment contains conjunction $\wedge$ it is convenient to restrict attention to *binary* sequents,

$$x_1 : A_1, \ldots, x_n : A_n \mid \varphi \vdash \psi,$$

by replacing $\varphi_1, \ldots, \varphi_m$ with $\varphi_1 \wedge \ldots \wedge \varphi_m$. When the fragment contains equality, we may replace the type-theoretic equality judgments

$$x_1 : A_1, \ldots, x_n : A_n \mid t = u : B$$

with the logical statements

$$x_1 : A_1, \ldots, x_n : A_n \mid \cdot \vdash t =_B u \ .$$

[2]We follow type-theoretic practice here by adding the tag `pred` to turn what would otherwise be an exhibited formula in context into a judgement concerning the formula.

The subscript at the equality sign indicates the type at which the equality is taken. In a theory $\mathbb{T}$ there are basic entailments, or axioms, which together with the inference rules for the opertations involved can be used for deriving valid judgments, as usual.

We shall consider several standard fragments of first-order logic, determined by selecting a subset of logical connectives and quantifiers. These are as follows:

1. *Full first-order logic* is built from logical operations
$$= \quad \top \quad \bot \quad \neg \quad \wedge \quad \vee \quad \Rightarrow \quad \forall \quad \exists \, .$$
2. *Cartesian logic* is the fragment built from
$$= \quad \top \quad \wedge \, .$$
3. *Regular logic* is the fragment built from
$$= \quad \top \quad \wedge \quad \exists \, .$$
4. *Coherent logic* is the fragment built from
$$= \quad \top \quad \wedge \quad \exists \quad \bot \quad \vee \, .$$
5. A *geometric formula* is a formula of the form
$$\forall \, x : A \, . \, (\varphi \Longrightarrow \psi) \ ,$$
where $\varphi$ and $\psi$ are coherent formulas.

The names for these fragments come from the names of various categorical structures in which they are interpreted.

The well-formed terms and formulas of a first-order theory $\mathbb{T}$ constitute its *language*. It may seem that we are doing things backwards, because we should have spoken of first-order languages before we spoke of first-order theories. While this is possible for simple theories, it becomes difficult to do when we consider more complicated theories in which types and logical formulas are intertwined. In such cases the typing judgments and logical entailments may be given by a mutual recursive definition. In order to find out whether a given term is well-formed, we might have to prove a logical statement. In everyday mathematics this occurs all the time, for example, to show that the term $\int_0^\infty f$ denotes a real number, it may be necessary to prove that $f : \mathbb{R} \to \mathbb{R}$ is an integrable function and that the integral has a finite value. This is why it does not always make sense to strictly differentiate a language from a theory.[3]

In order to focus on the logical part of first-order theories, we are going to limit attention to only two very simple kinds of type theory. A *single-sorted* first-order theory has as its underlying type theory a single type $A$, and for each $k \in \mathbb{N}$ a set of basic $k$-ary function symbols. The rules for typing judgments are:

[3]However, it *does* make sense to distinguish syntax from theory. Rules of substitution and the behaviour of free and bound variables are syntactic considerations, for example.

1. Variables in contexts:

$$\frac{}{x_1 : A, \ldots, x_n : A \mid x_i : A}$$

2. For each basic function symbol $f$ of arity $k$, there is an inference rule

$$\frac{\Gamma \mid t_1 : A \qquad \cdots \qquad \Gamma \mid t_n : A}{\Gamma \mid f(t_1, \ldots, t_n) : A}$$

This much is essentially an algebraic theory. In addition, however, a single-sorted first-order theory may contain relation symbols, formulas, axioms, and rules of inference which an algebraic theory does not.

A slight generalization of a single-sorted theory is a *multi-sorted* one. Its underlying type theory is given by a set of types, and a set of basic function symbols. Each function symbol $f$ has a *signature* $(A_1, \ldots, A_n; B)$, where $n$ is the arity of $f$ and $A_1, \ldots, A_n, B$ are types. The rules for typing judgments are:

1. Variables in contexts:

$$\frac{}{x_1 : A_1, \ldots, x_n : A_n \mid x_i : A_i}$$

2. For each basic function symbol $f$ with signature $(A_1, \ldots, A_n; B)$, there is an inference rule

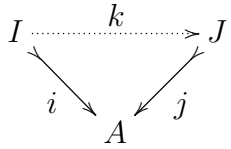
$$\frac{\Gamma \mid t_1 : A_1 \qquad \cdots \qquad \Gamma \mid t_n : A_n}{\Gamma \mid f(t_1, \ldots, t_n) : B}$$

We often write suggestively $f : A_1 \times \cdots \times A_n \to B$ to indicate that $(A_1, \ldots, A_n; B)$ is the signature of $f$. However, this does not mean that $A_1 \times \cdots \times A_n \to B$ is a type! A multi-sorted first-order theory does *not* have any type forming operations, such as $\times$ and $\to$.

## 2.2 Predicates as subobjects

Formulas of first-order logic will be interpreted as "generalized subsets", i.e. subobjects. We therefore need to review some of the basic theory of these.

Let $A$ be an object in a category $\mathcal{C}$. If $i : I \rightarrowtail A$ and $j : J \rightarrowtail A$ are monos into $A$, we say that $i$ is smaller than $j$, and write $i \leq j$, when there exists a morphism $k : I \to J$ such that the following diagram commutes:




If such a $k$ exists then it, too, is monic, since $i$ is, and it is unique, since $j$ is monic. The class $\mathsf{Mono}(A)$ of all monos into $A$ is this preordered by this relation $\leq$, it is the same as

the slice category $\mathsf{Mono}(\mathcal{C})/A$ of all monos in $\mathcal{C}$, sliced over the object $A$. Let $\mathsf{Sub}(A)$ be the poset reflection of this preorder. Thus the elements of $\mathsf{Sub}(A)$ are equivalence classes of monos into $A$, where monos $i : I \rightarrowtail A$ and $j : J \rightarrowtail A$ are equivalent when $i \leq j$ and $j \leq i$ (note that then $I \cong J$). The induced relation $\leq$ on $\mathsf{Sub}(A)$ is then a partial order.

We have to be a bit careful with the formation of $\mathsf{Sub}(A)$, since it is defined as a quotient of a *class* $\mathsf{Mono}(A)$. In many particular cases the general construction by quotients can be avoided. If we can demonstrate that the preorder $\mathsf{Mono}(A)$ is equivalent, as a category, to a poset $P$ then we can simply take $\mathsf{Sub}(A) = P$. At any rate, we usually require that $\mathsf{Sub}(A)$ is small.

**Definition 2.2.1.** A category $\mathcal{C}$ is *well-powered* when, for all $A \in \mathcal{C}$, there is at most a *set* of subobjects of $A$, so that the category $\mathsf{Mono}(A)$ is equivalent to a small poset. In other words, for every $A \in \mathcal{C}$, $\mathsf{Sub}(A)$ is a small category.

We shall often speak of subobjects as if they were monos rather than equivalence classes of monos. It is understood that we mean the subobjects represented by monos and not the monos themselves. Sometimes we refer to a mono $i : I \rightarrowtail A$ by its domain $I$ only, even though the object $I$ itself does not determine the morphism $i$. Hopefully this will not cause confusion, as it is always going to be clear which mono is meant to go along with the object $I$.

In a category $\mathcal{C}$ with finite limits the assignment $A \mapsto \mathsf{Sub}(A)$ is the object part of the *subobject functor*

$$\mathsf{Sub} : \mathcal{C}^{\mathsf{op}} \to \mathsf{Poset} \ .$$

The morphism part of $\mathsf{Sub}$ is pullback. More precisely, given a morphism $f : A \to B$, let $\mathsf{Sub}(f) = f^* : \mathsf{Sub}(B) \to \mathsf{Sub}(A)$ be the monotone map which maps the subobject $[i : I \rightarrowtail B]$ to the subobject $[f^*i : f^*I \rightarrowtail A]$, where $f^*i : f^*I \rightarrowtail A$ is a pullback of $i$ along $f$:

$$\begin{array}{ccc} f^*I & \longrightarrow & I \\ {\scriptstyle f^*i}\downarrow & \lrcorner & \downarrow{\scriptstyle i} \\ A & \xrightarrow[f]{} & B \end{array}$$

Recall that a pullback of a mono is again mono, so this definition makes sense. We need to verify (1) that if two monos $i : I \rightarrowtail A$ and $j : J \rightarrowtail A$ are equivalent, then their pullbacks are so as well; and (2) that $\mathsf{Sub}(1_A) = 1_{\mathsf{Sub}(A)}$ and $\mathsf{Sub}(g \circ f) = \mathsf{Sub}(f) \circ \mathsf{Sub}(g)$. These all follow easily from the fact that pullback is a functor $\mathcal{C}/B \to \mathcal{C}/A$, which reduces to the familiar "2-pullbacks" lemma:

**Lemma 2.2.2.** *Suppose both squares in the following diagram are pullbacks:*

$$\begin{array}{ccccc} \cdot & \longrightarrow & \cdot & \longrightarrow & \cdot \\ \downarrow & \lrcorner & \downarrow & \lrcorner & \downarrow \\ \cdot & \longrightarrow & \cdot & \longrightarrow & \cdot \end{array}$$

*Then the outer rectangle is a pullback diagram as well. Moreover, if the outer rectangle and the right square are pullbacks, then so is the left square.*

*Proof.* This is left as an exercise in diagram chasing. □

Of course, pullbacks are really only determined up to isomorphism, but this does not cause any problems because isomorphic monos represent the same subobject.

In the semantics to be given below, a formula

$$x : A \mid \varphi \ \texttt{pred}$$

will be interpreted as a subobject

$$[\![x : A \mid \varphi]\!] \rightarrowtail [\![A]\!].$$

Thus $\mathsf{Sub}(A)$ can be regarded as the poset of "predicates" on $A$, generalizing the powerset of a set $A$. Logical operations on formulas then correspond to operations on $\mathsf{Sub}(A)$. The structure of $\mathsf{Sub}(A)$ therefore determines which logical connectives can be interpreted. If $\mathsf{Sub}(A)$ is a Heyting algebra, then we can interpret the full intuitionistic propositional calculus (cf. Subsection **??**), but if $\mathsf{Sub}(A)$ only has binary meets then all that can be interpreted are $\top$ and $\wedge$. We will work out details of different operations in the following sections, but one common aspect that we require is the "stability" of the interpretation of the logical operations, in a sense that we now make clear.

### Stability and substitution

Let us consider the interpretation of substitution of terms for variables. There are two kinds of substitution, into a term, and into a formula. We may substitute a term $x : A \mid t : B$ for a variable $y$ in a term $y : B \mid u : C$ to obtain a new term $x : A \mid u[t/y] : C$. If $t$ and $u$ are interpreted as morphisms

$$[\![A]\!] \xrightarrow{[\![t]\!]} [\![B]\!] \xrightarrow{[\![u]\!]} [\![C]\!]$$

then $u[t/y]$ is interpreted as their composition:

$$[\![x : A \mid u[t/y] : C]\!] = [\![y : B \mid u : C]\!] \circ [\![x : A \mid t : B]\!] \,.$$

Thus, *substitution into a term is composition.*

The second kind of substitution occurs when we substitute a term $x : A \mid t : B$ for a variable $y$ in a formula $y : B \mid \varphi$ to obtain a new formula $x : A \mid \varphi[t/y]$. If $t$ is interpreted as a morphism $[\![t]\!] : [\![A]\!] \to [\![B]\!]$ and $\varphi$ is interpreted as a subobject $[\![\varphi]\!] \rightarrowtail [\![B]\!]$ then the interpretation of $\varphi[t/y]$ is the pullback of $[\![\varphi]\!]$ along $[\![t]\!]$:

$$\begin{array}{ccc} [\![\varphi[t/y]]\!] = [\![t]\!]^*[\![\varphi]\!] & \longrightarrow & [\![\varphi]\!] \\ \downarrow & \lrcorner & \downarrow \\ [\![A]\!] & \xrightarrow[[\![t]\!]]{} & [\![B]\!] \end{array}$$

Thus, *substitution into a formula is pullback*,

$$[\![x : A \mid \varphi[t/y]\,]\!] = [\![x : A \mid t : B]\!]^*[\![y : B \mid \varphi]\!].$$

Now, because substitution respects the syntactical, logical operations, e.g.

$$(\varphi \wedge \psi)[t/x] = \varphi[t/x] \wedge \psi[t/x],$$

the categorical structures used to interpret the various logical operations such as $\wedge$ must also behave well with respect to the interpretation of substitution, i.e. pullback. We say that a categorical property or structure is *stable (under pullbacks)* if it is preserved by pullbacks.

For example, a category $\mathcal{C}$ has *stable meets* if each poset $\mathsf{Sub}(A)$ has binary meets, and the pullback of a meet $I \wedge J \rightarrowtail A$ along any map $f : B \to A$ is the meet $f^*I \wedge f^*J \rightarrowtail A$ of the respective pullbacks,

$$f^*(I \wedge J) = f^*I \wedge f^*J.$$

This means that the meet operation,

$$\wedge : \mathsf{Sub}(A) \times \mathsf{Sub}(A) \longrightarrow \mathsf{Sub}(A)$$

is natural in $A$, in the sense that for any map $f : B \to A$ the following diagram commutes.

$$\begin{array}{ccc} \mathsf{Sub}(A) \times \mathsf{Sub}(A) & \xrightarrow{\wedge_A} & \mathsf{Sub}(A) \\ \downarrow{\scriptstyle f^* \times f^*} & & \downarrow{\scriptstyle f^*} \\ \mathsf{Sub}(B) \times \mathsf{Sub}(B) & \xrightarrow[\wedge_B]{} & \mathsf{Sub}(B) \end{array}$$

**Exercise 2.2.3.** Show that any category $\mathcal{C}$ with finite limits has stable meets in the foregoing sense: each poset $\mathsf{Sub}(A)$ has all finite meets (i.e. including the "empty meet" 1), and these are stable under pullbacks. Conclude that $\mathsf{Sub} : \mathcal{C}^{\mathsf{op}} \longrightarrow \mathsf{Posets}$ factors through the subcategory of $\wedge$-semi-lattices.

### Generalized elements

In any category, we sometimes consider arbitrary arrows $x : X \to C$ as *generalized elements* of $C$, thinking thereby of variable elements or parameters. With respect to a subobject $S \rightarrowtail C$, such an element is said to be *in the subobject*, writtten

$$x \in_C S,$$

if it factors through (any mono representing) the subobject,

$$\begin{array}{ccc} & & S \\ & \nearrow & \downarrow \\ X & \xrightarrow[x]{} & C \end{array}$$

which, observe, it then does uniquely. The following "generalized element semantics" can be a useful technique for "externalizing" the operations on subobjects into statements about generalized elements.

**Proposition 2.2.4.** *Let $C$ be any object in a category $\mathcal{C}$ with finite limits.*

1. *for the top element $1 \in \mathsf{Sub}(C)$ and any $S \in \mathsf{Sub}(C)$,*

$$S = 1 \iff x \in_C S \text{ for all } x : X \to C.$$

2. *for any $S, T \in \mathsf{Sub}(C)$,*

$$S \leq T \iff x \in_C S \text{ implies } x \in_C T, \text{ for all } x : X \to C.$$

3. *for any $S, T \in \mathsf{Sub}(C)$, and for all $x : X \to C$,*

$$x \in_C S \wedge T \iff x \in_C S \text{ and } x \in_C T.$$

4. *for the subobject $\Delta = [\langle 1_C, 1_C \rangle] \in \mathsf{Sub}(C \times C)$, and for all $x, y : X \to C$,*

$$\langle x, y \rangle \in \Delta \iff x = y.$$

5. *for the equalizer $E_{(f,g)} \rightarrowtail A$ of a pair of arrows $f, g : A \rightrightarrows B$, and for all $x : X \to A$,*

$$x \in_A E_{(f,g)} \iff fx = gx.$$

6. *for the pullback $f^*S \rightarrowtail A$ of a subobject $S \rightarrowtail B$ along any arrow $f : A \to B$, and for all $x : X \to A$,*

$$x \in_A f^*S \iff fx \in_B S.$$

**Exercise 2.2.5.** Prove the proposition.

## 2.3 Cartesian logic

As a first example we look at the logic of *cartesian categories*, which are categories with finite limits, to be called *cartesian logic*. This is a logic of formulas over a multi-sorted type theory with unit type 1. (See section **??** for multi-sorted type theories and the axioms for the unit type). The logical operations are $=$, $\top$, and $\wedge$.

### Formation rules for cartesian logic

Given a basic language consisting of a stock of relation and function symbols (with arities), the terms are built up as usual from the basic function symbols and variables (we take "constants" to be 0-ary function symbols). The rules for constructing formulas are as follows:

1. The 0-ary relation symbol $\top$ is a formula:

$$\frac{}{\Gamma \mid \top \ \texttt{pred}}$$

2. For each basic relation symbol $R$ with signature $(A_1, \ldots, A_n)$ there is a rule

$$\frac{\Gamma \mid t_1 : A_1 \qquad \cdots \qquad \Gamma \mid t_n : A_n}{\Gamma \mid R(t_1, \ldots, t_n) \ \texttt{pred}}$$

3. For each type $A$, there is a rule

$$\frac{\Gamma \mid s : A \qquad \Gamma \mid t : A}{\Gamma \mid s =_A t \ \texttt{pred}}$$

4. Conjunction:

$$\frac{\Gamma \mid \varphi \ \texttt{pred} \qquad \Gamma \mid \psi \ \texttt{pred}}{\Gamma \mid \varphi \wedge \psi \ \texttt{pred}}$$

5. Weakening:

$$\frac{\Gamma \mid \varphi \ \texttt{pred}}{\Gamma, x : A \mid \varphi \ \texttt{pred}}$$

Observe that, as usual, there is then a derived operation of substitution of terms for variables into formulas, defined by structural recursion on the above specification of formulas:

Substitution:

$$\frac{\Gamma \mid t : A \qquad \Gamma, x : A \mid \varphi \ \texttt{pred}}{\Gamma \mid \varphi[t/x] \ \texttt{pred}}$$

### Inference rules for cartesian logic

Although we do not yet need them, we state the rules of inference here, too, for the convenience of having the entire specification of cartesian logic in one place. As already mentioned, we can conveniently state this deductive calculus entirely in terms of *binary* sequents,

$$\Gamma \mid \psi \vdash \varphi.$$

We omit mention of the context $\Gamma$ when it is the same in the premisses and conclusion of a rule.

1. Weakening:
$$\frac{\Gamma \mid \psi \vdash \varphi}{\Gamma, x : A \mid \psi \vdash \varphi}$$
2. Substitution:
$$\frac{\Gamma \mid t : A \qquad \Gamma, x : A \mid \psi \vdash \varphi}{\Gamma \mid \psi[t/x] \vdash \varphi[t/x]}$$
3. Identity:
$$\frac{}{\varphi \vdash \varphi}$$
4. Cut:
$$\frac{\psi \vdash \theta \qquad \theta \vdash \varphi}{\psi \vdash \varphi}$$
5. Equality:
$$\frac{}{\psi \vdash t =_A t} \qquad \frac{\psi \vdash t =_A u \qquad \psi \vdash \varphi[t/z]}{\psi \vdash \varphi[u/z]}$$
6. Truth:
$$\frac{}{\psi \vdash \top}$$
7. Conjunction:
$$\frac{\psi \vdash \varphi \qquad \psi \vdash \psi}{\psi \vdash \varphi \wedge \psi} \qquad \frac{\psi \vdash \varphi \wedge \psi}{\psi \vdash \psi} \qquad \frac{\psi \vdash \varphi \wedge \psi}{\psi \vdash \varphi}$$

**Exercise 2.3.1.** Derive symmetry and transitivity of equality:
$$\frac{\Gamma \mid \Psi \vdash t =_A u}{\Gamma \mid \Psi \vdash u =_A t} \qquad \frac{\Gamma \mid \Psi \vdash t =_A u \qquad \Gamma \mid \Psi \vdash u =_A v}{\Gamma \mid \Psi \vdash t =_A v}$$

**Example 2.3.2.** The theory of a poset is a cartesian theory. There is one basic sort $\mathtt{P}$ and one binary relation symbol $\leq$ with signature $(\mathtt{P}, \mathtt{P})$. The axioms are the familiar axioms for reflexivity, transitivity, and antisymmetry:
$$x : \mathtt{P} \mid \cdot \vdash x \leq x$$
$$x : \mathtt{P}, y : \mathtt{P}, z : \mathtt{P} \mid x \leq y,\ y \leq z \vdash x \leq z$$
$$x : \mathtt{P}, y : \mathtt{P} \mid x \leq y,\ y \leq x \vdash x =_{\mathtt{P}} y$$

There are also many examples, such as ordered groups, ordered fields, etc., that are posets with further algebraic structure.

**Example 2.3.3.** An *equivalence relation* in a cartesian category is a model of the corresponding theory with one basic sort $\mathtt{A}$ and one binary relation symbol $\sim$ with signature $(\mathtt{A}, \mathtt{A})$. The axioms are the familiar axioms for reflexivity, symmetry, and transitivity:
$$x : \mathtt{A} \mid \cdot \vdash x \sim x$$
$$x : \mathtt{A}, y : \mathtt{A} \mid x \leq y \vdash y \leq x$$
$$x : \mathtt{A}, y : \mathtt{A}, z : \mathtt{A} \mid x \leq y,\ y \leq z \vdash x \leq z$$

Before we embark on the semantics of cartesian logic, we note a couple of useful propositions regarding cartesian categories.

**Proposition 2.3.4.** *If a category $\mathcal{C}$ has pullbacks then, for every $A \in \mathcal{C}$, $\mathsf{Sub}(A)$ has finite limits. Moreover, these are stable under pullback.*

*Proof.* The poset $\mathsf{Sub}(A)$ has finite limits if it has a top object and binary meets. The top object of $\mathsf{Sub}(A)$ is the subobject $[1_A : A \to A]$. The meet of subobjects $i : I \rightarrowtail A$ and $j : J \rightarrowtail A$ is the subobject $i \wedge j = i \circ (i^* j) = j \circ (j^* i) : I \wedge J \rightarrowtail A$ obtained by pullback, as in the following diagram:

$$\begin{array}{ccc} I \wedge J & \xrightarrow{j^* i} & J \\ {\scriptstyle i^* j}\downarrow & \lrcorner & \downarrow{\scriptstyle j} \\ I & \xrightarrow[i]{} & A \end{array}$$

It is easy to verify that $I \wedge J$ is the infimum of $I$ and $J$. Finally, stability follows from a familiar diagram chase on a cube of pullbacks. □

**Proposition 2.3.5.** *If a category has finite products and pullbacks of monos along monos then it has all finite limits.*

*Proof.* It is sufficient to show that the category has equalizers. To construct the equalizer of parallel arrows $f : A \to B$ and $g : A \to B$, first observe that the arrows

$$A \xrightarrow{\langle 1_A, f \rangle} A \times B \qquad\qquad A \xrightarrow{\langle 1_A, g \rangle} A \times B$$

are monos because the projection $\pi_0 : A \times B \to A$ is their left inverse. Therefore, we may construct the pullback

$$\begin{array}{ccc} P & \xrightarrow{p} & A \\ {\scriptstyle q}\downarrow & \lrcorner & \downarrow{\scriptstyle \langle 1_A, f \rangle} \\ A & \xrightarrow[\langle 1_A, g \rangle]{} & A \times B \end{array}$$

The morphisms $p$ and $q$ coincide because $\langle 1_A, f \rangle$ and $\langle 1_A, g \rangle$ have a common left inverse $\pi_0$:

$$p = 1_A \circ p = \pi_0 \circ \langle 1_A, f \rangle \circ p = \pi_0 \circ \langle 1_A, f \rangle \circ q = 1_A \circ q = q \,.$$

Let us show that $p : P \to A$ is the equalizer of $f$ and $g$. First, $p$ equalizes $f$ and $g$,

$$f \circ p = \pi_1 \circ \langle 1_A, f \rangle \circ p = \pi_1 \circ \langle 1_A, g \rangle \circ q = g \circ q = g \circ p \,.$$

If $k : K \to A$ also equalizes $f$ and $g$ then

$$\langle 1_A, f \rangle \circ k = \langle k, f \circ k \rangle = \langle k, g \circ k \rangle = \langle 1_A, g \rangle \circ k \,,$$

therefore by the universal property of the constructed pullback there exists a unique factorization $\overline{k} : K \to P$ such that $k = p \circ \overline{k}$, as required. □

We now explain how cartesian logic is interpreted in a cartesian category $\mathcal{C}$ (i.e. $\mathcal{C}$ is finitely complete). Let $\mathbb{T}$ be a multi-sorted cartesian theory. Recall that the type theory of $\mathbb{T}$ is specified by a set of sorts (types) and a set of basic function symbols together with their signatures, while the logic is given by a set of basic relation symbols with their signatures, and a set of axioms in the form of logical entailments,

$$\Gamma \mid \psi \vdash \varphi.$$

An interpretation of $\mathbb{T}$ in $\mathcal{C}$ is given by the following data, where $\Gamma$ stands for a typing context $x_1 : A_1, \ldots, x_n : A_n$, and $\psi$ and $\psi$ are formulas:

1. A sort $A$ is interpreted as an object $[\![A]\!]$.
2. The unit sort $\mathtt{1}$ is interpreted as the terminal object $1$.
3. A typing context $x_1 : A_1, \ldots, x_n : A_n$ is interpreted as the product $[\![A_1]\!] \times \cdots \times [\![A_n]\!]$. The empty context is interpreted as the terminal object $1$.
4. A basic function symbol $f$ with signature $(A_1, \ldots, A_m; B)$ is interpreted as a morphism $[\![f]\!] : [\![A_1]\!] \times \cdots [\![A_m]\!] \to [\![B]\!]$.
5. A term in a context $\Gamma \mid t : B$ is interpreted as a morphism $[\![\Gamma \mid t : B]\!] : [\![\Gamma]\!] \to [\![B]\!]$, as follows:
   (a) A variable $x_0 : A_1, \ldots, x_n : A_n \mid x_i : A_i$ is interpreted as the $i$-th projection $\pi_i : [\![A_1]\!] \times \cdots \times [\![A_n]\!] \to [\![A_i]\!]$.
   (b) The interpretation of $\Gamma \mid * : \mathtt{1}$ is the unique morphism $!_{[\![\Gamma]\!]} : [\![\Gamma]\!] \to 1$.
   (c) A composite term $\Gamma \mid f(t_1, \ldots, t_m) : B$, where $f$ is a basic function symbol with signature $(A_1, \ldots, A_m; B)$, is interpreted as the composition

   $$[\![\Gamma]\!] \xrightarrow{\langle [\![t_1]\!], \ldots, [\![t_m]\!] \rangle} [\![A_1]\!] \times \cdots \times [\![A_m]\!] \xrightarrow{[\![f]\!]} [\![B]\!]$$

   Here $[\![t_i]\!]$ is shorthand for $[\![\Gamma \mid t_i : A_i]\!]$.
6. A basic relation symbol $R$ with signature $(A_1, \ldots, A_n)$ is interpreted as a subobject $[\![R]\!] \in \mathsf{Sub}([\![A_1]\!] \times \cdots \times [\![A_n]\!])$.
7. A formula in a context $\Gamma \mid \varphi$ is interpreted as a subobject $[\![\Gamma \mid \varphi]\!] \in \mathsf{Sub}([\![\Gamma]\!])$. The details are given below.
8. A logical entailment $\Gamma \mid \psi \vdash \varphi$ is interpreted as an inequality $[\![\psi]\!] \leq [\![\varphi]\!]$ in $\mathsf{Sub}([\![\Gamma]\!])$.

It remains to explain how formulas are interpreted as subobjects. The logical constant $\top$ is interpreted as the maximal subobject:

$$[\![\Gamma \mid \top]\!] = [1_{[\![\Gamma]\!]} : [\![\Gamma]\!] \to [\![\Gamma]\!]] \,.$$

An atomic formula $\Gamma \mid R(t_1, \ldots, t_m)$, where $R$ is a basic relation symbol with signature $(A_1, \ldots, A_m)$ is interpreted as the left-hand side of the pullback pullback

$$\begin{array}{ccc} [\![R(t_1,\ldots,t_m)]\!] & \longrightarrow & [\![R]\!] \\ \downarrow & \lrcorner & \downarrow \\ [\![\Gamma]\!] & \xrightarrow[\langle [\![t_1]\!], \ldots, [\![t_m]\!] \rangle]{} & [\![A_1]\!] \times \cdots \times [\![A_m]\!] \end{array}$$

An equation $\Gamma \mid t =_A u$ `pred` is interpreted as the subobject represented by the equalizer of $[\![\Gamma \mid t : A]\!]$ and $[\![\Gamma \mid u : A]\!]$:

$$[\![\Gamma \mid t =_A u]\!] \rightarrowtail [\![\Gamma]\!] \underset{[\![u]\!]}{\overset{[\![t]\!]}{\rightrightarrows}} [\![A]\!]$$

By Proposition 2.3.4, each $\mathsf{Sub}(A)$ is a poset with binary meets. Thus we interpret a conjunction $\Gamma \mid \varphi \wedge \psi$ `pred` as the infimum of subobjects

$$[\![\Gamma \mid \varphi \wedge \psi]\!] = [\![\Gamma \mid \varphi]\!] \wedge [\![\Gamma \mid \psi]\!] \,.$$

A formula formed by weakening is interpreted as pullback along a projection:

$$\begin{array}{ccc} [\![\Gamma, x : A \mid \varphi]\!] & \longrightarrow & [\![\Gamma \mid \varphi]\!] \\ \downarrow & \lrcorner & \downarrow i \\ [\![\Gamma]\!] \times [\![A]\!] & \xrightarrow[\pi]{} & [\![\Gamma]\!] \end{array}$$

This pullback can be computed and the interpretation of $[\![\Gamma, x : A \mid \varphi]\!]$ turns out to be the subobject

$$[\![\Gamma \mid \varphi]\!] \times [\![A]\!] \xrightarrow{\;i \times 1_A\;} [\![\Gamma]\!] \times [\![A]\!]$$

This concludes the description of an interpretation of cartesian theory $\mathbb{T}$ in a cartesian category $\mathcal{C}$.

As was explained in the previous section, the operation of substitution of terms into formulas is then interpreted as pullback:

**Lemma 2.3.6.** *Let the formula $\Gamma, x : A \mid \varphi$ and the term $\Gamma \mid t : A$ be given. Then the substituted formula $\Gamma \mid \varphi[t/x]$ is interpreted as the pullback indicated in the following diagram:*

$$\begin{array}{ccc} [\![\Gamma \mid \varphi[t/x]]\!] & \longrightarrow & [\![\Gamma, x : A \mid \varphi]\!] \\ \downarrow & \lrcorner & \downarrow \\ [\![\Gamma]\!] & \xrightarrow[\langle 1_\Gamma, [\![t]\!] \rangle]{} & [\![\Gamma]\!] \times [\![A]\!] \end{array}$$

*Proof.* A simple induction on the structure of $\varphi$. □

**Exercise 2.3.7.** Prove this.

When we deal with many interpretations at once we name them $M$, $N$, ..., and subscript the semantic brackets accordingly, $[\![\Gamma]\!]_M$, $[\![\Gamma]\!]_N$, ...

If $\Gamma \mid \psi \vdash \psi$ is a logical entailment in $\mathbb{T}$ such that $[\![\Gamma \mid \psi]\!]_M \leq [\![\Gamma \mid \varphi]\!]_M$ holds in an interpretation $M$, then we say that $M$ *satisfies* or *models* $\Gamma \mid \psi \vdash \varphi$ and write

$$M \models \Gamma \mid \psi \vdash \varphi \,.$$

An interpretation $M$ is a *model* of $\mathbb{T}$ if it satisfies all the axioms of $\mathbb{T}$.

**Theorem 2.3.8** (Soundness of cartesian logic)**.** *If a cartesian theory $\mathbb{T}$ proves an entailment*

$$\Gamma \mid \psi \vdash \varphi$$

*then every model $M$ of $\mathbb{T}$ satisfies the entailment:*

$$M \models \Gamma \mid \psi \vdash \varphi \,.$$

*Proof.* The proof proceeds by induction on the proof of the entailment. In the following we often omit the typing context $\Gamma$ to simplify notation, and all inequalities are interpreted in $\mathsf{Sub}([\![\Gamma]\!])$. We consider all possible last steps in the proof of the entailment:

1. Weakening: if $[\![\Gamma \mid \psi]\!] \leq [\![\Gamma \mid \varphi]\!]$ in $\mathsf{Sub}([\![\Gamma]\!])$ then

   $$[\![\Gamma, x : A \mid \psi]\!] = [\![\Gamma \mid \psi]\!] \times A \leq [\![\Gamma \mid \varphi]\!] \times A = [\![\Gamma, x : A \mid \varphi]\!] \quad \text{in } \mathsf{Sub}([\![\Gamma, x : A]\!]).$$

2. Substitution: by lemma 2.3.6, substitution is interpreted by pullback so that $[\![\varphi[t/x]]\!] = \langle 1_{[\![\psi]\!]}, [\![t]\!]\rangle^* [\![\varphi]\!]$ and $[\![\psi[t/x]]\!] = \langle 1_{[\![\psi]\!]}, [\![t]\!]\rangle^* [\![\psi]\!]$. Because

   $$\langle 1_{[\![\psi]\!]}, [\![t]\!]\rangle^* : \mathsf{Sub}([\![\psi]\!]) \to \mathsf{Sub}([\![\psi]\!] \times [\![A]\!])$$

   is a functor it is a monotone map, therefore $[\![\psi]\!] \leq [\![\varphi]\!]$ implies

   $$\langle 1_{[\![\psi]\!]}, [\![t]\!]\rangle^* [\![\psi]\!] \leq \langle 1_{[\![\psi]\!]}, [\![t]\!]\rangle^* [\![\varphi]\!] \,.$$

3. Identity: trivially

   $$[\![\varphi]\!] \leq [\![\varphi]\!] \,.$$

4. Cut: if $[\![\psi]\!] \leq [\![\theta]\!]$ and $[\![\theta]\!] \leq [\![\varphi]\!]$ then clearly $[\![\psi]\!] \leq [\![\varphi]\!]$, since $\mathsf{Sub}([\![\Gamma, x : A]\!])$ is a poset.

5. Truth: trivially $[\![\psi]\!] \leq [\![\top]\!]$.

6. The rules for conjunction clearly hold because by the definition of infimum $[\![\Psi]\!] \leq [\![\varphi \wedge \psi]\!]$ if, and only if, $[\![\Psi]\!] \leq [\![\varphi]\!]$ and $[\![\Psi]\!] \leq [\![\psi]\!]$.

7. Equality: the axiom $t =_A t$ is satisfied because an equalizer of $[\![t]\!]$ with itself is the maximal subobject:
$$[\![\psi]\!] \leq [1_{[\![\Gamma]\!]} : [\![\Gamma]\!] \to [\![\Gamma]\!]] = [\![t =_A t]\!] .$$
For the other axiom, suppose $[\![\psi]\!] \leq [\![t =_A u]\!]$ and $[\![\psi]\!] \leq [\![\varphi[t/z]]\!]$. It suffices to show $[\![t =_A u]\!] \wedge [\![\varphi[t/z]]\!] \leq [\![\varphi[u/z]]\!]$ for then
$$[\![\psi]\!] \leq [\![t =_A u]\!] \wedge [\![\varphi[t/z]]\!] \leq [\![\varphi[u/z]]\!] .$$
The interpretation of $P = [\![t =_A u]\!] \wedge [\![\varphi[t/z]]\!]$ is obtained by two successive pullbacks, as in the following diagram:
$$\begin{array}{ccccc} P & \longrightarrow & [\![\varphi[t/z]]\!] & \longrightarrow & [\![\varphi]\!] \\ \downarrow & \lrcorner & \downarrow & \lrcorner & \downarrow \\ [\![t =_A u]\!] & \xrightarrow[e]{} & [\![\Gamma]\!] & \xrightarrow[\langle 1_\Gamma, [\![t]\!]\rangle]{} & [\![\Gamma]\!] \times [\![A]\!] \end{array}$$
Here $e$ is the equalizer of $[\![u]\!]$ and $[\![t]\!]$. Observe that $e$ equalizes $\langle 1_{[\![\Gamma]\!]}, [\![t]\!]\rangle$ and $\langle 1_{[\![\Gamma]\!]}, [\![u]\!]\rangle$ as well:
$$\langle 1_{[\![\Gamma]\!]}, [\![t]\!]\rangle \circ e = \langle e, [\![t]\!] \circ e\rangle = \langle e, [\![u]\!] \circ e\rangle = \langle 1_{[\![\Gamma]\!]}, [\![u]\!]\rangle \circ e .$$
Therefore, if we replace $\langle 1_{[\![\Gamma]\!]}, [\![t]\!]\rangle$ with $\langle 1_{[\![\Gamma]\!]}, [\![u]\!]\rangle$ in the above diagram, the outer rectangle still commutes. By the universal property of the pullback
$$\begin{array}{ccc} [\![\Gamma \mid \varphi[u/z]]\!] & \longrightarrow & [\![\Gamma, z : A \mid \varphi]\!] \\ \downarrow & \lrcorner & \downarrow \\ [\![\Gamma]\!] & \xrightarrow[\langle 1_\Gamma, [\![u]\!]\rangle]{} & [\![\Gamma]\!] \times [\![A]\!] \end{array}$$
it follows that $P$ factors through $[\![\varphi[u/z]]\!]$, as required.

□

**Example 2.3.9.** Recall the cartesian theory of posets (example 2.3.2). There is one basic sort $\mathtt{P}$ and one binary relation symbol $\leq$ with signature $(\mathtt{P}, \mathtt{P})$ and the axioms of reflexivity, transitivity, and antisymmetry. A poset in a cartesian category $\mathcal{C}$ is thus given by an object $P$, which is the interpretation of the sort $\mathtt{P}$, and a subobject $r : R \rightarrowtail P \times P$, which the interpretation of $\leq$, such that the axioms are satisfied. As an example we spell out when the reflexivity axiom is satisfied. The interpretation of $x : P \mid x \leq x$ is obtained by the following pullback:
$$\begin{array}{ccc} [\![x \leq x]\!] & \rightarrowtail & R \\ \downarrow & \lrcorner & \downarrow r \\ P & \xrightarrow[\delta_P]{} & P \times P \end{array}$$

where $\delta_P = \langle 1_P, 1_P \rangle$ is the diagonal. The first axiom is satisfied when $[\![x \leq x]\!] = P$, which happens if, and only if, $\delta_P$ factors through $r$. Therefore, reflexivity can be expressed as follows: there exists a "reflexivity" morphism $\rho : P \to R$ such that $r \circ \rho = \delta_P$. Equivalently, morphisms $\pi_0 \circ r$ and $\pi_1 \circ r$ have a common right inverse $\rho$.

Since the definition of a poset in a cartesian category is thus stated entirely in terms of finite limits, and these are computed pointwise in functor categories $\mathsf{Set}^{\mathbb{C}}$, it follows that a poset $P$ in $\mathsf{Set}^{\mathbb{C}}$ is the same thing as a functor $P : \mathbb{C} \to \mathsf{Poset}$. Indeed, as was the case for algebraic theories, we have an equivalence (and isomorphism, actually) of categories,

$$\mathsf{Poset}(\mathsf{Set}^{\mathbb{C}}) \;\cong\; \mathsf{Poset}(\mathsf{Set})^{\mathbb{C}} \;\cong\; \mathsf{Poset}^{\mathbb{C}}.$$

### 2.3.1 Subset types

Let us consider whether the theory of a category is a cartesian theory. We begin by expressing the definition of a category so that it can be interpreted in any cartesian category $\mathcal{C}$. An *internal category* in $\mathcal{C}$ consists of an *object of morphisms* $C_1$, an *object of objects* $C_0$, and *domain*, *codomain*, and *identity* morphisms,

$$\mathsf{dom} : C_1 \to C_0 \,, \qquad \mathsf{cod} : C_1 \to C_0 \,, \qquad \mathsf{id} : C_0 \to C_1 \,.$$

There is also a *composition* morphism $c : C_2 \to C_1$, where $C_2$ is obtained by the pullback

$$\begin{array}{ccc} C_2 & \xrightarrow{p_1} & C_1 \\ {\scriptstyle p_0}\downarrow & \lrcorner & \downarrow{\scriptstyle \mathsf{dom}} \\ C_1 & \xrightarrow[\mathsf{cod}]{} & C_0 \end{array}$$

The following equations must hold:

$$\begin{gathered} \mathsf{dom} \circ i = 1_{C_0} = \mathsf{cod} \circ i \,, \\ \mathsf{cod} \circ p_1 = \mathsf{cod} \circ c \,, \qquad \mathsf{dom} \circ p_0 = \mathsf{dom} \circ c \,. \\ c \circ \langle 1_{C_1}, i \circ \mathsf{dom} \rangle = 1_{C_1} = c \circ \langle i \circ \mathsf{cod}, 1_{C_1} \rangle \,, \end{gathered}$$

The first two equations state that the domain and codomain of an identity morphism $1_A$ are both $A$. The second equation states that $\mathsf{cod}\,(f \circ g) = \mathsf{cod}\, f$ and the third one that $\mathsf{dom}\,(f \circ g) = \mathsf{dom}\, g$. The fourth equation states that $f \circ 1_{\mathsf{dom}\, f} = f = 1_{\mathsf{cod}\, f} \circ f$. It remains to express associativity of composition. For this purpose we construct the pullback

$$\begin{array}{ccc} C_3 & \xrightarrow{q_2} & C_1 \\ {\scriptstyle q_{01}}\downarrow & \lrcorner & \downarrow{\scriptstyle \mathsf{dom}} \\ C_2 & \xrightarrow[\mathsf{cod} \circ p_1]{} & C_0 \end{array}$$

The object $C_3$ can be thought of as the set of triples of morphisms $(f, g, h)$ such that $\mathsf{cod}\, f = \mathsf{dom}\, g$ and $\mathsf{cod}\, g = \mathsf{dom}\, h$. We denote $q_0 = p_0 \circ q_{01}$ and $q_1 = p_1 \circ q_{01}$. The morphisms $q_0, q_1, q_2 : C_3 \to C_1$ are like three projections which select the first, second, and third element of a triple, respectively. With this notation we can write $q_{01} = \langle q_0, q_1 \rangle_{C_2}$ because $q_{01}$ is the unique morphism such that $p_0 \circ q_{01} = q_0$ and $p_1 \circ q_{01} = q_1$. The subscript $C_2$ reminds us that the "pair" $\langle q_0, q_1 \rangle_{C_2}$ is obtained by the universal property of the pullback $C_2$.

Morphisms $c \circ q_{01} : C_3 \to C_1$ and $q_2 : C_3 \to C_1$ factor through the pullback $C_2$ because

$$\mathsf{cod} \circ c \circ q_{01} = \mathsf{cod} \circ p_1 \circ q_0 = \mathsf{dom} \circ q_2 \ .$$

Thus let $r : C_3 \to C_2$ be the unique factorization for which $p_0 \circ r = c \circ q_{01}$ and $p_1 \circ r = q_2$. Because $p_0$ and $p_1$ are like projections from $C_2$ to $C_1$, morphism $r$ can be thought of as a pair of morphisms, so we write $r = \langle c \circ q_{01}, q_2 \rangle_{C_2}$. Morphism $c \circ \langle c \circ q_{01}, q_2 \rangle_{C_2} : C_3 \to C_1$ corresponds to the operations $\langle f, g, h \rangle \mapsto (f, g) \circ h$, whereas the morphism corresponding to $\langle f, g, h \rangle \mapsto f \circ (g \circ h)$ is obtained in a similar way and is equal to

$$c \circ \langle q_0, c \circ \langle q_1, q_2 \rangle_{C_2} \rangle_{C_2} : C_3 \to C_1 \ .$$

Thus associativity is expressed by the equation

$$c \circ \langle c \circ \langle q_0, q_1 \rangle_{C_2}, q_2 \rangle_{C_2} = c \circ \langle q_0, c \circ \langle q_1, q_2 \rangle_{C_2} \rangle_{C_2} \ .$$

**Example 2.3.10.** An internal category in $\mathsf{Set}$ is a small category.

We have successfully formulated the theory of a category so that it makes sense in any cartesian category. In fact, the definition of an internal category refers only to certain pullbacks, hence the notion of an internal category makes sense in any category with pullbacks. However, if we try to formulate it as a multi-sorted cartesian theory, there is a problem. Obviously, there ought to be a basic sort of objects $\mathtt{C}_0$ and a basic sort of morphisms $\mathtt{C}_1$. There are also basic function symbols with signatures

$$\mathtt{dom} : (\mathtt{C}_1; \mathtt{C}_0) \qquad \mathtt{cod} : (\mathtt{C}_1; \mathtt{C}_0) \qquad \mathtt{id} : (\mathtt{C}_0, \mathtt{C}_1) \ .$$

However, it is not clear what the signature for composition should be. It is not $(\mathtt{C}_1, \mathtt{C}_1; \mathtt{C}_1)$ because composition is undefined for non-composable pairs of morphisms. We might be tempted to postulate another basic sort $\mathtt{C}_2$ but then we would have no way of stating that $\mathtt{C}_2$ is the pullback of $\mathtt{dom}$ and $\mathtt{cod}$. And even if we somehow axiomatized the fact that $\mathtt{C}_2$ is a pullback, we would then still have to formalize the object $\mathtt{C}_3$ of composable triples, $\mathtt{C}_4$ of composable quadruples, and so on. What we lack is the ability to define the type $\mathtt{C}_2$ as a *subset type* of $\mathtt{C}_1 \times \mathtt{C}_1$.

In order to remedy the situation we need to use a richer type theory, namely one that allows *simple subset types*. We explain what these are. The formation rule for simple subset types is

$$\frac{x : A \mid \varphi \ \mathtt{pred}}{\{x : A \mid \varphi\} \ \mathtt{type}}$$

We can think of $\{x : A \mid \varphi\}$ as the subset of all those $x : A$ that satisfy $\varphi$. Note that we did not allow an arbitrary context $\Gamma$ to be present. This means that we cannot define subset types that depend on parameters, which why they are called "simple".

Inference rules for subset types are as follows:

$$\frac{\Gamma \mid t : \{x : A \mid \varphi\}}{\Gamma \mid \mathtt{in}_\varphi\, t : A} \qquad \frac{\Gamma \mid t : \{x : A \mid \varphi\}}{\Gamma \mid \cdot \vdash \varphi[t/x]} \qquad \frac{\Gamma \mid t : A \qquad \Gamma \mid \cdot \vdash \varphi[t/x]}{\Gamma \mid \mathtt{rs}_\varphi\, t : \{x : A \mid \varphi\}}$$

$$\frac{\Gamma, x : A \mid \Psi, \varphi \vdash \theta}{\Gamma, y : \{x : A \mid \varphi\} \mid \Psi[\mathtt{in}_\varphi\, y/x] \vdash \theta[\mathtt{in}_\varphi\, y/x]}$$

The first rule states that a term $t$ of subset type $\{x : A \mid \varphi\}$ can be converted to a term $\mathtt{in}_\varphi\, t$ of type $A$. We can think of the constant $\mathtt{in}_\varphi$ as the *inclusion* $\mathtt{in}_\varphi : \{x : A \mid \varphi\} \to A$. The second rule states that every term of a subset type $\{x : A \mid \varphi\}$ satisfies the defining predicate $\varphi$. The third rule states that a term $t$ of type $A$ which satisfies $\varphi$ can be converted to a term $\mathtt{rs}_\varphi\, t$ of type $\{x : A \mid \varphi\}$. A good way to think of the constant $\mathtt{rs}_\varphi$ is as a partially defined *restriction*, or a type-casting operations, $\mathtt{rs}_\varphi : A \rightharpoonup \{x : A \mid \varphi\}$.[4] The last rule tells us how to replace a variable $x$ of type $A$ and an assumption $\varphi$ about it with a variable $y$ of type $\{x : A \mid \varphi\}$ and remove the assumption. Note that this is a two-way rule.

There are two more axioms that relate inclusions and restrictions:

$$\frac{\Gamma \mid t : \{x : A \mid \varphi\}}{\Gamma \mid \cdot \vdash \mathtt{rs}_\varphi\, (\mathtt{in}_\varphi\, t) = t} \qquad\qquad \frac{\Gamma \mid t : A \qquad \Gamma \mid \cdot \vdash \varphi[t/x]}{\Gamma \mid \cdot \vdash \mathtt{in}_\varphi\, (\mathtt{rs}_\varphi\, t) = t\ .}$$

In an informal discussion it is customary for the inclusions and restrictions to be omitted, or at least for the subscript $\varphi$ to be missing.[5]

**Exercise 2.3.11.** Suppose $x : A \mid \psi$ and $x : A \mid \varphi$ are formulas. Show that

$$x : A \mid \psi \vdash \varphi$$

is provable if, and only if, $\{x : A \mid \psi\}$ factors through $\{x : A \mid \varphi\}$, which means that there exists a term $k$,

$$y : \{x : A \mid \psi\} \mid k : \{x : A \mid \varphi\}\ ,$$

such that

$$y : \{x : A \mid \psi\} \mid \cdot \vdash \mathtt{in}_\psi\, y =_A \mathtt{in}_\varphi\, k$$

is provable. Show also that $k$ is determined uniquely up to provable equality.

[4] Inclusions and restrictions are like type-casting operations in some programming languages. For example in Java, an inclusion corresponds to an (implicit) type cast from a class to its superclass, whereas a restriction corresponds to a type cast from a class to a subclass. Must I write that Java is a registered trademark of Sun Microsystems?

[5] Strictly speaking, even the notation $\mathtt{in}_\varphi\, t$ is imprecise because it does not indiciate that $\phi$ stands in the context $x : A$. The correct notation would be $\mathtt{in}_{(x:A\mid\varphi)}\, t$, where $x$ is bound in the subscript. A similar remark holds for $\mathtt{rs}_\varphi\, t$.

**Example 2.3.12.** We are now able to formulate the theory of a category as a cartesian theory whose underlying type theory has product types and subset types. The basic types are the type of objects $\mathtt{C}_0$ and the type of morphisms $\mathtt{C}_1$. We define the type $\mathtt{C}_2$ to be

$$\mathtt{C}_2 \equiv \{p : C_1 \times C_1 \mid \mathtt{cod}(\mathtt{fst}\, p) = \mathtt{dom}(\mathtt{snd}\, p)\} .$$

The basic function symbols and their signatures are:

$$\mathtt{dom} : \mathtt{C}_1 \to \mathtt{C}_0 , \qquad \mathtt{cod} : \mathtt{C}_1 \to \mathtt{C}_0 , \qquad \mathtt{id} : \mathtt{C}_0 \to \mathtt{C}_1 , \qquad \mathtt{c} : \mathtt{C}_2 \to \mathtt{C}_1 .$$

The axioms are:

$$\begin{aligned}
a : \mathtt{C}_0 \mid \cdot &\vdash \mathtt{dom}(\mathtt{id}(a)) = a \\
a : \mathtt{C}_0 \mid \cdot &\vdash \mathtt{cod}(\mathtt{id}(a)) = a \\
f : \mathtt{C}_1, g : \mathtt{C}_1 \mid \mathtt{cod}(f) = \mathtt{dom}(g) &\vdash \mathtt{dom}(\mathtt{c}(\mathtt{rs}\, \langle f, g\rangle)) = f \\
f : \mathtt{C}_1, g : \mathtt{C}_1 \mid \mathtt{cod}(f) = \mathtt{dom}(g) &\vdash \mathtt{cod}(\mathtt{c}(\mathtt{rs}\, \langle f, g\rangle)) = g \\
f : \mathtt{C}_1 \mid \cdot &\vdash \mathtt{c}(\mathtt{rs}\, \langle \mathtt{id}(\mathtt{dom}(f)), f\rangle) = f \\
f : \mathtt{C}_1 \mid \cdot &\vdash \mathtt{c}(\mathtt{rs}\, \langle f, \mathtt{id}(\mathtt{cod}(f))\rangle) = f
\end{aligned}$$

Lastly, the associativity axiom is

$$f : \mathtt{C}_1, g : \mathtt{C}_1, h : \mathtt{C}_1 \mid \mathtt{cod}(f) = \mathtt{dom}(g), \mathtt{cod}(g) = \mathtt{dom}(h) \vdash \mathtt{c}(\mathtt{rs}\, \langle \mathtt{c}(\mathtt{rs}\, \langle f, g\rangle), h\rangle) = \mathtt{c}(\mathtt{rs}\, \langle f, \mathtt{c}(\mathtt{rs}\, \langle g, h\rangle)\rangle) .$$

This notation is quite unreadable. If we write $g \circ f$ instead of $\mathtt{c}(\mathtt{rs}\, \langle f, g\rangle)$ then the axioms take on a more familiar form. For example, associativity is just $h \circ (g \circ f) = (h \circ g) \circ f$. However, we need to remember that we may form the term $g \circ f$ only if we first prove $\mathtt{dom}(g) = \mathtt{cod}(f)$.

A subset type $\{x : A \mid \varphi\}$ is interpreted as the domain of a monomorphism representing $x : A \mid \varphi$:

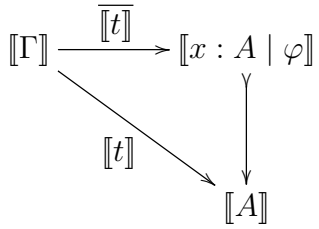
$$[\![\{x : A \mid \varphi\}]\!] \rightarrowtail [\![A]\!] \quad \text{(over } [\![x : A \mid \varphi]\!]\text{)}$$

Some care must be taken here because monos representing a given subobject are only determined up to isomorphism. We assume that a suitable canonical choice of monos can be made.

An inclusion $\Gamma \mid \mathtt{in}_\varphi\, t : A$ is interpreted as the composition

$$[\![\Gamma]\!] \xrightarrow{[\![t]\!]} [\![\{x : A \mid \varphi\}]\!] \rightarrowtail [\![A]\!] \quad \text{(over } [\![x : A \mid \varphi]\!]\text{)}$$

A restriction $\Gamma \mid \mathtt{rs}_\varphi\, t : \{x : A \mid \varphi\}$ is interpreted as the unique $\overline{[\![t]\!]}$ which makes the following diagram commute:




**Exercise 2.3.13.** Formulate and prove a soundness theorem for subset types. Pay attention to the interpretation of restrictions, where you need to show unique existence of $\overline{[\![t]\!]}$.