Introduction to Part 0

In Part 0 we recall the basic background in category theory which may be required in later portions of this book. The reader who is familiar with category theory should certainly skip Part 0, but even the reader who is not is advised to consult it only in addition to standard texts.

Most of the material in Part 0 is standard and may also be found in other books. Therefore, on the whole we shall refrain from making historical remarks. However, our exposition differs from treatments elsewhere in several respects.

Firstly, our exposition is slanted towards readers with some acquaintance with logic. Quite early we introduce the notion of a 'deductive system'. For us, this is just a category without the usual equations between arrows. In particular, we do not insist that a deductive system is freely generated from certain axioms, as is customary in logic. In fact, we really believe that logicians should turn attention to categories, which are deductive systems with suitable equations between proofs.

Secondly, we have summarized some of the main thrusts of category theory in the form of succinct slogans. Most of these are due to Bill Lawvere (whose influence on the development of category theory is difficult to overestimate), even if we do not use his exact words. Slogan V represents the point of view of a series of papers by one of the authors in collaboration with Basil Rattray.

Thirdly, we have emphasized the algebraic or equational nature of many of the systems studied in category theory. Just as groups or rings are algebraic over sets, it has been known for a long time that categories with finite products are equational over graphs. More recently, Albert Burroni made the surprising discovery that categories with equalizers are also algebraic over graphs. We have included this result, without going into his more technical concept of 'graphical algebra'.

In Part 0, as in the rest of this book, we have been rather cavalier about set theoretical foundations. Essentially, we are using Gödel–Bernays, as do most mathematicians, but occasionally we refer to universes in the sense of Grothendieck. The reason for our lack of enthusiasm in presenting the foundations properly is our belief that mathematics should be based on a version of type theory, a variant of which adequate for arithmetic and analysis is developed in Part II. For a detailed discussion of these foundational questions see Hatcher (1982, Chapter 8.)

1 Categories and functors

In this section we present what our reader is expected to know about category theory. We begin with a rather informal definition.

Definition 1.1. A *concrete category* is a collection of two kinds of entities, called *objects* and *morphisms*. The former are sets which are endowed with some kind of structure, and the latter are mappings, that is, functions from one object to another, in some sense preserving that structure. Among the morphisms, there is attached to each object A the *identity mapping* 1_A: $A \to A$ such that $1_A(a) = a$ for all $a \in A$. Moreover, morphisms $f: A \to B$ and $g: B \to C$ may be *composed* to produce a morphism $gf: A \to C$ such that $(gf)(a) = g(f(a))$ for all $a \in A$. (See also Exercise 2 below.)

Examples of concrete categories abound in mathematics; here are just three:

Example C1. The category of *sets*. Its objects are arbitrary sets and its morphisms are arbitrary mappings. We call this category '**Sets**'.

Example C2. The category of *monoids*. Its objects are monoids, that is, semigroups with unity element, and its morphisms are homomorphisms, that is, mappings which preserve multiplication (the semigroup operation) and the unity element.

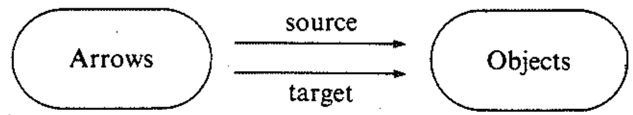
Example C3. The category of *preordered sets*. Its objects are preordered sets, that is, sets with a transitive and reflexive relation on them, and its morphisms are monotone mappings, that is, mappings which preserve this relation.

The reader will be able to think of many other examples: the categories of rings, topological spaces and Banach algebras, to name just a few. In fact, one is tempted to make a generalization, which may be summed up as follows, provided we understand 'object' to mean 'structured set'.

Slogan I. Many objects of interest in mathematics congregate in concrete categories.

We shall now progress from concrete categories to abstract ones, in three easy stages.

Definition 1.2. A *graph* (usually called a *directed graph*) consists of two classes: the class of *arrows* (or *oriented edges*) and the class of *objects* (usually called *nodes* or *vertices*) and two mappings from the class of arrows to the class of objects, called *source* and *target* (often also *domain* and *codomain*).



One writes '$f: A \to B$' for 'source $f = A$ and target $f = B$'. A graph is said to be *small* if the classes of objects and arrows are sets.

Example C4. The category of small *graphs* is another concrete category. Its objects are small graphs and its morphisms are functions F which send arrows to arrows and vertices to vertices so that, whenever $f: A \to B$, then $F(f): F(A) \to F(B)$.

A *deductive system* is a graph in which to each object A there is associated an arrow $1_A: A \to A$, the *identity* arrow, and to each pair of arrows $f: A \to B$ and $g: B \to C$ there is associated an arrow $gf: A \to C$, the *composition* of f with g. A logician may think of the objects as *formulas* and of the arrows as *deductions* or *proofs*, hence of

$$\frac{f: A \to B \quad g: B \to C}{gf: A \to C}$$

as a *rule of inference*. (Deductive systems will be discussed further in Part I.)

A *category* is a deductive system in which the following equations hold, for all $f: A \to B$, $g: B \to C$ and $h: C \to D$:

$$f1_A = f = 1_B f, \quad (hg)f = h(gf).$$

Of course, all concrete categories are categories. A category is said to be *small* if the classes of arrows and objects are sets. While the concrete categories described in examples 1 to 4 are not small, a somewhat surprising observation is summarized as follows:

Slogan II. Many objects of interest to mathematicians are themselves small categories.

Example C1′. Any set can be viewed as a category: a small *discrete* category. The objects are its elements and there are no arrows except the obligatory identity arrows.

Example C2′. Any monoid can be viewed as a category. There is only one object, which may remain nameless, and the arrows of the monoid are its elements. In particular, the identity arrow is the unity element. Composition is the binary operation of the monoid.

Example C3′. Any preordered set can be viewed as a category. The objects are its elements and, for any pair of objects (a, b), there is at most one arrow $a \to b$, exactly one when $a \leqslant b$.

It follows from slogans I and II that small categories themselves should be the objects of a category worthy of study.

Example C5. The category **Cat** has as objects small categories and as morphisms functors, which we shall now define.

Definition 1.3. A *functor* $F: \mathscr{A} \to \mathscr{B}$ is first of all a morphism of graphs (see Example C4), that is, it sends objects of $\mathscr{A}$ to objects of $\mathscr{B}$ and arrows of $\mathscr{A}$ to arrows of $\mathscr{B}$ such that, if $f: A \to A'$, then $F(f): F(A) \to F(A')$. Moreover, a functor preserves identities and composition; thus

$$F(1_A) = 1_{F(A)}, \quad F(gf) = F(g)F(f).$$

In particular, the identity functor $1_{\mathscr{A}}: \mathscr{A} \to \mathscr{A}$ leaves objects and arrows unchanged and the composition of functors $F: \mathscr{A} \to \mathscr{B}$ and $G: \mathscr{B} \to \mathscr{C}$ is given by

$$(GF)(A) = G(F(A)), \quad (GF)(f) = G(F(f)),$$

for all objects A of $\mathscr{A}$ and all arrows $f: A \to A'$ in $\mathscr{A}$.

The reader will now easily check the following assertion.

Proposition 1.4. When sets, monoids and preordered sets are regarded as small categories, the morphisms between them are the same as the functors between them.

The above definition of a functor $F: \mathscr{A} \to \mathscr{B}$ applies equally well when $\mathscr{A}$ and $\mathscr{B}$ are not necessarily small, provided we allow mappings between classes. Of special interest is the situation when $\mathscr{B} = \mathbf{Sets}$ and $\mathscr{A}$ is small.

Slogan III. Many objects of interest to mathematicians may be viewed as functors from small categories to **Sets**.

Example F1. A set may be viewed as a functor from a discrete one-object category to **Sets**.

Example F2. A small graph may be viewed as a functor from the small category $\cdot \rightrightarrows \cdot$ (with identity arrows not shown) to **Sets**.

Example F3. If $\mathscr{M} = (M, 1, \cdot)$ is a monoid viewed as a one-object category, an $\mathscr{M}$-set may be regarded as a functor from $\mathscr{M}$ to **Sets**. (An *$\mathscr{M}$-set* is a set A together with a mapping $M \times A \to A$, usually denoted by $(m, a) \mapsto ma$, such that $1a = a$ and $(m \cdot m')a = m(m'a)$ for all $a \in A$, m and $m' \in M$.)

Once we admit that functors $\mathscr{A} \to \mathscr{B}$ are interesting objects to study, we should see in them the objects of yet another category. We shall study such functor categories in the next section. For the present, let us mention two other ways of forming new categories from old.

Example C6. From any category (or graph) $\mathscr{A}$ one forms a new category (respectively graph) $\mathscr{A}^{op}$ with the same objects but with arrows reversed, that is, with the two mappings 'source' and 'target' interchanged. $\mathscr{A}^{op}$ is called the *opposite* or *dual* of $\mathscr{A}$. A functor from $\mathscr{A}^{op}$ to $\mathscr{B}$ is often called a *contravariant* functor from $\mathscr{A}$ to $\mathscr{B}$, but we shall avoid this terminology except for occasional emphasis.

Example C7. Given two categories $\mathscr{A}$ and $\mathscr{B}$, one forms a new category $\mathscr{A} \times \mathscr{B}$ whose objects are pairs (A, B), A in $\mathscr{A}$ and B in $\mathscr{B}$, and whose arrows are pairs $(f, g): (A, B) \to (A', B')$, where $f: A \to A'$ in $\mathscr{A}$ and $g: B \to B'$ in $\mathscr{B}$. Composition of arrows is defined componentwise.

Definition 1.5. An arrow $f: A \to B$ in a category is called an *isomorphism* if there is an arrow $g: B \to A$ such that $gf = 1_A$ and $fg = 1_B$. One writes $A \cong B$ to mean that such an isomorphism exists and says that A is *isomorphic* with B.

In particular, a functor $F: \mathscr{A} \to \mathscr{B}$ between two categories is an isomorphism if there is a functor $G: \mathscr{B} \to \mathscr{A}$ such that $GF = 1_{\mathscr{A}}$ and $FG = 1_{\mathscr{B}}$. We also remark that a group is a one-object category in which all arrows are isomorphisms.

To end this section, we shall record three basic isomorphisms. Here **1** is the category with one object and one arrow.

Proposition 1.6. For any categories $\mathscr{A}$, $\mathscr{B}$ and $\mathscr{C}$,

$$\mathscr{A} \times \mathbf{1} \cong \mathscr{A}, \quad (\mathscr{A} \times \mathscr{B}) \times \mathscr{C} \cong \mathscr{A} \times (\mathscr{B} \times \mathscr{C}), \quad \mathscr{A} \times \mathscr{B} \cong \mathscr{B} \times \mathscr{A}.$$

Exercises

1. Prove Propositions 1.4 and 1.6.
2. Show that for any concrete category $\mathscr{A}$ there is a functor $U: \mathscr{A} \to \mathbf{Sets}$

which 'forgets' the structure, often called the *forgetful* functor. Clearly U is *faithful* in the sense that, for all $f, g: A \rightrightarrows B$, if $U(f) = U(g)$ then $f = g$. (A more formal version of Definition 1.1 describes a *concrete* category as a pair $(\mathscr{A}, U)$, where $\mathscr{A}$ is a category and $U: \mathscr{A} \to \mathbf{Sets}$ is a faithful functor.)

3. Show that for any category $\mathscr{A}$ there are functors $\Delta: \mathscr{A} \to \mathscr{A} \times \mathscr{A}$ and $\bigcirc_{\mathscr{A}}: \mathscr{A} \to \mathbf{1}$ given on objects A of $\mathscr{A}$ by $\Delta(A) = (A, A)$ and $\bigcirc_{\mathscr{A}}(A) =$ the object of $\mathbf{1}$.

2 Natural transformations

In this section we shall investigate morphisms between functors.

Definition 2.1. Given functors $F, G: \mathscr{A} \rightrightarrows \mathscr{B}$, a *natural transformation* $t: F \to G$ is a family of arrows $t(A): F(A) \to G(A)$ in $\mathscr{B}$, one arrow for each object A of $\mathscr{A}$, such that the following square commutes for all arrows $f: A \to B$ in $\mathscr{A}$:

$$\begin{array}{ccc} F(A) & \xrightarrow{t(A)} & G(A) \\ {\scriptstyle F(f)}\downarrow & & \downarrow{\scriptstyle G(f)} \\ F(B) & \xrightarrow[t(B)]{} & G(B) \end{array}$$

that is to say, such that

$$G(f)t(A) = t(B)F(f).$$

It is this concept about which it has been said that it necessitated the invention of category theory. We shall give examples of natural transformations later. For the moment, we are interested in another example of a category.

Example C8. Given categories $\mathscr{A}$ and $\mathscr{B}$, the *functor category* $\mathscr{B}^{\mathscr{A}}$ has as objects functors $F: \mathscr{A} \to \mathscr{B}$ and as arrows natural transformations. The *identity* natural transformation $1_F: F \to F$ is of course given by stipulating that $1_F(A) = 1_{F(A)}$ for each object A of $\mathscr{A}$. If $t: F \to G$ and $u: G \to H$ are natural transformations, their *composition* $u \circ t$ is given by stipulating that $(u \circ t)(A) = u(A)t(A)$ for each object A of $\mathscr{A}$.

To appreciate the usefulness of natural transformations, the reader should prove for himself the following, which supports Slogan III.

Proposition 2.2. When objects such as sets, small graphs and $\mathscr{M}$-sets are

viewed as functors into **Sets** (see Examples F1 to F3 in Section 1), the morphisms between two objects are precisely the natural transformations. Thus, the categories of sets, small graphs and $\mathscr{M}$-sets may be identified with the functor categories $\mathbf{Sets}^{\mathbf{1}}$, $\mathbf{Sets}^{\cdot\rightrightarrows\cdot}$ and $\mathbf{Sets}^{\mathscr{M}}$ respectively.

Of course, morphisms between sets are mappings, morphisms between graphs were described in Definition 1.3 and morphisms between $\mathscr{M}$-sets are $\mathscr{M}$-homomorphisms. (An *$\mathscr{M}$-homomorphism* $f: A \to B$ between $\mathscr{M}$-sets is a mapping such that $f(ma) = mf(a)$ for all $m \in M$ and $a \in A$.)

We record three more basic isomorphisms in the spirit of Proposition 1.6.

Proposition 2.3. For any categories $\mathscr{A}$, $\mathscr{B}$ and $\mathscr{C}$,

$$\mathscr{A}^{\mathbf{1}} \cong \mathscr{A}, \quad \mathscr{C}^{\mathscr{A} \times \mathscr{B}} \cong (\mathscr{C}^{\mathscr{B}})^{\mathscr{A}}, \quad (\mathscr{A} \times \mathscr{B})^{\mathscr{C}} \cong \mathscr{A}^{\mathscr{C}} \times \mathscr{B}^{\mathscr{C}}.$$

We shall leave the lengthy proof of this to the reader. We only mention here the functor $\mathscr{C}^{\mathscr{A} \times \mathscr{B}} \to (\mathscr{C}^{\mathscr{B}})^{\mathscr{A}}$, which will be used later. We describe its action on objects by stipulating that it assigns to a functor $F: \mathscr{A} \times \mathscr{B} \to \mathscr{C}$ the functor $F^*: \mathscr{A} \to \mathscr{C}^{\mathscr{B}}$ which is defined as follows:

For any object A of $\mathscr{A}$, the functor $F^*(A): \mathscr{B} \to \mathscr{C}$ is given by $F^*(A)(B) = F(A, B)$ and $F^*(A)(g) = F(1_A, g)$, for any object B of $\mathscr{B}$ and any arrow $g: B \to B'$ in $\mathscr{B}$.

For any arrow $f: A \to A'$, $F^*(f): F^*(A) \to F^*(A')$ is the natural transformation given by $F^*(f)(B) = F(f, 1_B)$, for all objects B of $\mathscr{B}$.

Finally, to any natural transformation $t: F \to G$ between functors F, G: $\mathscr{A} \times \mathscr{B} \rightrightarrows \mathscr{C}$ we assign the natural transformation $t^*: F^* \to G^*$ which is given by $t^*(A)(B) = t(A, B)$ for all objects A of $\mathscr{A}$ and B of $\mathscr{B}$.

This may be as good a place as any to mention that natural transformations may also be composed with functors.

Definition 2.4. In the situation

$$\mathscr{D} \xrightarrow{L} \mathscr{A} \underset{G}{\overset{F}{\rightrightarrows}} \mathscr{B} \xrightarrow{K} \mathscr{C},$$

if $t: F \to G$ is a natural transformation, one obtains natural transformations $Kt: KF \to KG$ between functors from $\mathscr{A}$ to $\mathscr{C}$ and $tL: FL \to GL$ between functors from $\mathscr{D}$ to $\mathscr{B}$ defined as follows:

$$(Kt)(A) = K(t(A)), \quad (tL)(D) = t(L(D)),$$

for all objects A of $\mathscr{A}$ and D of $\mathscr{D}$.

If $H: \mathscr{A} \to \mathscr{B}$ is another functor and $u: G \to H$ another natural transformation, then the reader will easily check the following 'distributive laws':

$$K(u \circ t) = (Ku) \circ (Kt), \quad (u \circ t)L = (uL) \circ (tL).$$

If we compare Slogans I and III, we are led to ask: which categories may be viewed as categories of functors into **Sets**? In preparation for an answer to that question we need another definition.

Definition 2.5. If A and B are objects of a category $\mathscr{A}$, we denote by $\mathrm{Hom}_{\mathscr{A}}(A, B)$ the class of arrows $A \to B$. (Later, the subscript $\mathscr{A}$ will often be omitted.) If it so happens that $\mathrm{Hom}_{\mathscr{A}}(A, B)$ is a set for all objects A and B, $\mathscr{A}$ is said to be *locally small.*

One purpose of this definition is to describe the following functor.

Example F4. If $\mathscr{A}$ is a locally small category, then there is a functor $\mathrm{Hom}_{\mathscr{A}}: \mathscr{A}^{op} \times \mathscr{A} \to \mathrm{Sets}$. For an object (A, B) of $\mathscr{A}^{op} \times \mathscr{A}$, the value of this functor is $\mathrm{Hom}_{\mathscr{A}}(A, B)$, as suggested by the notation. For an arrow $(g, h): (A, B) \to (A', B')$ of $\mathscr{A}^{op} \times \mathscr{A}$, where $g: A' \to A$ and $h: B \to B'$ in $\mathscr{A}$, $\mathrm{Hom}_{\mathscr{A}}(g, h)$ sends $f \in \mathrm{Hom}_{\mathscr{A}}(A, B)$ to $hfg \in \mathrm{Hom}_{\mathscr{A}}(A', B')$.

Applying the isomorphism $\mathrm{Sets}^{\mathscr{A}^{op} \times \mathscr{A}} \to (\mathrm{Sets}^{\mathscr{A}})^{\mathscr{A}^{op}}$ of Proposition 2.3, we obtain a functor $\mathrm{Hom}^*_{\mathscr{A}}: \mathscr{A}^{op} \to \mathrm{Sets}^{\mathscr{A}}$ and, dually, a functor $\mathrm{Hom}^*_{\mathscr{A}^{op}}: \mathscr{A} \to \mathrm{Set}^{\mathscr{A}^{op}}$. We shall see that the latter functor allows us to assert that $\mathscr{A}$ is isomorphic to a 'full' subcategory of $\mathrm{Sets}^{\mathscr{A}^{op}}$.

Definition 2.6. A *subcategory* $\mathscr{C}$ of a category $\mathscr{B}$ is any category whose class of objects and arrows is contained in the class of objects and arrows of $\mathscr{C}$ respectively and which is closed under the 'operations' source, target, identity and composition. By saying that a subcategory $\mathscr{C}$ of $\mathscr{B}$ is *full* we mean that, for any objects C, C' of $\mathscr{C}$, $\mathrm{Hom}_{\mathscr{C}}(C, C') = \mathrm{Hom}_{\mathscr{B}}(C, C')$.

For example, a proper subgroup of a group is a subcategory which is not full, but the category of Abelian groups is a full subcategory of the category of all groups.

The arrows $F \to G$ in $\mathbf{Sets}^{\mathscr{A}^{op}}$ are natural transformations. We therefore write $\mathrm{Nat}(F, G)$ in place of $\mathrm{Hom}(F, G)$ in $\mathbf{Sets}^{\mathscr{A}^{op}}$.

Objects of the latter category are sometimes called 'contravariant' functors from $\mathscr{A}$ to **Sets**. Among them is the functor $h_A \equiv \mathrm{Hom}_{\mathscr{A}}(-, A)$ which sends the object A' of $\mathscr{A}$ onto the set $\mathrm{Hom}_{\mathscr{A}}(A', A)$ and the arrow $f: A' \to A''$ onto the mapping $\mathrm{Hom}_{\mathscr{A}}(f, 1_A)$: $\mathrm{Hom}_{\mathscr{A}}(A'', A) \to \mathrm{Hom}_{\mathscr{A}}(A', A)$.

The following is known as Yoneda's Lemma.

Proposition 2.7. If $\mathscr{A}$ is locally small and $F: \mathscr{A}^{op} \to \mathbf{Sets}$, then $\mathrm{Nat}(h_A, F)$ is in one-to-one correspondence with $F(A)$.

Proof. If $a \in F(A)$, we obtain a natural transformation $\check{a}: h_A \to F$ by stipulating that $\check{a}(B): \mathrm{Hom}_{\mathscr{A}}(B, A) \to F(B)$ sends $g: B \to A$ onto $F(g)(a)$. (Note that F is contravariant, so $F(g): F(A) \to F(B)$.)

Conversely, if $t: h_A \to F$ is a natural transformation, we obtain the element $t(A)(1_A) \in F(A)$. It is a routine exercise to check that the mappings $a \mapsto \check{a}$ and $t \mapsto t(A)(1_A)$ are inverse to one another.

Definition 2.8. A functor $H: \mathscr{A} \to \mathscr{B}$ is said to be *faithful* if the induced mappings $\mathrm{Hom}_{\mathscr{A}}(A, A') \to \mathrm{Hom}_{\mathscr{B}}(H(A), H(A'))$ sending $f: A \to A'$ onto $H(f)$: $H(A) \to H(A')$ for all $A', A \in \mathscr{A}$ are injective and *full* if they are surjective. A *full embedding* is a full and faithful functor which is also injective on objects, that is, for which $H(A) = H(A')$ implies $A = A'$.

Corollary 2.9. If $\mathscr{A}$ is locally small, the Yoneda functor $\mathrm{Hom}^*_{\mathscr{A}^{op}}: \mathscr{A} \to \mathbf{Sets}^{\mathscr{A}^{op}}$ is a full embedding.

Proof. Writing $H \equiv \mathrm{Hom}^*_{\mathscr{A}^{op}}$, we see that the induced mapping $\mathrm{Hom}(A, A') \to \mathrm{Nat}(H(A), H(A'))$ sends $f: A \to A'$ onto the natural transformation $H(f): H(A) \to H(A')$ which, for all objects B of $\mathscr{A}$, gives rise to the mapping $H(f)(B) = \mathrm{Hom}(1_B, f)$: $\mathrm{Hom}(B, A) \to \mathrm{Hom}(B, A')$. Now $f \in H(A')(A)$, hence $\check{f}: H(A) \to H(A')$, as defined in the proof of Proposition 2.7, is given by

$$\begin{aligned}\check{f}(B)(g) &= H(A')(g)(f) = \mathrm{Hom}_{\mathscr{A}}(g, 1_{A'})(f) \\ &= fg = \mathrm{Hom}_{\mathscr{A}}(1_B, f)(g) = H(f)(B)(g),\end{aligned}$$

hence $\check{f} = H(f)$. Thus the mapping $f \mapsto H(f)$ is a bijection and so H is full and faithful.

Finally, to show that H is injective on objects, assume $H(A) = H(A')$, then $\mathrm{Hom}(A, A) = \mathrm{Hom}(A, A')$, so A' must be the target of the identity arrow 1_A, thus $A' = A$.

Exercises

1. Prove propositions 2.2 and 2.3.
2. If **2** is the category $\cdot \to \cdot$ (with identity arrows not shown), show that the objects of $\mathscr{A}^2$ are essentially the arrows of $\mathscr{A}$ and that 'source' and 'target' may be viewed as functors $\delta, \delta': \mathscr{A}^2 \rightrightarrows \mathscr{A}$.
3. If $F, G: \mathscr{A} \rightrightarrows \mathscr{B}$ are given functors, show that a natural transformation $t: F \to G$ is essentially the same as a functor $t: \mathscr{A} \to \mathscr{A}^2$ such that $\delta t = F$ and $\delta' t = G$.
4. Show that the isomorphism in Yoneda's Lemma (Proposition 2.7) is natural in both A and F, that is, if $f: B \to A$ and $t: F \to G$ then the relevant diagrams commute.

3 Adjoint functors

Perhaps the most important concept which category theory has helped to formulate is that of adjoint functors. Aspects of this idea were known even before the advent of category theory and we shall begin by looking at one such.

We recall from Proposition 1.4 that a functor $\mathscr{A} \to \mathscr{B}$ between two preordered sets $\mathscr{A} = (A, \leqslant)$ and $\mathscr{B} = (B, \leqslant)$ regarded as categories is an order preserving mapping $F: A \to B$, that is, such that, for all elements a, a' of A, if $a \leqslant a'$ then $F(a) \leqslant F(a')$. A functor $G: \mathscr{B} \to \mathscr{A}$ in the opposite direction is said to be *right adjoint* to F provided, for all $a \in A$ and $b \in B$,

$$F(a) \leqslant b \quad \text{if and only if} \quad a \leqslant G(b).$$

Classically, a pair of order preserving mappings (F, G) is called a covariant *Galois correspondence* if it satisfies this condition.

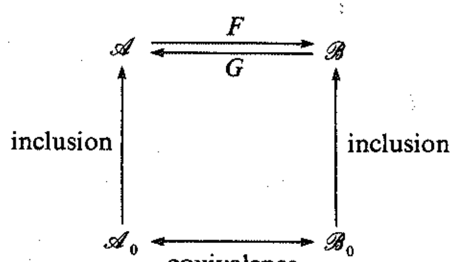
Once we have such a Galois correspondence, we see immediately that $GF: \mathscr{A} \to \mathscr{A}$ is a *closure operation*, that is, for all $a, a' \in A$,

$$a \leqslant GF(a),$$
$$GFGF(a) \leqslant GF(a),$$
$$\text{if } a \leqslant a' \quad \text{then} \quad GF(a) \leqslant GF(a').$$

Similarly, $FG: \mathscr{B} \to \mathscr{B}$ may be called an *interior operation*: it satisfies the conditions dual to the above.

In a preordered set an isomorphism $a \cong a'$ just means that $a \leqslant a'$ and $a' \leqslant a$. (In a *poset*, or *partially ordered* set, one has the antisymmetry law: if $a \cong a'$ then $a = a'$.) We note that it follows from the above that $GFGF(a) \cong GF(a)$ and, dually, $FGFG(b) \cong FG(b)$, for all $a \in A$ and $b \in B$.

The most interesting consequence of a Galois correspondence is this: the functors F and G set up a one-to-one correspondence between isomorphism classes of 'closed' elements a of A such that $GF(a) \cong a$ and isomorphism classes of 'open' elements b of B such that $FG(b) \cong b$. We also say that F and G determine an *equivalence* between the preordered set $\mathscr{A}_0$ of closed elements of $\mathscr{A}$ and the preordered set $\mathscr{B}_0$ of open elements of $\mathscr{B}$. The following picture illustrates this principle of 'unity of opposites', which will be generalized later in this section.



Before carrying out the promised generalization, let us look at a couple of examples of Galois correspondence; others will be found in the exercises.

Example G1. Take both $\mathscr{A}$ and $\mathscr{B}$ to be $(\mathbb{N}, \leqslant)$, the set of natural numbers with the usual ordering, and let

$$F(0) = 0, F(a) = p_a = \text{the } a\text{th prime number when } a > 0,$$
$$G(b) = \pi(b) = \text{the number of primes } \leqslant b.$$

Then F and G form a pair of adjoint functors and the 'unity of opposites' describes the biunique correspondence between positive integers and prime numbers.

Many examples arise from a binary relation $R \subseteq X \times Y$ between two sets X and Y. Take $\mathscr{A} = (\mathscr{P}(X), \subseteq)$, the set of subsets of X ordered by inclusion, and $\mathscr{B} = (\mathscr{P}(Y), \supseteq)$, ordered by inverse inclusion, and put

$$F(A) = \{y \in Y \mid \forall_{x \in A}(x, y) \in R\},$$
$$G(B) = \{x \in X \mid \forall_{y \in B}(x, y) \in R\},$$

for all $A \subseteq X$ and $B \subseteq Y$. This situation is called a *polarity*; it gives rise to an isomorphism between the lattice $\mathscr{A}_0$ of 'closed' subsets of X and the lattice $\mathscr{B}_0$ of 'closed' subsets of Y. (Note that the open elements of $\mathscr{B}$ are closed subsets of Y.)

Example G2. Take X to be the set of points of a plane, Y the set of halfplanes, and write $(x, y) \in R$ for $x \in y$. Then, for any set A of points, $GF(A)$ is the intersection of all halfplanes containing A, in other words, the *convex hull* of A. The 'unity of opposites' here asserts that there are two equivalent ways of describing a convex set: by the points on it or by the halfplanes containing it.

We shall now generalize the notion of adjoint functor from preordered sets to arbitrary categories. In so doing, we shall bow to a notational prejudice of many categorists and replace the letter 'G' by the letter 'U'. ('U' is for 'underlying', 'F' for 'free'.)

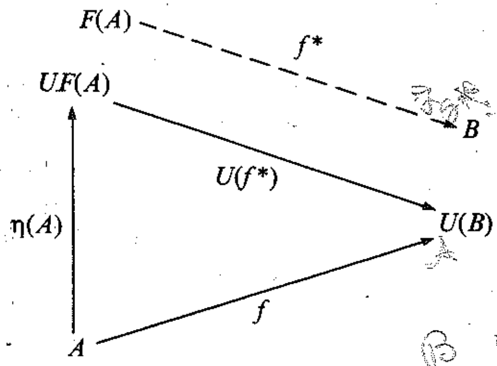
Definition 3.1. An *adjointness* between categories $\mathscr{A}$ and $\mathscr{B}$ is given by a quadruple $(F, U, \eta, \varepsilon)$, where $F: \mathscr{A} \to \mathscr{B}$ and $U: \mathscr{B} \to \mathscr{A}$ are functors and $\eta: 1_{\mathscr{A}} \to UF$ and $\varepsilon: FU \to 1_{\mathscr{B}}$ are natural transformations such that

$$(U\varepsilon) \circ (\eta U) = 1_U, \quad (\varepsilon F) \circ (F\eta) = 1_F.$$

One says that U is *right adjoint* to F or that F is *left adjoint* to U and one calls η and ε the two *adjunctions*.

Before going into examples, let us give another formulation of what will turn out to be an equivalent concept (in Proposition 3.3 below).

Definition 3.2. A solution to the *universal mapping problem* for a functor $U:\mathscr{B}\to\mathscr{A}$ is given by the following data: for each object A of $\mathscr{A}$ an object $F(A)$ of $\mathscr{B}$ and an arrow $\eta(A): A\to UF(A)$ such that, for each object B of $\mathscr{B}$ and each arrow $f:A\to U(B)$ in $\mathscr{A}$, there exists a unique arrow $f^*:F(A)\to B$ in $\mathscr{B}$ such that $U(f^*)\eta(A)=f$.



Example U1. Let $\mathscr{B}$ be the category of monoids, $\mathscr{A}$ the category of sets, $U:\mathscr{B}\to\mathscr{A}$ the forgetful (= underlying) functor, $F(A)$ the free monoid generated by the set A and $\eta(A)$ the obvious mapping of A into the underlying set of the monoid $F(A)$.

Definition 3.2′. Of special interest is the case of Definition 3.2 in which $\mathscr{B}$ is a full subcategory of $\mathscr{A}$ and $U:\mathscr{B}\to\mathscr{A}$ is the inclusion. Then $\eta(A): A\to F(A)$ may be called the *best approximation* of A by an object of $\mathscr{B}$ in the sense that, for each arrow $f:A\to B$ with B in $\mathscr{B}$, there is a unique arrow $f^*:F(A)\to B$ such that $f^*\eta(A)=f$. One then says that $\mathscr{B}$ is a full *reflective* subcategory of $\mathscr{A}$ with *reflector* F and *reflection* η.

Example U2. Let $\mathscr{A}$ be the category of Abelian groups, $\mathscr{B}$ the full subcategory of torsion free Abelian groups and $F(A)=A/T(A)$, where $T(A)$ is the torsion subgroup of A.

Proposition 3.3. Given two categories $\mathscr{A}$ and $\mathscr{B}$, there is a one-to-one correspondence between adjointnesses (F,U,η,ε) and solutions $(F,\eta,*)$ of the universal mapping problem for $U:\mathscr{B}\to\mathscr{A}$.

Proof. If (F,U,η,ε) is given, put $f^*=\varepsilon(B)F(f)$. Conversely, if U and $(F,\eta,*)$ are given, for each $f:A\to A'$, put $F(f)=(\eta(A')f)^*$ and check that this makes F a functor and η a natural transformation; moreover define $\varepsilon(B)=(1_{U(B)})^*$.

It follows from symmetry considerations that an adjointness is also equivalent to a 'co-universal mapping problem', obtained by dualizing Definition 3.2. (A left adjoint to $\mathscr{B}\to\mathscr{A}$ is a right adjoint to $\mathscr{B}^{op}\to\mathscr{A}^{op}$.)

In view of Proposition 3.3, Examples U1 and U2 are examples of adjoint functors. We shall give other examples later.

There is yet another way of looking at adjoint functors, at least when $\mathscr{A}$ and $\mathscr{B}$ are locally small.

Proposition 3.4. An adjointness (F,U,η,ε) between locally small categories $\mathscr{A}$ and $\mathscr{B}$ gives rise to and is determined by a natural isomorphism $\mathrm{Hom}_{\mathscr{B}}(F(-),-)\cong \mathrm{Hom}_{\mathscr{A}}(-,U(-))$ between functors $\mathscr{A}^{op}\times\mathscr{B}\rightrightarrows \mathbf{Sets}$.

We leave the proof of this to the reader.

Even if $\mathscr{A}$ is not locally small, there is a natural bijection between arrows $FA\to B$ in $\mathscr{B}$ and arrows $A\to UB$ in $\mathscr{A}$. Logicians may think of such a bijection as comprising two rules of inference; and this point of view has been quite influential in the development of categorical logic. An analogous situation in the propositional calculus would be the bijection between proofs of the entailments $C\wedge A\vdash B$ and $A\vdash C\Rightarrow B$ (see Exercise 4 below). Inasmuch as implication is a more sophisticated notion than conjunction, the adjointness here explains the emergence of one concept from another. This point of view, due to Lawvere, may be summarized by yet another slogan, illustrations of which will be found throughout this book (see, for instance, Exercise 6 below).

Slogan IV. Many important concepts in mathematics arise as adjoints, right or left, to previously known functors.

We summarize two important properties of adjoint functors, which will be useful later.

Proposition 3.5. (i) Adjoint functors determine each other uniquely up to natural isomorphisms.

(ii) If (U,F) and (U',F') are pairs of adjoint functors, as in the diagram

$$\mathscr{C}\ \underset{F'}{\overset{U'}{\rightleftarrows}}\ \mathscr{B}\ \underset{F}{\overset{U}{\rightleftarrows}}\ \mathscr{A},$$

then $(UU',F'F)$ is also an adjoint pair.

Exercise

1. If (F,G) is a Galois correspondence between posets $\mathscr{A}$ and $\mathscr{B}$, show that F preserves supremums and G preserves infimums. If $\mathscr{A}$ has and F preserves supremums, show that its right adjoint $G:\mathscr{B}\to\mathscr{A}$ can be calculated by the

formula

$$G(b) = \sup\{a \in \mathscr{A} \mid F(a) \leqslant b\}.$$

2. In Example G1, show that the two sets $\{F(a)+a \mid a \in \mathbb{N}\}$ and $\{G(b)+b+1 \mid b \in \mathbb{N}\}$ are complementary sets.
3. Given a commutative ring C, take X to be the set of elements of C, Y the set of prime ideals of C and define $R \subseteq X \times Y$ by writing $(x, y) \in R$ for $x \in y$. If F and G are defined as for any polarity, show that, for any subset A of X, $GF(A) = \{x \in X \mid \exists_{n \in \mathbb{N}} x^n \in A\}$, the so-called *radical* of A. Also show that the closure operation FG on the set of subsets of Y makes Y into a compact topological space called the *spectrum* of C. The 'unity of opposites' here describes a one-to-one correspondence between closed subspaces of the spectrum and ideals which are equal to their radical.
4. Take $\mathscr{A}$ and $\mathscr{B}$ to be the preordered sets of formulas of the propositional calculus, the order being entailment. For a fixed formula C, show that $F: \mathscr{A} \to \mathscr{B}$ and $G: \mathscr{B} \to \mathscr{A}$ defined by $F(A) \equiv C \wedge A$ and $G(B) \equiv C \Rightarrow B$ are a pair of adjoint functors. What is the 'unity of opposites' in this case?
5. Prove propositions 3.4 and 3.5.
6. If $\mathscr{A} = \mathscr{B} = \mathbf{Sets}$, C a given set, let $F(A) = C \times A$ and $U(B) = B^C$ for any sets A and B. Extend U and F to functors and show that U is right adjoint to F.
7. Show that the forgetful functor from **Cat** to **Sets** which sends every small category onto its set of objects has both a left and a right adjoint.
8. Show that the forgetful functor from **Cat** to the category of graphs has a left adjoint, which assigns to each graph the category 'generated by it'.

4 Equivalence of categories

We shall extend the 'unity of opposites' to general categories, but first we need to extend the notion of 'equivalence'.

Definition 4.1. An adjointness $(F, U, \eta, \varepsilon)$ is an *adjoint equivalence* if η and ε are natural isomorphisms. More generally, an *equivalence* between categories $\mathscr{A}$ and $\mathscr{B}$ is given by a pair of functors $F: \mathscr{A} \to \mathscr{B}$ and $U: \mathscr{B} \to \mathscr{A}$ such that $UF \cong 1_{\mathscr{A}}$ and $FU \cong 1_{\mathscr{B}}$.

The extra generality is an illusion: given that $\eta: 1_{\mathscr{A}} \to UF$ and $\varepsilon': FU \to 1_{\mathscr{B}}$ are isomorphisms, one obtains an adjoint equivalence by putting

$$\varepsilon(B) \equiv \varepsilon'(B)F(U\varepsilon'(B)\eta U(B))^{-1}.$$

Proposition 4.2. An adjointness $(F, U, \eta, \varepsilon)$ between categories $\mathscr{A}$ and $\mathscr{B}$ induces an adjoint equivalence between full subcategories $\mathscr{A}_0$ of $\mathscr{A}$ and $\mathscr{B}_0$ of $\mathscr{B}$, where

$$\mathscr{A}_0 \equiv \text{Fix}\,\eta \equiv \{A \in \mathscr{A} \mid \eta(A) \text{ is an isomorphism}\},$$
$$\mathscr{B}_0 \equiv \text{Fix}\,\varepsilon \equiv \{B \in \mathscr{B} \mid \varepsilon(B) \text{ is an isomorphism}\}.$$

Moreover, ηU is an isomorphism if and only if εF is.

The significance of the last statement is this: if ηU is an isomorphism, $\mathscr{B}_0$ becomes a reflective subcategory of $\mathscr{B}$; if εF is an isomorphism $\mathscr{A}_0$ becomes a coreflective subcategory of $\mathscr{A}$. (See Definition 3.2′, 'coreflective' being the dual of 'reflective'.)

Proof. Only the last statement requires proof. It is a consequence of the following.

Lemma 4.3. Given an adjointness $(F, U, \eta, \varepsilon)$ between categories $\mathscr{A}$ and $\mathscr{B}$, the following statements are equivalent:

(1) $\eta UF = UF\eta$,
(2) ηU is an isomorphism,
(3) $\varepsilon FU = FU\varepsilon$,
(4) εF is an isomorphism.

Proof. We show that (1)⇒(2)⇒(3)⇒(4)⇒(1).

(1)⇒(2). Suppose for the moment that $\eta(A)$ has a left inverse g, we claim that, in the presence of (1), g is also a right inverse. For

$$\begin{aligned}\eta(A)g &= UF(g)\eta UF(A) \text{ by naturality of } \eta \\ &= UF(g)UF\eta(A) \text{ by (1)} \\ &= UF(g\eta(A)) = UF(1_A) = 1_{UF(A)}.\end{aligned}$$

Now, by Definition 3.1, $\eta U(B)$ has a left inverse $U\varepsilon(B)$, hence $\eta U(B)$ is an isomorphism, which proves (2).

(2)⇒(3). Assume that $\eta U(B)$ is an isomorphism, then its inverse is $U\varepsilon(B)$, by Definition 3.1. Hence

$$\begin{aligned}\varepsilon FU(B) &= \varepsilon FU(B)F(1_{U(B)}) \\ &= \varepsilon FU(B)F(\eta U(B)U\varepsilon(B)) \\ &= \varepsilon FU(B)F\eta U(B)FU\varepsilon(B) \\ &= 1_{FU(B)}FU\varepsilon(B) \text{ by Definition 3.1} \\ &= FU\varepsilon(B).\end{aligned}$$

(3)⇒(4). This is proved exactly like (1)⇒(2). In fact, we may quote (1)⇒(2), since there is an adjointness between $\mathscr{B}^{\text{op}}$ and $\mathscr{A}^{\text{op}}$.

(4)⇒(1). This is proved like (2)⇒(3) or by duality quoting (2)⇒(3).

Examples of Proposition 4.2 abound in mathematics. The main problem is usually the identification of $\mathscr{A}_0$ and $\mathscr{B}_0$. The following examples require some knowledge of mathematics that has not been developed in this book. (The same will be true for exercises 2 and 3 below.)

Example A1. Let $\mathscr{A}$ be the category of Abelian groups and $\mathscr{B}$ the opposite of the category of topological Abelian groups. Let K be the compact group of the reals modulo the integers: $K \equiv \mathbb{R}/\mathbb{Z}$. For any abstract Abelian group A, define $F(A)$ as the group of all homomorphisms of A into K, with the topology induced by K. For any topological Abelian group B, define $U(B)$ as the group of all continuous homomorphisms of B into K. Then U and F are easily seen to be the object parts of a pair of adjoint functors. Here $\mathscr{A}_0$ is $\mathscr{A}$, while $\mathscr{B}_0$ is the opposite of the category of compact Abelian groups. The 'unity of opposites' asserts the well-known Pontrjagin duality between abstract and compact Abelian groups. The last statement of Proposition 4.2 tells us that the compact Abelian groups form a reflective subcategory of the category of all topological Abelian groups.

Example A2. Let $\mathscr{A}$ be the category of rings and $\mathscr{B}$ the opposite of the category of topological spaces. For any ring A, $F(A)$ is the topological space of homomorphisms of A into $\mathbb{Z}/(2)$, the ring of integers modulo 2, the topology being induced by the discrete topology of $\mathbb{Z}/(2)$. For any topological space B, $U(B)$ is the ring of continuous functions of B into $\mathbb{Z}/(2)$ (with the discrete topology), with the ring structure inherited by that of $\mathbb{Z}/(2)$. Here $\mathscr{A}_0$ is the category of Boolean rings and $\mathscr{B}_0$ is the opposite of the category of zero-dimensional compact Hausdorff spaces. The 'unity of opposites' asserts the well-known Stone duality. Both $\mathscr{A}_0$ and $\mathscr{B}_0^{op}$ are full reflective subcategories.

We summarize the 'unity of opposites' principle in another slogan. (The reader will have noticed that a *duality* between categories $\mathscr{A}$ and $\mathscr{B}$ is nothing but an equivalence between $\mathscr{A}$ and $\mathscr{B}^{op}$.)

Slogan V. Many equivalence and duality theorems in mathematics arise as an equivalence of fixed subcategories induced by a pair of adjoint functors.

Exercises

1. Prove the statement following Definition 4.1 that every equivalence gives rise to an adjoint equivalence. (Hint: first show that $\eta UF = UF\eta$.)

2. Give a presentation of the well-known Gelfand duality between commutative C*-algebras and compact Hausdorff spaces in a manner similar to Example A2. (Let $\mathscr{A}$ be the category of commutative Banach algebras.)
3. If $\mathscr{A}$ is the category of presheaves on a topological space X and $\mathscr{B}$ is the category of spaces over X, show that there is a pair of adjoint functors between $\mathscr{A}$ and $\mathscr{B}$ which induces an equivalence between sheaves and local homeomorphisms. (See also Part II, Theorem 10.3.)
4. Prove that $U: \mathscr{B} \to \mathscr{A}$ is (half of) an equivalence if and only if it is full and faithful and every object of $\mathscr{A}$ is isomorphic to one of the form $U(B)$, for some object B to $\mathscr{B}$.

5 Limits in categories

In this section we shall study limits in categories. They contain as special cases many important constructions, for example products, equalizers and pullbacks, as well as their duals. Moreover, they serve as an illustration of Slogan IV. We begin with the following special case.

Definition 5.1. An object T of a category $\mathscr{A}$ is said to be a *terminal* object if for each object A of $\mathscr{A}$ there is a unique arrow $\bigcirc_A: A \to T$. (Later, we shall usually write 1 for T.)

We note that the uniqueness of $\bigcirc_A$ may be expressed equationally by saying that, for all arrows $h: A \to T, h = \bigcirc_A$.

It is easily seen that T is unique up to isomorphism: if T' is another terminal object, then $T' \cong T$. Hence, one often speaks of *the* terminal object. For example, in the category of sets, any one element set $\{*\}$ is terminal and, in the category of groups, any one element group is terminal. A terminal object in $\mathscr{A}^{op}$ is also called an *initial* object in $\mathscr{A}$. In **Sets**, the only initial object is the empty set $\varnothing$, while, in the category of groups, any terminal object is also initial.

As an illustration of Slogan IV, we note that to say that $\mathscr{A}$ has a terminal (respectively initial) object is the same as saying that the functor $\bigcirc_{\mathscr{A}}: \mathscr{A} \to \mathbf{1}$ has a right (respectively left) adjoint.

Definition 5.2. Given a set I and a family $\{A_i | i \in I\}$ of objects in a category $\mathscr{A}$, their *product* is given by an object P and a family of *projections* $\{p_i: P \to A_i | i \in I\}$ with the following universal property: given any object Q and any family of arrows $\{q_i: Q \to A_i | i \in I\}$, there is a unique arrow $f: Q \to P$ such that $p_i f = q_i$ for all $i \in I$.
We may also say that the family $\{p_i: P \to A_i | i \in I\}$ is a terminal object in the

category of all families $\{q_i : Q \to A_i \mid i \in I\}$ (with appropriate arrows).

It is easily seen that the object P is unique up to isomorphism. Hence, one speaks of *the* product. It is often denoted by $\prod_{i \in I} A_i$. In the category of sets, products are 'cartesian' products. In many concrete categories, products are constructed on the underlying sets with an obvious induced structure. This is true for the categories of monoids, groups, rings etc., in fact all 'algebraic' categories (that is, varieties of universal algebras), as well as for the categories of posets and topological spaces.

A product in $\mathscr{A}^{op}$ is also called a *coproduct* in $\mathscr{A}$. There is no one preferred name for coproducts in the literature; in **Sets**, coproducts are disjoint unions, while, in the category of groups, they are free products.

What if I is the empty set? Then the universal property asserts that, for each object Q, there is a unique arrow $Q \to P$, in other words, that P is a terminal object.

Again we have an illustration of Slogan IV: to say that all I-indexed families in $\mathscr{A}$ have products (respectively coproducts) is the same as saying that the functor $\mathscr{A} \to \mathscr{A}^I$ which sends an object A of $\mathscr{A}$ onto the constant family $\{A \mid i \in I\}$ has a right (respectively left) adjoint.

It may be worth looking at the product of two objects A and B of $\mathscr{A}$ in some detail. It is given by an object $A \times B$ with projections $\pi_{A,B} : A \times B \to A$ and $\pi'_{A,B} : A \times B \to B$ such that, for all arrows $f : C \to A$ and $g : C \to B$, there is a unique arrow $\langle f, g \rangle : C \to A \times B$ satisfying the equations:

$$\pi_{A,B}\langle f, g \rangle = f, \quad \pi'_{A,B}\langle f, g \rangle = g.$$

Note that the uniqueness of $\langle f, g \rangle$ may also be expressed by an equation, namely:

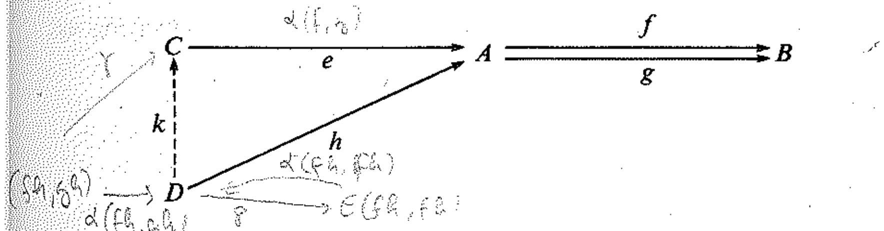
$$\langle \pi_{A,B} h, \pi'_{A,B} h \rangle = h,$$

for all $h : C \to A \times B$.

Evidently, the defining property of $A \times B$ establishes a bijection between pairs of arrows $(C \to A, C \to B)$ and arrows $C \to A \times B$. To say that all such products exist is the same as saying that the diagonal functor $\Delta : \mathscr{A} \to \mathscr{A} \times \mathscr{A}$ has a right adjoint. Dually, all binary coproducts exist if and only if Δ has a left adjoint.

Definition 5.3. A pair of arrows $f, g : A \rightrightarrows B$ is said to have an *equalizer* $e : C \to A$ provided $fe = ge$ and, for all $h : D \to A$ such that $fh = gh$, there is a unique arrow $k : D \to C$ satisfying $ek = h$. Another way of expressing this is to say that $e : C \to A$ is terminal in the category of all arrows $h : D \to A$ such that $fh = gh$.

It is easily seen that the equalizing object C is unique up to isomorphism.



In the category of sets or groups, one may take $C \equiv \{a \in A \mid f(a) = g(a)\}$ and $e : C \to A$ as the inclusion. As is the case for products, equalizers in many concrete categories are formed on the underlying sets. An equalizer in $\mathscr{A}^{op}$ is also called a *coequalizer* in $\mathscr{A}$. In **Sets**, the coequalizer of two mappings $f, g : B \rightrightarrows A$ is given by $e : A \to C$, where C is obtained from A by identifying all elements $f(b)$ and $g(b)$ with $b \in B$, and where e is the obvious surjection. (More precisely, $C = A/\equiv$, where $\equiv$ is the smallest equivalence relation on A such that $f(b) \equiv g(b)$ for all $b \in B$.) In the category of groups, the coequalizer of two homomorphisms $f, g : B \rightrightarrows A$ is obtained similarly from a suitable congruence relation on A (or normal subgroup of A).

While it was evident how finite products could be presented equationally, it is by no means obvious how this can be done for equalizers. The following discussion is our version of Burroni's pioneering ideas.

With any diagram $A \overset{f}{\underset{g}{\rightrightarrows}} B$ we associate another diagram $E(f, g) \xrightarrow{\alpha(f, g)} A$ which is to serve as its equalizer. Clearly, we must stipulate the equation

(B1) $\quad f\alpha(f, g) = g\alpha(f, g).$

Next, let us consider the universal property of $\alpha(f, g)$. Given an arrow $h : D \to A$ such that $fh = gh$, we seek a unique arrow $\beta(f, g, h) : D \to E(f, g)$ such that

(*) $\quad \alpha(f, g)\beta(f, g, h) = h.$

While (*) is an equation, it depends on the condition $fh = gh$, which we would like to get rid of. We shall consider two special cases of $\beta(f, g, h)$ in which the condition $fh = gh$ is automatically satisfied.

First special case: consider any arrow $h : D \to A$, then surely

$$fh\alpha(fh, gh) = gh\alpha(fh, gh).$$

Hence we stipulate an arrow $\gamma(f, g, h)$ $(\equiv \beta(f, g, h\alpha(fh, gh)))$: $E(fh, gh) \to E(f, g)$ satisfying as a special case of (*):

(B2) $\quad \alpha(f, g)\gamma(f, g, h) = h\alpha(fh, gh).$

Second special case: consider any arrow $f : A \to B$, then surely $f1_A = f1_A$.

Hence we stipulate an arrow $\delta(f)$ ($\equiv \beta(f,f,1_A)$): $A \to E(f,f)$ satisfying as a special case of (*):

(B3) $\alpha(f,f)\delta(f) = 1_A$.

From the two special cases we can define $\beta(f,g,h)$ in general: Assuming $fh = gh$, put

(**) $\beta(f,g,h) \equiv \gamma(f,g,h)\delta(fh)$.

Then

$$\alpha(f,g)\beta(f,g,h) = \alpha(f,g)\gamma(f,g,h)\delta(fh) = h\alpha(fh,gh)\delta(fh),$$

by (B2). As it so happens that $fh = gh$, this becomes equal to

$$h1_D = h$$

by (B3), and so we recapture (*).

It remains to express the uniqueness of $\beta(f,g,h)$ equationally. So suppose that $\alpha(f,g)k = h$, we want this to imply that $k = \beta(f,g,h)$. This is evidently done by

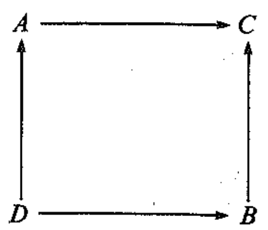
(B4) $\beta(f,g,\alpha(f,g)k) = k$.

Here β can be eliminated in favour of γ and δ using (**).

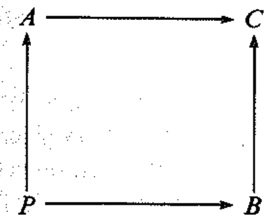
We summarize the preceding discussion of equalizers as follows.

Proposition 5.4. (Burroni). Equalizers for all pairs of arrows $f, g: A \rightrightarrows B$ are given by the following data: an arrow $\alpha(f,g): E(f,g) \to A$ for each such pair, a family of arrows $\gamma(f,g,h): E(fh,gh) \to E(f,g)$ one for each $h: D \to A$, and an arrow $\delta(f): A \to E(f,f)$ satisfying (B1) to (B4) (with β eliminated from (B4) by (**)).

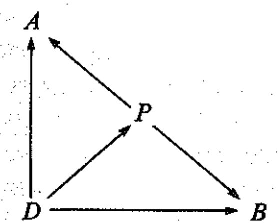
Definition 5.5. A *pullback* of a diagram $A \to C \leftarrow B$ is given by a diagram $A \leftarrow P \to B$ which is terminal in the category of all diagrams $A \leftarrow D \to B$ such that



commutes. In other words,



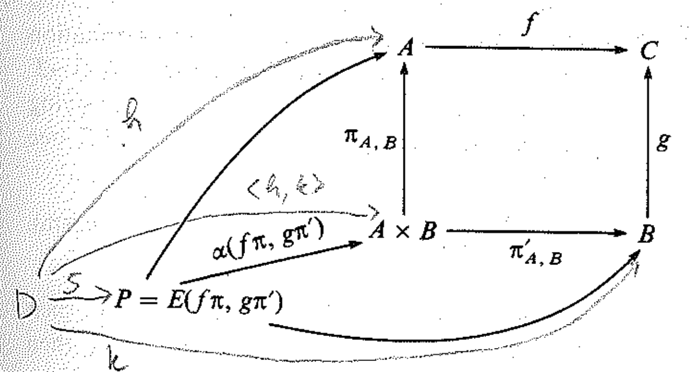
commutes and, for any other commutative square as above, there is a unique arrow $D \to P$ such that the two triangles



commute.

It is easily seen that P is unique up to isomorphism. A pullback in $\mathscr{A}^{op}$ is called a *pushout* in $\mathscr{A}$. In a category with a terminal object T, binary products are special cases of pullbacks, namely when $C \equiv T$. Instead of describing pullbacks in other special categories, we shall show how, in general, they may be constructed from products and equalizers.

Proposition 5.6. If a category has binary products and equalizers, pullbacks may be constructed as follows:



Proof. Note that $f\pi\alpha(f\pi, g\pi') = g\pi'\alpha(f\pi, g\pi')$, by (B1). Suppose $h: D \to A$ and $k: D \to B$ are such that $fh = gk$. Then there is a unique arrow $\langle h,k\rangle: D \to A \times B$ such that $\pi\langle h,k\rangle = h$ and $\pi'\langle h,k\rangle = k$, hence a unique

$P \xrightarrow{\alpha} A\times B$ equalizer for $f\pi$, $g\pi'$

arrow $s: D \to P$ such that $\pi\alpha(f\pi, g\pi)s = h$ and $\pi'\alpha(f\pi, g\pi)s = k$, that is, $\alpha(f\pi, g\pi)s = \langle h, k \rangle$.

Definition 5.7. Let there be given a category $\mathscr{I}$ (the *index* category) and a functor $\Gamma: \mathscr{I} \to \mathscr{A}$ (called an $\mathscr{I}$*-diagram*). A *limit* of Γ is given by a terminal object in the category of all pairs (A, t) with A an object of $\mathscr{A}$ and $t: K(A) \to \Gamma$ a natural transformation, where $K(A): \mathscr{I} \to \mathscr{A}$ is the functor with *constant* value A. In other words, $(A_0, t_0: K(A_0) \to \Gamma)$ is a limit of Γ if for all $(A, t: K(A) \to \Gamma)$ there is unique $f: A \to A_0$ such that $t_0(I)f = t(I)$ for all objects I of $\mathscr{I}$.

It is easily seen that A_0 is unique up to isomorphism. Special cases of limits are products ($\mathscr{I}$ discrete), equalizers ($\mathscr{I}$ is $\cdot\rightrightarrows\cdot$) and pullbacks ($\mathscr{I}$ is $\genfrac{}{}{0pt}{}{\cdot}{\cdot}\!\rightrightarrows\!\cdot$). Limits may be constructed from products and equalizers as are pullbacks (Proposition 5.6). Limits in $\mathscr{A}^{\mathrm{op}}$ are also called *colimits* in $\mathscr{A}$. If $\mathscr{I}$ is a directed poset, limits are usually called *inverse* or *projective* limits, while colimits are called *direct* or *inductive* limits. The limit of Γ (or rather the object A_0) is sometimes denoted by $\varprojlim \Gamma$ and the colimit by $\varinjlim \Gamma$.

The following connection between limits and adjoint functors illustrates Slogan IV.

Proposition 5.8. To say for given categories $\mathscr{I}$ and $\mathscr{A}$ that every $\mathscr{I}$-diagram $\Gamma: \mathscr{I} \to \mathscr{A}$ has a limit (respectively colimit) is equivalent to saying that the *constancy* functor $K: \mathscr{A} \to \mathscr{A}^{\mathscr{I}}$, which associates to every object A of $\mathscr{A}$ the functor $K(A): \mathscr{I} \to \mathscr{A}$ with constant value A, has a right adjoint (respectively left adjoint).

Proof. One way of asserting that K has a right adjoint $L: \mathscr{A}^{\mathscr{I}} \to \mathscr{A}$ is by the solution to the universal mapping problem (dualize Definition 3.2): for each object Γ of $\mathscr{A}^{\mathscr{I}}$ there is an object $L(\Gamma)$ and a natural transformation $\varepsilon(\Gamma): KL(\Gamma) \to \Gamma$ such that, for every natural transformation $t: K(A) \to \Gamma$ there is a unique natural transformation $t^*: A \to L(\Gamma)$ satisfying $\varepsilon(\Gamma)K(t^*) = t$. But this says precisely that $(L(\Gamma), \varepsilon(\Gamma))$ is a limit of Γ (see Definition 4.7).

Many functors occurring in nature preserve limits (up to isomorphism). We shall mention two examples.

Proposition 5.9. If A is an object of the locally small category $\mathscr{A}$, then $\mathrm{Hom}(A, -): \mathscr{A} \to \mathbf{Sets}$ preserves limits: if $\Gamma: \mathscr{I} \to \mathscr{A}$ has limit A_0 then $\mathrm{Hom}(A, \Gamma(-)): \mathscr{I} \to \mathbf{Sets}$ has limit $\mathrm{Hom}(A, A_0)$.

Proof. Write $h^A \equiv \mathrm{Hom}(A, -)$ and assume that $(A_0, t_0: K(A_0) \to \Gamma)$ is terminal in the category of all pairs $(A, t: K(A) \to \Gamma)$. We assert that $(h^A(A_0),\ h^A t_0: h^A K(A_0) \to h^A \Gamma)$ is terminal in the category of all pairs

$(X, \tau: K(X) \to h^A \Gamma)$, X being a set. (Note that $h^A K(A_0) = K(h^A(A_0))$.) In other words, we claim that there is a unique mapping $\psi: X \to h^A(A_0)$ such that $(h^A t_0) \circ K(\psi) = \tau$. To see what this last equation means, apply it to any object I of $\mathscr{I}$, then it asserts

$$\mathrm{Hom}(1_A, t_0(I))\psi = \tau(I).$$

Again, applying this equation to any $x \in X$, we obtain

$$t_0(I)\psi(x) = \tau(I)(x).$$

If $t_x: K(A) \to \Gamma$ is defined by $t_x(I) \equiv \tau(I)(x)$, we see that this means

$$t_0 \circ K(\psi(x)) = t_0.$$

The existence of a unique $\psi(x): A_0 \to A$ with this property is assured by the fact that (A_0, t_0) was terminal.

Proposition 5.10. If $F: \mathscr{A} \to \mathscr{B}$ is left adjoint to $U: \mathscr{B} \to \mathscr{A}$, then U preserves limits and F preserves colimits.

Proof. If $\mathscr{A}$ and $\mathscr{B}$ are locally small, this is an easy corollary of Proposition 5.9. However, one may just as well prove the result directly, without assuming local smallness, and we shall do so for U.

While it is easy to give a precise argument as in the proof of Proposition 5.9, the reader may find the following sketch more intuitive.

Let $\mathscr{C}$ (respectively $\mathscr{D}$) be the full subcategory of $\mathscr{A}^{\mathscr{I}}$ (respectively $\mathscr{B}^{\mathscr{I}}$) consisting of those $\mathscr{I}$-diagrams which have limits. Evidently, $\mathscr{C}$ contains all constant $\mathscr{I}$-diagrams $K_{\mathscr{A}}(A)$, with A in $\mathscr{A}$, such that $K_{\mathscr{A}}(A)(I) = A$ for all I in $\mathscr{I}$. Hence we may factor the constancy functor $K_{\mathscr{A}}: \mathscr{A} \to \mathscr{A}^{\mathscr{I}}$ through $K'_{\mathscr{A}}: \mathscr{A} \to \mathscr{C}$. As in Proposition 5.8, we may regard $\varprojlim_{\mathscr{A}}$ as right adjoint to $K'_{\mathscr{A}}$. Now $F^{\mathscr{I}}: \mathscr{A}^{\mathscr{I}} \to \mathscr{B}^{\mathscr{I}}$ (respectively $U^{\mathscr{I}}: \mathscr{B}^{\mathscr{I}} \to \mathscr{A}^{\mathscr{I}}$) factors through $F': \mathscr{C} \to \mathscr{D}$ (respectively $U': \mathscr{D} \to \mathscr{C}$) and U' is right adjoint to F'. Then, clearly, $F'K'_{\mathscr{A}} = K'_{\mathscr{B}}F$.

$$\begin{array}{ccc} \mathscr{A} & \xrightarrow{F} & \mathscr{B} \\ {\scriptstyle K'_{\mathscr{A}}}\downarrow & & \downarrow{\scriptstyle K'_{\mathscr{B}}} \\ \mathscr{C} & \xrightarrow{F'} & \mathscr{D} \\ {\scriptstyle \text{inclusion}}\downarrow & & \downarrow{\scriptstyle \text{inclusion}} \\ \mathscr{A}^{\mathscr{I}} & \xrightarrow{F^{\mathscr{I}}} & \mathscr{B}^{\mathscr{I}} \end{array}$$

Taking right adjoints, we obtain, in view of Proposition 3.5, $\varprojlim_{\mathscr{A}} U' \cong U \varprojlim_{\mathscr{B}}$. Applying both sides to any diagram $\Delta: \mathscr{I} \to \mathscr{B}$ and noting that $U'(\Delta) = U\Delta$, we finally obtain $\varprojlim (U\Delta) \cong U(\varprojlim(\Delta))$.

Definition 5.11. A category $\mathscr{A}$ is said to be *complete* (cocomplete) if it has all limits (colimits) of diagrams $\Gamma: \mathscr{I} \to \mathscr{A}$, $\mathscr{I}$ being small. This means that products (coproducts) and equalizers (coequalizers) exist.

Assuming completeness of $\mathscr{A}$ or $\mathscr{B}$ one can prove a kind of converse of Proposition 5.9 and of 5.10. For example, if $U: \mathscr{B} \to \mathscr{A}$ preserves limits and $\mathscr{A}$ is complete, one can construct a left adjoint $F: \mathscr{A} \to \mathscr{B}$, as in Exercise 1 of Section 3, provided a certain 'solution set condition' holds; this is the content of Freyd's Adjoint Functor Theorem. These converse results will be brought out in the exercises; they depend on the following lemma, the proof of which is a bit tricky.

Lemma 5.12. If $\mathscr{A}$ is complete, then $\mathscr{A}$ has an initial object if and only if it has a small *pre-initial* full subcategory $\mathscr{C}$, that is to say, for any object A of $\mathscr{A}$ there is an object C of $\mathscr{C}$ and an arrow $f: C \to A$ in $\mathscr{A}$.

Proof. The necessity of the condition is obvious. To prove its sufficiency, let $(A_0, u: K(A_0) \to \Gamma)$ be the limit of the inclusion functor $\Gamma: \mathscr{C} \to \mathscr{A}$. In particular, for each object C of $\mathscr{C}$ there is an arrow $u(C): A_0 \to C$. Take any object A of $\mathscr{A}$, then, by assumption, we can find C in $\mathscr{C}$ and an arrow $f: C \to A$, hence an arrow $fu(C): A_0 \to A$. It remains to show that there is only one arrow $A_0 \to A$.

Suppose we have two arrows $g, h: A_0 \rightrightarrows A$ and let $k: K \to A_0$ be their equalizer. It will follow that $g = h$ if we can show that k has a right inverse. By assumption, there exists C' in $\mathscr{C}$ and an arrow $f': C' \to K$. It will suffice to show that $kf'u(C') = 1_{A_0}$.

Now, for any object C of $\mathscr{C}$,

$$u(C)kf'u(C') = u(C),$$

by naturality of u and because $u(C)kf': C' \to C$ is an arrow in the full subcategory $\mathscr{C}$. Since (A_0, u) is the limit of the inclusion $\mathscr{C} \to \mathscr{A}$, there exists a unique arrow $e: A_0 \to A_0$ such that $u(C)e = u(C)$. Hence

$$kf'u(C') = e = 1_{A_0},$$

and our argument is complete.

Exercises

1. Prove that limits can be constructed from products and equalizers, generalizing the proof of Proposition 5.6.

2. Deduce from Proposition 5.10 that, in the propositional calculus regarded as a preordered set (see Exercise 4 of Section 3), the distributive law holds: $p \wedge (a \vee b) \cong (p \wedge a) \vee (p \wedge b)$.

3. Given two functors $F, G: \mathscr{A} \rightrightarrows \mathscr{B}$, let $(F; G)$ be the category whose objects are pairs $(A, b: F(A) \to G(A))$, A any object of $\mathscr{A}$, and whose arrows $(A, b) \to (A', b')$ are arrows $a: A \to A'$ in $\mathscr{A}$, such that $G(a)b = b'F(a)$. Assuming that $\mathscr{A}$ is complete and that G preserves limits, show that $(F; G)$ has an initial object if and only if it has a small pre-initial full subcategory. (Hint: Use Proposition 5.12.)

4. If $\mathscr{A}$ is locally small, a functor $U: \mathscr{A} \to \mathbf{Sets}$ is said to be *representable* if $U \cong \mathrm{Hom}(A, -)$ for some objects A of $\mathscr{A}$. Show that U is representable if and only if the category $(K(\{*\}); U)$ has a small pre-initial full subcategory. (Hint: Use Exercise 3 with $\mathscr{B} = \mathbf{Sets}$.)

5. Let $\mathscr{A}$ be a complete category. Show that a functor $U: \mathscr{A} \to \mathscr{B}$ has a left adjoint if and only if U preserves limits and, for each object B of $\mathscr{B}$, the category $(K(B); U)$ has a small pre-initial full sub-category.

6. Let $\mathscr{A}$ be a complete category. Show that a functor $\Gamma: \mathscr{I} \to \mathscr{A}$ has a colimit if and only if the category $(K(\Gamma); K)$ has a small pre-initial full subcategory. (Here the first K denotes the constancy functor $\mathscr{A}^{\mathscr{I}} \to (\mathscr{A}^{\mathscr{I}})^{\mathscr{A}}$, while the second K denotes the constancy functor $\mathscr{A} \to \mathscr{A}^{\mathscr{I}}$.)

7. Given a small category $\mathscr{A}$ and any functor $F: \mathscr{A}^{\mathrm{op}} \to \mathbf{Sets}$, show that F is a colimit of representable functors as follows. Let $\mathscr{I}_F$ be the category whose objects are pairs (A, t), A an object of $\mathscr{A}$ and $t: \mathrm{Hom}_{\mathscr{A}}(-, A) \to F$ a natural transformation, and whose arrows $(A, t) \to (A', t')$ are arrows $a: A \to A'$ in $\mathscr{A}$ such that $t' \circ \mathrm{Hom}_{\mathscr{A}}(-, a) = t$. Then F is the colimit of the functor $\Gamma_F: \mathscr{I}_F \to \mathbf{Sets}^{\mathscr{A}^{\mathrm{op}}}$ obtained by composing the Yoneda embedding $\mathscr{A} \to \mathbf{Sets}^{\mathscr{A}^{\mathrm{op}}}$ with the obvious forgetful functor $\mathscr{I}_F \to \mathscr{A}$. (The associated natural transformation $t_0: \Gamma_F \to K(F)$ is defined by $t_0(A, t) \equiv t$.)

6 Triples

We recall that a *closure operation* on a preordered set $\mathscr{A} = (|\mathscr{A}|, \leqslant)$ is a mapping $T: |\mathscr{A}| \to |\mathscr{A}|$ with the following properties:

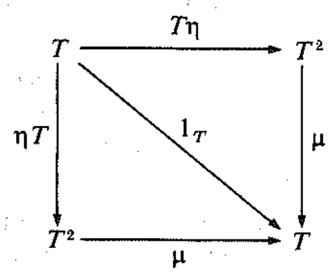
$$\frac{A \leqslant B}{T(A) \leqslant T(B)}, \quad A \leqslant T(A), \quad TT(A) \leqslant T(A),$$

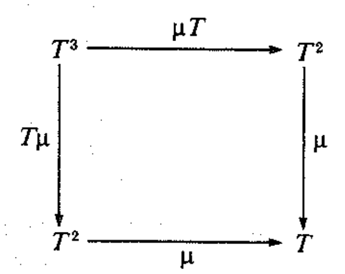
for all elements A, B of $|\mathscr{A}|$. The first of these says, of course, that T is order preserving. This notion has been generalized from preordered sets to arbitrary categories and is then called a 'standard construction', 'triple' or 'monad'. Reluctantly, we choose the second term, as it appears to be the most widely used.

Definition 6.1. A *triple* (T, η, μ) on a category $\mathscr{A}$ consists of a functor T: $\mathscr{A} \to \mathscr{A}$ and natural transformations $\eta: 1_{\mathscr{A}} \to T$ and $\mu: T^2 \to T$ satisfying the equations

$$\mu \circ T\eta = 1_T = \mu \circ \eta T, \quad \mu \circ \mu T = \mu \circ T\mu.$$

These equations are sometimes called the *unity laws* and *associative law* respectively and are illustrated by the following commutative diagrams:





The reader will recall how natural transformations are composed (see Example C8); for example, the associative law asserts that, for every object A of $\mathscr{A}$,

$$\mu(A)\mu(T(A)) = \mu(A)T(\mu(A)).$$

Proposition 6.2. (Huber). If $F: \mathscr{A} \to \mathscr{B}$ is left adjoint to the functor $U: \mathscr{B} \to \mathscr{A}$ with adjunctions $\eta: 1_{\mathscr{A}} \to UF$ and $\varepsilon: FU \to 1_{\mathscr{B}}$, then $(UF, \eta, U\varepsilon F)$ is a triple on $\mathscr{A}$.

Proof. For example, let us prove one of the unity laws:

$$\mu \circ T\eta = U\varepsilon F \circ UF\eta = U(\varepsilon F \circ F\eta) = U1_F = 1_{UF},$$

by Definition 3.1, and since

$$(U1_F)(A) = U(1_{F(A)}) = 1_{U(F(A))} = 1_{UF}(A).$$

We leave the proofs of the other two laws to the reader.

We shall see that the converse of this proposition is also true; but first we shall look at a number of examples of triples, which, on the face of it, do not seem to arise from a pair of adjoint functors.

Example T1. Let there be given a monoid $\mathscr{M} = (M, 1, \cdot)$. For each set A define the set $T(A) \equiv M \times A$ and the mappings

$$\eta(A): A \to M \times A, \quad \mu(A): M \times (M \times A) \to M \times A$$
$$a \mapsto (1, a) \qquad (m, (m', a)) \mapsto (m \cdot m', a).$$

One easily makes T into a functor **Sets** $\to$ **Sets** and checks that η and μ are

natural transformations. Moreover, one obtains a triple (T, η, μ) on **Sets**, the unity laws and associative law here following from the equations

$$m \cdot 1 = m = 1 \cdot m, \quad (m \cdot m') \cdot m'' = m \cdot (m' \cdot m'')$$

for all m, m' and $m'' \in M$, which will explain their names.

Example T2. Let $T \equiv P$ be the covariant power set functor **Sets** $\to$ **Sets**, that is, for any set A,

$$P(A) \equiv \{X \mid X \subseteq A\}$$

and, for any mapping $f: A \to B$, and any subset $X \subseteq A$,

$$P(f)(X) = \{f(x) \mid x \in X\}.$$

Furthermore, let the natural transformations η and μ be given by the mappings $\eta(A): A \to P(A)$ and $\mu(A): P(P(A)) \to P(A)$ defined by

$$\eta(A)(a) \equiv \{a\}, \quad \mu(A)(\mathscr{X}) \equiv \bigcup \mathscr{X} \equiv \bigcup_{X \in \mathscr{X}} X,$$

for any set A, any element $a \in A$ and any set $\mathscr{X}$ of subsets of A. The reader is invited to show that (T, η, μ) is a triple by verifying the unity and associative laws in this case.

We now return to the question: does every triple on $\mathscr{A}$ arise from a pair of adjoint functors $\mathscr{A} \xrightarrow{F} \mathscr{B} \xrightarrow{U} \mathscr{A}$ as in Proposition 6.2? The answer is 'yes', but the category $\mathscr{B}$ is not unique. In fact, we shall present two extremes for the construction of $\mathscr{B}$.

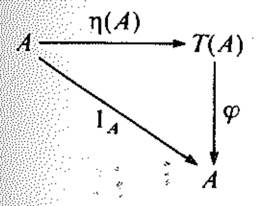
Definition 6.3. Given a triple (T, η, μ) on a category $\mathscr{A}$, the *Eilenberg–Moore* category $\mathscr{A}^T$ of the triple is defined as follows. Its objects, called *algebras*, are pairs (A, φ), where $\varphi: T(A) \to A$ is an arrow of $\mathscr{A}$ satisfying the equations

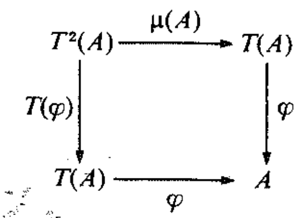
$$\varphi\eta(A) = 1_A, \quad \varphi\mu(A) = \varphi T(\varphi)$$

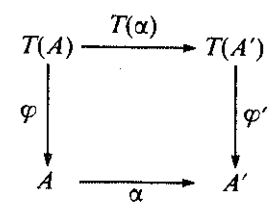
for all objects A of $\mathscr{A}$. Its arrows, called *homomorphisms*, $(A, \varphi) \to (A', \varphi')$ are arrows $\alpha: A \to A'$ of $\mathscr{A}$ satisfying the equation

$$\varphi' T(\alpha) = \alpha\varphi.$$

These equations are illustrated by the following commutative diagrams:







Example T1 (continued). An element of $T(A) \equiv M \times A$ is a pair (m, a) with $m \in M$ and $a \in A$. One usually writes $ma \equiv \varphi(m, a)$. The equations of an algebra then read

$$1a = a, \quad (m \cdot m')a = m(m'a),$$

for all $a \in A, m$ and $m' \in M$. In other words, an algebra is an $\mathscr{M}$-set (see Example F3 in Section 1). The equation satisfied by a homomorphism reads

$$\alpha(ma) = m\alpha(a),$$

for all $a \in A$, so we recapture the usual homomorphisms of $\mathscr{M}$-sets (see Proposition 2.2).

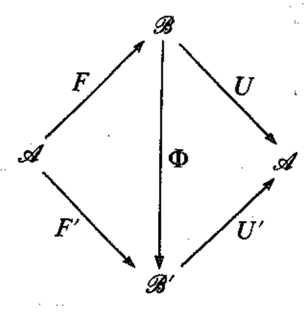
Example T2 (continued). The algebras of the power set triple on **Sets** are sup-complete (hence inf-complete) lattices and the homomorphisms are sup-preserving (hence also order preserving) mappings.

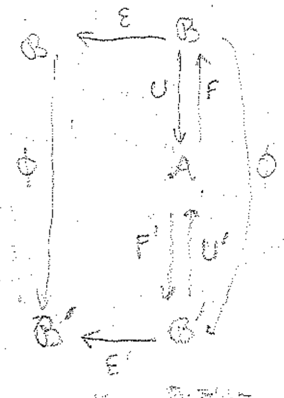
In view of these examples and many others like them, we enunciate our final slogan.

Slogan VI. Many categories of interest are the Eilenberg–Moore categories of triples on familiar categories.

In both examples above, the familiar category is **Sets**, but in Exercise 2 below it is **Ab**, the category of abelian groups. Categories, on the other hand, may be viewed as algebras over **Grph**, the category of graphs.

Definition 6.4. Given a triple (T, η, μ) on a category $\mathscr{A}$, by a *resolution* $(\mathscr{B}, U, F, \varepsilon)$ of this triple we mean a category $\mathscr{B}$ and a pair of adjoint functors $\mathscr{A} \xrightarrow{F} \mathscr{B} \xrightarrow{U} \mathscr{A}$ such that $UF = T$ with adjunctions η (as given) and ε such that $U\varepsilon F = \mu$ (as in Proposition 6.2). The resolutions of the given triple form a category whose arrows $\Phi: (\mathscr{B}, U, F, \varepsilon) \to (\mathscr{B}', U', F', \varepsilon')$ are functors $\Phi: \mathscr{B} \to \mathscr{B}'$ such that $\Phi F = F'$, $U'\Phi = U$ and $\Phi\varepsilon = \varepsilon'\Phi$. In particular, the following two triangles commute:





Proposition 6.5. The Eilenberg–Moore category $\mathscr{A}^T$ of the triple (T, η, μ) on $\mathscr{A}$ gives rise to a resolution $(\mathscr{A}^T, U^T, F^T, \varepsilon^T)$, which is a terminal object in the category of all resolutions. Thus, given any resolution $(\mathscr{B}, U, F, \varepsilon)$, there is a unique functor $K^T: \mathscr{B} \to \mathscr{A}^T$, called the *comparison functor*, such that $K^T F = F^T$, $U^T K^T = U$ and $K^T \varepsilon = \varepsilon^T K^T$. Moreover, U^T is faithful.

Proof. (1) We define $U^T: \mathscr{A}^T \to \mathscr{A}$ by

$$U^T(A, \varphi) \equiv A, \quad U^T(\alpha) \equiv \alpha,$$

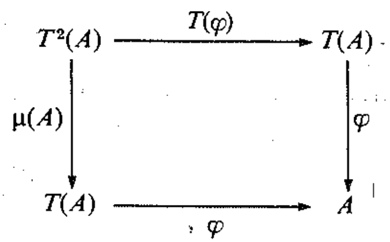
for any algebra (A, φ) and any homomorphism α. Evidently, U^T is faithful.

(2) We define $F^T: \mathscr{A} \to \mathscr{A}^T$ by

$$F^T(A) \equiv (T(A), \mu(A)), \quad F^T(f) \equiv T(f),$$

for any object A and any arrow f of $\mathscr{A}$. It is easily checked that $(T(A), \mu(A))$ is an algebra, that $T(f)$ is a homomorphism and that $U^T F^T = T$.

(3) We define the natural transformation ε^T from $F^T U^T$ to the identity functor on $\mathscr{A}^T$ by its action on the algebra (A, φ) as follows: the homomorphism $\varepsilon^T(A, \varphi) \equiv \varphi$. Indeed, the square



commutes by Definition 6.3. To see that $U^T \varepsilon^T F^T = \mu$, one calculates

$$(U^T \varepsilon^T F^T)(A) = (U^T \varepsilon^T)(T(A), \mu(A)) = U^T(\mu(A)) = \mu(A).$$

We let the reader check that

$$(\varepsilon^T F^T \circ F^T \eta)(A) = A, \quad (U^T \varepsilon^T \circ \eta U^T)(A, \varphi) = (A, \varphi),$$

for any object A of $\mathscr{A}$ and any algebra (A, φ), whence it follows that $(\mathscr{A}^T, U^T, F^T, \varepsilon^T)$ is a resolution of the given triple.

(4) Let $(\mathscr{B}, U, F, \varepsilon)$ be another resolution of the same triple, we shall construct the comparison functor $K^T: \mathscr{B} \to \mathscr{A}^T$ and show that it is the unique functor with the desired properties. For any object B and any arrow g of $\mathscr{B}$, we put

$$K^T(B) \equiv (U(B), U\varepsilon(B)), \quad K^T(g) \equiv U(g).$$

Then surely $U^T K^T = U$; in fact, this result forces the definitions of $K^T(g)$ and of the first component of $K^T(B)$. Moreover, $\varepsilon^T K^T(B) = U\varepsilon(B)$, and this forces the definition of the second component of $K^T(B)$. It remains to check that $K^T F = F^T$. Indeed, for any object A of $\mathscr{A}$,

$$K^T F(A) = (UF(A), U\varepsilon F(A)) = (T(A), \mu(A)) = F^T(A).$$

This completes the proof.

We remark that, in view of Slogan VI, it is of interest to know when the comparison functor is an equivalence of categories. Conditions for this to be the case were found by Beck. Without going into these conditions here, let us only mention that a functor $U: \mathscr{B} \to \mathscr{A}$ is called *tripleable* or *monadic* if it has a left adjoint and if the comparison functor K^T is an equivalence. Examples of tripleable concrete categories $U: \mathscr{B} \to$ **Sets** are all algebraic categories, that is, varieties of universal algebras, and the category of compact Hausdorff spaces.

The category of resolutions of a triple also has an initial object.

Definition 6.6. The *Kleisli* category $\mathscr{A}_T$ of a triple (T, η, μ) on a category $\mathscr{A}$ is defined as follows. Its objects are the same as those of $\mathscr{A}$; however, arrows $A \to A'$ in $\mathscr{A}_T$ are not the same as they would be in $\mathscr{A}$, instead they are arrows $A \to T(A')$ in $\mathscr{A}$. How do we compose arrows $f: A \to T(A')$ and $g: A' \to T(A'')$? Denoting their composition in $\mathscr{A}_T$ by $g * f: A \to T(A'')$ in $\mathscr{A}$, we define

$$g * f \equiv \mu(A'')T(g)f.$$

In particular,

$$f * \eta(A) = \mu(A')T(f)\eta(A) = \mu(A')\eta T(A')f = 1_T(A')f = f$$

and

$$\eta(A') * f = \mu(A')T\eta(A')f = 1_T(A')f = f,$$

hence $\eta(A): A \to T(A)$ serves as the identity arrow $A \to A$ in $\mathscr{A}_T$. We leave it to the reader to check the associativity of composition in $\mathscr{A}_T$.

Example T2 (continued). What is the Kleisli category of the power set triple on **Sets**? An arrow $A \to P(B)$ in **Sets** may be regarded as a multi-valued function from A to B or, equivalently, as a relation between A and B. More precisely, let $f: A \to P(B)$ correspond to $R_f \subseteq A \times B$, where $(a, b) \in R_f$ means $b \in f(a)$. What about the composition of f with $g: B \to P(C)$? According to Definition 6.6,

$$\begin{aligned}(g * f)(a) &\equiv \mu(C)(P(g)(f(a))) \\ &\equiv \bigcup \{g(b) \mid b \in f(a)\},\end{aligned}$$

hence

$$\begin{aligned}(a, c) \in R_{g*f} &\Leftrightarrow c \in (g * f)(a) \\ &\Leftrightarrow \exists_{b \in B}(b \in f(a) \wedge c \in g(b)) \\ &\Leftrightarrow \exists_{b \in B}((a, b) \in R_f \wedge (b, c) \in R_g) \\ &\Leftrightarrow (a, c) \in R_g R_f,\end{aligned}$$

according to one way of defining the 'relative product'. Moreover, the identity arrow 1_A in the Kleisli category is represented by the mapping $\eta(A): A \to P(A)$ in **Sets**, which sends $a \in A$ onto $\{a\} \subseteq A$. Hence $(a, a') \in R_{\eta(A)} \Leftrightarrow a' \in \{a\}$, so $R_{\eta(A)}$ is the identity relation on A. We conclude that the Kleisli category of the power set triple on **Sets** is (isomorphic to) the category whose objects are sets and whose arrows are binary relations.

Proposition 6.7. The Kleisli category $\mathscr{A}_T$ of the triple (T, η, μ) on $\mathscr{A}$ gives rise to a resolution $(\mathscr{A}_T, U_T, F_T, \varepsilon_T)$, which is an initial object in the category of all resolutions. Thus, given any resolution $(\mathscr{B}, U, F, \varepsilon)$, there is a unique functor $K_T: \mathscr{A}_T \to \mathscr{B}$ such that $K_T F_T = F$, $UK_T = U_T$ and $K_T \varepsilon_T = \varepsilon K_T$. Moreover, F_T is bijective on objects.

Proof. (1) We define $U_T: \mathscr{A}_T \to \mathscr{A}$ by

$$U_T(A) \equiv T(A), \quad U_T(f) \equiv \mu(B)T(f),$$

for any object A of $\mathscr{A}_T$, that is of $\mathscr{A}$, and for any arrow $f: A \to B$ in $\mathscr{A}_T$, that is, $f: A \to T(B)$ in $\mathscr{A}$. It is easily verified that U_T is a functor.

(2) We define $F_T: \mathscr{A} \to \mathscr{A}_T$ by

$$F_T(A) \equiv A, \quad F_T(f) \equiv \eta(B)f,$$

for any object A and any arrow $f: A \to B$ in $\mathscr{A}$. Evidently, F_T is bijective on objects and it is easily checked that $U_T F_T = T$ and that F_T is a functor.

(3) We define the natural transformation ε_T from $F_T U_T$ to the identity functor on $\mathscr{A}_T$ by putting $\varepsilon_T(A) \equiv 1_{T(A)}$ in $\mathscr{A}$. To see that $U_T \varepsilon_T F_T = \mu$ one calculates

$$(U_T \varepsilon_T F_T)(A) = (U_T \varepsilon_T)(A) = U_T(1_{T(A)}) = \mu(A).$$

We let the reader check that

$$(\varepsilon_T F_T \circ F_T \eta)(A) = A, \quad (U_T \varepsilon_T \circ \eta U_T)(A) = A,$$

for any object A of $\mathscr{A}$, hence of $\mathscr{A}_T$, whence it follows that $(\mathscr{A}_T, U_T, F_T, \varepsilon_T)$ is a resolution of the given triple.

(4) Let $(\mathscr{B}, U, F, \varepsilon)$ be another resolution of the same triple. We shall construct a functor $K_T: \mathscr{A}_T \to \mathscr{B}$ and show that is the unique functor with the desired properties.

For any object A of $\mathscr{A}_T$ and any arrow $g: A \to A'$ in $\mathscr{A}_T$, that is, $g: A \to T(A')$ in $\mathscr{A}$, we put

$$K_T(A) \equiv F(A), \quad K_T(g) \equiv \varepsilon F(A')F(g).$$

Then surely $K_T F_T(A) = K_T(A) = F(A)$, and this forces the definition of K_T on objects. Moreover, for any $f: A \to B$ in $\mathscr{A}$, $K_T F_T(f) = K_T(\eta(B)f) = \varepsilon F(B)F\eta(B)F(f) = F(f)$. Thus $K_T F_T = F$.

Conversely, $K_T F_T = F$ implies that $K_T(\eta(B)f) = F(f)$; in particular, it implies for $g: A \to T(A')$ in $\mathscr{A}$ that $K_T(\eta T(A')g) = F(g)$. We shall see later that this forces the definition of K_T on arrows, once we know what it does to the arrow $1_{T(A')}$.

We calculate

$$K_T \varepsilon_T(A') = K_T(1_{T(A')}) = \varepsilon F(A') = \varepsilon K_T(A')$$

as required, and this forces the definition of $K_T(1_{T(A')})$. Now if $g: A \to T(A')$ in $\mathscr{A}$ is any arrow $A \to A'$ in $\mathscr{A}_T$,

$$g = \mu(A')\eta T(A')g = \mu(A')T(1_{T(A')})\eta T(A')g = 1_{T(A')} * \eta T(A')g,$$

where $*$ denotes composition in $\mathscr{A}_T$, hence

$$\begin{aligned} K_T(g) &= K_T(1_{T(A')})K_T(\eta T(A')g) \\ &= \varepsilon F(A')F(g), \end{aligned}$$

which finally establishes the uniqueness of K_T.

It remains to check that

$$\begin{aligned} UK_T(A) &= UF(A) = T(A) = U_T(A), \\ UK_T(g) &= U\varepsilon F(A')UF(g) = \mu(A')T(g) = U_T(g), \end{aligned}$$

and this completes the proof.

Corollary 6.8. Let $L_T: \mathscr{A}_T \to \mathscr{A}^T$ be the special case of the comparison functor K^T when $\mathscr{B} = \mathscr{A}_T$ (or of K_T when $\mathscr{B} = \mathscr{A}^T$), then we have functors

$$\mathscr{A} \xrightarrow{F_T} \mathscr{A}_T \xrightarrow{L_T} \mathscr{A}^T \xrightarrow{U^T} \mathscr{A}$$

with $F^T = L_T F_T$ left adjoint to U^T and $U_T = U^T L_T$ right adjoint to F_T. Moreover, F_T is bijective on objects, U^T is faithful and L_T is full and faithful.

Proof. In view of Propositions 6.5 and 6.7, it only remains to show that L_T is full and faithful. This follows from the following calculation: for any $g: A \to T(A')$ in $\mathscr{A}$,

$$L_T(g) = K^T(g) = U_T(g) = \mu(A')T(g),$$

hence

$$g = \mu(A')\eta T(A')g = \mu(A')T(g)\eta(A) = L_T(g)\eta(A).$$

Corollary 6.9. The Kleisli category of a triple is equivalent to the full subcategory of the Eilenberg–Moore category consisting of all free algebras.

Proof. The full and faithful functor L_T establishes an equivalence between $\mathscr{A}_T$ and a full subcategory of $\mathscr{A}^T$. Since, for any object A of $\mathscr{A}_T$,

$$L_T(A) = K^T(A) = (U_T(A), U_T\varepsilon_T(A)) = (T(A), \mu(A)) = F^T(A),$$

it follows that the objects of this subcategory are precisely the 'free' algebras of the triple.

Example T1 (continued). The Kleisli category of the triple associated with a monoid $\mathscr{M}$ is equivalent to the category of all free $\mathscr{M}$-sets regarded as a full subcategory of the category of all $\mathscr{M}$-sets.

Exercises.

1. Complete the proofs of Propositions 6.2 and 6.4 and the proofs in Examples T1 and T2.
2. Given a ring R (associative with unity element), construct a triple (T, η, μ) on the category **Ab** of abelian groups with $T(A) = R \otimes A$ for any abelian group A. What is the Eilenberg–Moore category of this triple?
3. Prove the associativity of composition in the Kleisli category of a triple.
4. (Linton). Show that the Eilenberg–Moore category may be constructed from the Kleisli category as a pullback:

$$\begin{array}{ccc} \mathscr{A}^T & \longrightarrow & \mathbf{Sets}^{\mathscr{A}_T^{op}} \\ {\scriptstyle U^T}\downarrow & & \downarrow{\scriptstyle \mathbf{Sets}^{L_T^{op}}} \\ \mathscr{A} & \xrightarrow[\text{Yoneda}]{} & \mathbf{Sets}^{\mathscr{A}^{op}} \end{array}$$

7 Examples of cartesian closed categories

In Part I we shall talk at length about 'cartesian closed categories', which will be defined equationally. In preparation, it may be useful to give a less formal definition and to present some examples.

A *cartesian closed* category is a category $\mathscr{C}$ with finite products (hence having a terminal object) such that, for each object B of $\mathscr{C}$, the functor $(-) \times B: \mathscr{C} \to \mathscr{C}$ has a right adjoint, denoted by $(-)^B: \mathscr{C} \to \mathscr{C}$. This means that, for all objects A, B and C of $\mathscr{C}$, there is an isomorphism

$$(*) \qquad \mathrm{Hom}_{\mathscr{C}}(A \times B, C) \simeq \mathrm{Hom}_{\mathscr{C}}(A, C^B)$$

and, moreover, that this isomorphism is natural in A, B and C.

Example 7.1. The category **Sets** is cartesian closed. Here $A \times B$ is the usual cartesian product of sets and C^B is the set of all functions $B \to C$. The bijection $(*)$ sends the function $f: A \times B \to C$ onto the function $f^*: A \to C^B$, where $f^*(a)(b) = f(a,b)$ for all $a \in A$ and $b \in B$. (See Section 3, Exercise 6.)

Example 7.2. More generally, for any small category $\mathscr{X}$, the functor category $\mathbf{Sets}^{\mathscr{X}}$ is cartesian closed. Also cartesian closed is the category of sheaves on a topological space and, in fact, every so-called topos (see Part II, Sections 9 and 10, even without natural numbers object).

Example 7.3. We recall from Section 1 that a poset $(P, \leqslant)$ (that is, preordered set satisfying the antisymmetry law) may be regarded as a category. As such, it has finite products if and only if it has a largest element 1 and a binary operation $\wedge$ such that $c \leqslant a \wedge b$ if and only if $c \leqslant a$ and $c \leqslant b$ for all elements a, b and c of P. In fact, $(P, 1, \wedge)$ is then a monoid satisfying the commutative and idempotent laws:

$$a \wedge b = b \wedge a, \quad a \wedge a = a.$$

Such a monoid is usually called a *semilattice*, and one may recapture the partial order by defining $a \leqslant b$ to mean $a \wedge b = a$. For $(P, 1, \wedge)$ to be cartesian closed there must be another binary operation $\Leftarrow$ such that $a \wedge b \leqslant c$ if and only if $a \leqslant c \Leftarrow b$ for all elements a, b and c of P. $(P, 1, \wedge, \Leftarrow)$ is then called a *Heyting semilattice.*

Example 7.4. A *Heyting algebra* $(P, 0, 1, \wedge, \vee, \Leftarrow)$ also has a smallest element 0 and a binary operation $\vee$ such that $a \vee b \leqslant c$ if and only if $a \leqslant c$ and $b \leqslant c$ for all elements a, b and c of P (hence $(P, \wedge, \vee)$ is a *lattice*), it being assumed that $(P, 1, \wedge, \Leftarrow)$ is a Heyting semilattice. When the underlying poset $(P, \leqslant)$ is viewed as a category, $\vee$ becomes a coproduct and the category is called *bicartesian closed*. Incidentally, the distributive law

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$

then follows from general categorical principles (see Section 5, Exercise 2).

A typical example of a Heyting algebra is the lattice of open subsets of a topological space X, with the following structure:

$$1 \equiv X, 0 \equiv \varnothing, U \wedge V \equiv U \cap V, U \vee V \equiv U \cup V,$$

$$V \Leftarrow U \equiv \mathrm{int}((X - U) \cup V),$$

for all open subsets U and V of X, where 'int' denotes the interior operation.

Another example of a Heyting algebra will be the lattice of subobjects of an object in a topos (see Part II, Section 5, Exercise 3). Many other examples are found in the literature (see the books by Balbes and Dwinger and by Rasiowa and Sikorski).

Example 7.5. **Cat**, the category of small categories, is cartesian closed. For any small categories $\mathscr{A}$ and $\mathscr{B}$, $\mathscr{A} \times \mathscr{B}$ is their product and $\mathscr{B}^{\mathscr{A}}$ is the category of all functors $\mathscr{A} \to \mathscr{B}$. (See: Section 1, Example C7; Section 2, Example C8; Proposition 2.3.)

Example 7.6. Although the category **top** of topological spaces and continuous mappings is not itself cartesian closed, various full subcategories of **top** are. For example, the category of Kelley spaces (that is, compactly generated Hausdorff spaces) is cartesian closed if products are defined in the usual way and Y^X is the set of all continuous functions $X \to Y$ with the compact–open topology. (See the book by MacLane for more details.)

Example 7.7. The category of ω-posets is cartesian closed. An *ω-poset* is a poset in which every countable ascending chain $a_0 \leqslant a_1 \leqslant a_2 \leqslant \ldots$ of elements has a supremum. Morphisms of ω-posets are mappings which preserve supremums of countable ascending chains (such mappings necessarily preserve order). The product structure is inherited from **Sets** and B^A is $\mathrm{Hom}(A, B)$ with order and supremum being defined componentwise. (For details see Part I, Proposition 18.1. For related cartesian closed categories see the book by Gierz *et al.*)

Example 7.8. The category of Kuratowski limit spaces is cartesian closed. A *limit space* is a set X with a partial ω-ary operation (that is, an operation defined on a subset of $X^{\mathbb{N}}$, the set of all countable sequences of elements of X) satisfying the following conditions:

(i) the constant sequence $(x, x, \ldots)$ has limit x;
(ii) if a sequence has limit x, then so does every subsequence;
(iii) if every subsequence of a sequence has a subsequence with limit x, then the sequence itself has limit x.

A morphism $f: X \to Y$ between limit spaces is a function such that, whenever $\{x_n | n \in \mathbb{N}\}$ is a sequence of elements of X with limit x, then $\{f(x_n) | n \in \mathbb{N}\}$ has limit $f(x)$. The product is defined as for sets, with limits given componentwise, and Y^X is the set of all morphisms $X \to Y$, where the limit of $\{f_n | n \in \mathbb{N}\}$ is said to be f provided the limit of $\{f_n(x_n) | n \in \mathbb{N}\}$ is $f(x)$

whenever the limit of $\{x_n | n \in \mathbb{N}\}$ is x. (For details see the book by Kuratowski, Chapter 2.)

Exercises

1. Carry out the detailed proof in any of the above examples.
2. Show that Heyting semilattices may be defined equationally.

I

Cartesian closed categories and λ-calculus

Introduction to Part I

λ-calculus or combinatory logic is a topic that logicians have studied since 1924. Cartesian closed categories are more recent in origin, having been invented by Lawvere (1964, see also Eilenberg and Kelly, 1966). Both are attempts to describe axiomatically the process of substitution, so it is not surprising to find that these two subjects are essentially the same. More precisely, there is an equivalence of categories between the category of cartesian closed categories and the category of typed λ-calculi with surjective pairing. This remains true if cartesian closed categories are provided with a weak natural numbers object and if typed λ-calculi are assumed to have a natural numbers type with iterator.

This result depends crucially on the *functional completeness* of cartesian closed categories, which goes back to the functional completeness of combinatory logic due to Schönfinkel and Curry. It asserts, in particular, that every arrow $\varphi(x): 1 \to B$ expressible as a polynomial in an indeterminate arrow $x: 1 \to A$ over a cartesian closed category $\mathscr{A}$ (with given objects A and B) is uniquely of the form $1 \xrightarrow{x} A \xrightarrow{f} B$, where f is an arrow in $\mathscr{A}$ not depending on x.

Functional completeness is closely related to the *deduction theorem* for positive intuitionistic propositional calculi presented as deductive systems. In our version, it associates with each proof of $\mathrm{T} \vdash \mathrm{B}$ on the assumption $\mathrm{T} \vdash \mathrm{A}$ a proof of $\mathrm{A} \vdash \mathrm{B}$ without assumptions. However, functional completeness goes beyond this; it asserts that the proof of $\mathrm{T} \vdash \mathrm{B}$ on the assumption $\mathrm{T} \vdash \mathrm{A}$ is, in some sense, *equivalent* to the proof by transitivity:

$$\frac{\mathrm{T} \vdash \mathrm{A} \quad \mathrm{A} \vdash \mathrm{B}}{\mathrm{T} \vdash \mathrm{B}}.$$

Deductive systems are also used to construct free cartesian closed categories generated by graphs, whose arrows $\mathrm{A} \to \mathrm{B}$ are equivalence classes of proofs.

We present a decision procedure for equality of arrows in the free cartesian closed category (with weak natural numbers object) generated by the empty graph; equivalently, for convertibility of expressions in the pure typed λ-calculus under consideration. This is the coherence problem for cartesian closed categories, the solution of which goes back to early work in the λ-calculus.

Finally, we study *C-monoids,* essentially monoids which may be viewed as one-object cartesian closed categories without terminal object. The category of C-monoids is shown to be equivalent (even isomorphic) to the category of untyped λ-calculi with surjective pairing. Again, this result depends on functional completeness of C-monoids.

It is shown that every C-monoid may be regarded as the monoid of endomorphisms of an object U in a cartesian closed category such that $U \times U \cong U \cong U^U$. An example of such a category with U not isomorphic to 1, due to Dana Scott, is presented.

The reader who wishes to see these results in their historical perspective is advised to look at the following comments.

Historical perspective on Part I

For the purpose of this discussion, it will suffice to define a *cartesian closed category* as a category with an object 1 and operations $(-) \times (-)$ and $(-)^{(-)}$ on objects satisfying conditions which assure that

(i) $\mathrm{Hom}(A, 1) \cong \{*\}$,
(ii) $\mathrm{Hom}(C, A \times B) \cong \mathrm{Hom}(C, A) \times \mathrm{Hom}(C, B)$,
(iii) $\mathrm{Hom}(A, C^B) \cong \mathrm{Hom}(A \times B, C)$.

Here $\{*\}$ is supposed a typical one-element set, chosen once and for all.

It will be instructive to reverse the historical process and see how combinatory logic could have been discovered by rigorous application of Occam's razor.

Condition (i) says that, for each object A, there is only one arrow $A \to 1$, hence we might as well forget about the object 1 and the arrow leading to it. However, the arrows $1 \to A$ must be preserved, let us call them *entities of type A.*

Condition (ii) says that the arrows $C \to A \times B$ are in one-to-one correspondence with pairs of arrows $C \to A$ and $C \to B$, hence we might as well forget about the arrows going into $A \times B$.

Condition (iii) says that the arrows $A \times B \to C$ are in one-to-one correspondence with the arrows $A \to C^B$, hence we might as well forget about the arrows coming out of $A \times B$ too. Consequently, we might as well forget about $A \times B$ altogether.

We end up with a category with a binary operation 'exponentiation' on objects. Of course, this will have to satisfy some conditions, but these may be a little difficult to state. It is interesting to note that Eilenberg and Kelly went on a similar *tour de force* and ended up with a category with exponentiation in which some monstrous diagrams had to commute.

We may go a little further and forget about the category structure as well, since arrows $A \to B$ are in one-to-one correspondence with entities of type B^A, which we shall write $B \Leftarrow A$ for typographical reasons. Composition of arrows is then represented by a single entity of type $((C \Leftarrow A) \Leftarrow (C \Leftarrow B)) \Leftarrow (B \Leftarrow A)$. However, we do need a binary operation on entities called 'application': given entities f of type B^A and a of type A, there is an entity $f^{\varsigma}a$ (read 'f of a') of type B.

We have now arrived at typed combinatory logic. But even this came rather late in the thinking of logicians, although type theory had already been introduced by Russell and Whitehead. Let us continue on our journey backwards in time and apply Occam's razor still further.

An arrow $A \to B$ in a category has a source A and a target B. But what if there is only one object? Such a category is called a monoid and, indeed, the original presentation of combinatory logic by Curry does describe a monoid with additional structure. (The binary operation of multiplication is defined in terms of the primitive operation of application.) Underlying untyped combinatory logic there is a tacit ontological assumption, namely that all entities are functions and that each function can be applied to any entity.

To present the work of Schönfinkel and Curry in the modern language of universal algebra, one should think of an algebra $A = (|A|, {}^{\varsigma}, I, K, S)$, where $|A|$ is a set, ${}^{\varsigma}$ is a binary operation and I, K and S are elements of $|A|$ or nullary operations. According to Schönfinkel, these had to satisfy the following identities:

$$I^{\varsigma}a = a,$$
$$(K^{\varsigma}a)^{\varsigma}b = a,$$
$$((S^{\varsigma}f)^{\varsigma}g)^{\varsigma}c = (f^{\varsigma}c)^{\varsigma}(g^{\varsigma}c),$$

for all elements a, b, c, f and g of $|A|$. (Actually, he defined I in terms of K and S, but this is beside the point here.) The reader may think of I as the identity function and of K as the function which assigns to every entity a the function with constant value a. It is a bit more difficult to put S into words and we shall refrain from doing so.

Schönfinkel (1924) discovered a remarkable result, usually called 'functional completeness'. In modern terms this may be expressed as follows: every polynomial $\varphi(x)$ in an indeterminate x over a Schönfinkel algebra A can be written in the form $f^{\varsigma}x$, where $f \in |A|$.

From now on in our exposition, the arrow of time will point in its customary direction.

Curry (1930) rediscovered Schönfinkel's results, but went further in his thinking. He discovered that a finite set of additional identities would assure that the element f representing the polynomial $\varphi(x)$ was uniquely determined. We shall not reproduce these identities here, but reserve the name 'Curry algebra' for a Schönfinkel algebra which satisfies them.

Using the terminology of Church (1941), one writes f as $\lambda_x\varphi(x)$, which must then satisfy two equations:

$$(\beta) \qquad (\lambda_x\varphi(x))^{\varsigma}a = \varphi(a),$$

$$(\eta) \qquad \lambda_x(f^{\varsigma}x) = f.$$

(Many mathematicians write $x \mapsto \varphi(x)$ in place of $\lambda_x\varphi(x)$.) A λ-calculus is a formal language built up from variables $x, y, z, \ldots$ by means of term forming operations $(-)^{\varsigma}(-)$ and $\lambda_x(-)$, the latter being assumed to bind all free occurrences of the variable x occurring in $(-)$, such that the two given identities hold. The basic entities I, K and S may then be defined formally by

$$I \equiv \lambda_x x,$$
$$K \equiv \lambda_x\lambda_y x,$$
$$S \equiv \lambda_u\lambda_v\lambda_z((u^{\varsigma}z)^{\varsigma}(v^{\varsigma}z)).$$

(Actually, Church would have called such a language a λK-calculus and Curry might have called it a $\lambda\beta\eta$-calculus, but never mind.)

Both Curry and Church realized the importance of introducing types into combinatory logic or λ-calculus. To do this one just has to observe that, if f has type $B \Leftarrow A$ and a has type A, then $f^{\varsigma}a$ has type B, as already pointed out. In particular, the basic entities I, K and S, suitably equipped with subscripts, should have prescribed types. Thus I_A, $K_{A,B}$ and $S_{A,B,C}$ have types $A \Leftarrow A$, $(A \Leftarrow B) \Leftarrow A$ and $((A \Leftarrow C) \Leftarrow (B \Leftarrow C)) \Leftarrow ((A \Leftarrow B) \Leftarrow C)$ respectively.

As pointed out in the book by Curry and Feys, these three types are precisely the axioms of intuitionistic implicational logic. Moreover, the rule which computes the type of $f^{\varsigma}a$ from those of f and a corresponds to modus ponens: from $B \Leftarrow A$ and A one may infer B. In fact, Schönfinkel's definition of I in terms of K and S is exactly the same as the known proof that $A \Leftarrow A$ may be derived from the other two axioms.

Incidentally, several early texts on propositional logic used only implication and negation as primitive connectives, having eliminated conjunction and other connectives by suitable definitions, again inspired by Occam's razor. The observation that it is more natural to retain conjunction and other connectives as primitive is probably due to Gentzen and was made again by Lawvere in a categorical context.

Curry and Feys also realized that the proof of Schönfinkel's version of functional completeness was really the same as the proof of the usual deduction theorem: if one can prove B on the assumption A then one can prove $\mathrm{B} \Leftarrow \mathrm{A}$ without any assumption. In fact, it asserts that the proof of B on the assumption A is 'equivalent' to the proof by modus ponens:

$$\frac{\mathrm{B} \Leftarrow \mathrm{A} \quad \mathrm{A}}{\mathrm{B}}.$$

From our viewpoint, Curry's version of functional completeness, which insists on the uniqueness of f such that $\varphi(x)$ equals $f^{\varsigma}x$, then presupposes that entities are not proofs but equivalence classes of proofs.

In connection with cartesian closed categories, the analogy with propositional logic requires that 1, $A \times B$ and B^A be written as T, $A \wedge B$ and $B \Leftarrow A$ respectively. (For other structured categories, the senior author had pointed out and exploited a similar analogy with certain deductive systems, beginning with the so-called 'syntactic calculus' (see Lambek 1961b, Appendix II), which traces the idea back to joint work with George D. Findlay in 1956.) The relation between λ-calculi with product types and cartesian closed categories then suggests the observation: types = formulas, terms = proofs, or rather equivalence classes of proofs. Independently, W. Howard in 1969 privately circulated an influential manuscript on the equivalence of typed λ-terms (there called 'constructions') and derivations in various calculi, which finally appeared in the 1980 Curry Festschrift (see also Stenlund 1972).

Up to this point we have avoided discussing natural numbers. In an untyped λ-calculus natural numbers are easily defined (Church 1941). Writing

$$f \circ g \equiv \lambda_x(f^{\varsigma}(g^{\varsigma}x)),$$

one regards 2 as the process which assigns to every function f its iterate $f \circ f$, so $2^{\varsigma}f \equiv f \circ f$. Formally, one defines

$$0 \equiv \lambda_x I, \quad 1 \equiv \lambda_x x = I, \quad 2 \equiv \lambda_x(x \circ x), \ldots.$$

The successor function and the usual operations on natural numbers are defined by

$$S^{\int} n \equiv \lambda_y(y \circ (n^{\int} y)),$$
$$m + n \equiv \lambda_y((m^{\int} y) \circ (n^{\int} y)),$$
$$mn \equiv m \circ n,$$
$$m^n \equiv n^{\int} m.$$

Unfortunately, there are difficulties with this as soon as one introduces types. For, if a has type A, then f and g in $(f \circ g)^{\int} a$ both have types $A^A = B$ say. For $n^{\int} f$ to make sense, n will have to be of type B^B, and for $n^{\int} m$ to make sense, m will have to be of type B. If m and n are to have the same type, we are thus led to require that $B^B = B$, which is certainly not true in general, although Dana Scott (1972) showed that one may have $B^B \cong B$.

One way to get around this difficulty is to postulate a type N of natural numbers, a term 0 of type N and term forming operations $S(-)$ (successor) and $I(-,-,-)$ (iterator) such that $S(n)$ has type n and $I(a, h, n)$ has type A for all n of type N, a of type A and h of type A^A. These must satisfy suitable equations to assure that $I(a, h, n)$ means $h^{n\int} a$.

The analogous concept for cartesian closed categories is a *weak natural numbers object*: an object N with arrows $0: 1 \to N$ and $S: N \to N$ and a process which assigns to all arrows $a: 1 \to A$ and $h: A \to A$ an arrow $g: N \to A$ such that the following diagram commutes:

$$\begin{array}{ccccc} 1 & \xrightarrow{0} & N & \xrightarrow{S} & N \\ \downarrow & & \downarrow g & & \downarrow g \\ 1 & \xrightarrow[a]{} & A & \xrightarrow[h]{} & A \end{array}$$

Lawvere had defined a (strong) natural numbers object to be such that the arrow $g: N \to A$ with the above property is unique.

For us, a typed λ-calculus contains by definition the structure given by N, 0, S and I. In stating Theorem 11.3 on the equivalence between typed λ-calculi and cartesian closed categories, we stipulate that the latter be equipped with a weak natural numbers object. Such categories were first studied formally by Marie-France Thibault (1977, 1982), who called them 'prerecursive categories', although they are implicit in the work of logicians, e.g. in Gödel's functionals of finite type (1958).

We would have preferred to state Theorem 11.3 for strong natural numbers objects in Lawvere's sense. Unfortunately, we do not yet know how to handle the corresponding notion in typed λ-calculus equationally. As far as we can see, the iterators appearing in the literature (e.g. Troelstra 1973) mostly correspond to weak natural numbers objects. See however Sanchis (1967).

For further historical comments the reader is referred to the end of Part I.

1 Propositional calculus as a deductive system

We recall (Part 0; Definition 1.2) that, for categorists, a *graph* consists of two classes and two mappings between them:

$$\text{Arrows} \;\underset{\text{target}}{\overset{\text{source}}{\rightrightarrows}}\; \text{Objects}$$

In graph theory the arrows are usually called 'oriented edges' and the objects 'nodes' or 'vertices', but in various branches of mathematics other words may be used. Instead of writing

$$\text{source}(f) = A, \quad \text{target}(f) = B,$$

one often writes $f: A \to B$ or $A \xrightarrow{f} B$. We shall look at graphs with additional structure which are of interest in logic.

A *deductive system* is a graph with a specified arrow

R1a. $A \xrightarrow{1_A} A,$

and a binary operation on arrows (*composition*)

R1b. $\dfrac{A \xrightarrow{f} B \quad B \xrightarrow{g} C}{A \xrightarrow{gf} C}.$

Logicians will think of the objects of a deductive system as *formulas*, of the arrows as *proofs* (or *deductions*) and of an operation on arrows as a *rule of inference*.

Logicians should note that a deductive system is concerned not just with unlabelled entailments or sequents $A \to B$ (as in Gentzen's proof theory), but with deductions or proofs of such entailments. In writing $f: A \to B$ we think of f as the 'reason' why A entails B.

A *conjunction calculus* is a deductive system dealing with truth and conjunction. Thus we assume that there is given a formula T (= true) and a binary operation $\wedge$ (= and) for forming the conjunction $A \wedge B$ of two given formulas A and B. Moreover, we specify the following additional arrows and rules of inference:

R2. $A \xrightarrow{\bigcirc_A} \mathrm{T}$;

R3a. $A \wedge B \xrightarrow{\pi_{A,B}} A$,

R3b. $A \wedge B \xrightarrow{\pi'_{A,B}} B$,

R3c. $$\frac{C \xrightarrow{f} A \quad C \xrightarrow{g} B}{C \xrightarrow{\langle f,g\rangle} A \wedge B}.$$

Here is a sample proof of the so-called commutative law for conjunction:

$$\frac{A \wedge B \xrightarrow{\pi'_{A,B}} B \quad A \wedge B \xrightarrow{\pi_{A,B}} A}{A \wedge B \xrightarrow{\langle \pi'_{A,B}, \pi_{A,B}\rangle} B \wedge A}.$$

The presentation of this proof in tree-form, while instructive, is superfluous. It suffices to denote it by $\langle \pi'_{A,B}, \pi_{A,B}\rangle$ or even by $\langle \pi', \pi\rangle$ when the subscripts are understood.

Another example is the proof of the associative law $\alpha_{A,B,C}:(A \wedge B) \wedge C \to A \wedge (B \wedge C)$. It is given by

$$\alpha_{A,B,C} \equiv \langle \pi_{A,B}\pi_{A\wedge B,C}, \langle \pi'_{A,B}\pi_{A\wedge B,C}, \pi'_{A\wedge B,C}\rangle\rangle \tag{1.1}$$

or just by $\alpha \equiv \langle \pi\pi, \langle \pi'\pi, \pi'\rangle\rangle$.

If we compose operations on proofs, we obtain 'derived' rules of inference. For example, consider the derived rule:

$$\frac{\dfrac{A \wedge C \xrightarrow{\pi_{A,C}} A \quad A \xrightarrow{f} B}{A \wedge C \to B} \quad \dfrac{A \wedge C \xrightarrow{\pi'_{A,C}} C \quad C \xrightarrow{g} D}{A \wedge C \to D}}{A \wedge C \xrightarrow{f \wedge g} B \wedge D}.$$

It asserts that from proofs f and g one can construct the proof

$$f \wedge g = \langle f\pi_{A,C}, g\pi'_{A,C}\rangle.$$

Thus we may write simply

$$\frac{A \xrightarrow{f} B \quad C \xrightarrow{g} D}{A \wedge C \xrightarrow{f \wedge g} B \wedge D}.$$

A *positive intuitionistic propositional calculus* is a conjunction calculus with an additional binary operation $\Leftarrow$ (= if). Thus, if A and B are formulas,

so are T, $A \wedge B$ and $A \Leftarrow B$. (Yes, most people write $B \Rightarrow A$ instead.) We also specify the following new arrow and rule of inference.

R4a. $(A \Leftarrow B) \wedge B \xrightarrow{\varepsilon_{A,B}} A$,

R4b. $$\frac{C \wedge B \xrightarrow{h} A}{C \xrightarrow{h^*} A \Leftarrow B}.$$

Actually we should have written $h^* = \Lambda^C_{A,B}(h)$, but the subscripts are usually understood from the context.

We note that from R4b, with the help of R4a, one may derive

R′4b. $C \xrightarrow{\eta_{C,B}} (C \wedge B) \Leftarrow B$,

R′4c. $$\frac{D \xrightarrow{g} A}{(D \Leftarrow B) \xrightarrow{g \Leftarrow 1_B} (A \Leftarrow B)}.$$

To derive these, we put

$$\eta_{C,B} \equiv 1^*_{C\wedge B}, \quad g \Leftarrow 1_B \equiv (g\varepsilon_{D,B})^*.$$

Conversely, one may derive R4b from R′4b and R′4c by putting

$$h^* \equiv (h \Leftarrow 1_B)\eta_{C,B}.$$

For future reference, we also note the following two derived rules of inference:

$$\frac{A \xrightarrow{f} B}{\mathrm{T} \xrightarrow{\ulcorner f \urcorner} B \Leftarrow A}, \quad \frac{\mathrm{T} \xrightarrow{g} B \Leftarrow A}{A \xrightarrow{g^\int} B},$$

where

$$\ulcorner f \urcorner \equiv (f\pi'_{1,A})^*, \quad g^\int \equiv \varepsilon_{B,A}\langle g\bigcirc_A, 1_A\rangle.$$

An intuitionistic (propositional) calculus is more than a positive one; it requires also falsehood and disjunction, that is, a formula $\perp$ (= false) and an operation $\vee$ (= or) on formulas, together with the following additional arrows:

R5. $\perp \xrightarrow{\square_A} A$;

R6a. $A \xrightarrow{\kappa_{A,B}} A \vee B$, R6b. $B \xrightarrow{\kappa'_{A,B}} A \vee B,$

R6c. $(C \Leftarrow A) \wedge (C \Leftarrow B) \xrightarrow{\zeta^C_{A,B}} C \Leftarrow (A \vee B).$

The last mentioned arrow gives rise to and may be derived from the rule

R′6c. $$\frac{A \xrightarrow{f} C \quad B \xrightarrow{g} C}{A \vee B \xrightarrow{[f,g]} C}.$$

Indeed, we may put

$$[f,g] \equiv (\zeta^C_{A,B}\langle \ulcorner f \urcorner, \ulcorner g \urcorner \rangle)^{\int}.$$

If we want *classical* propositional logic, we must also require

R7. $\bot \Leftarrow (\bot \Leftarrow A) \to A.$

Exercises

1. For the appropriate deductive systems, obtain proofs of the following and their converses:

   $A \wedge \mathrm{T} \to A, A \Leftarrow \mathrm{T} \to A, \mathrm{T} \Leftarrow A \to \mathrm{T};$

   $(A \wedge B) \Leftarrow C \to (A \Leftarrow C) \wedge (B \Leftarrow C);$

   $A \Leftarrow (B \wedge C) \to (A \Leftarrow C) \Leftarrow B;$

   $A \wedge \bot \to \bot, A \Leftarrow \bot \to \mathrm{T}, A \vee \bot \to A;$

   $(A \wedge C) \vee (B \wedge C) \to (A \vee B) \wedge C.$

2. For the appropriate deductive systems, deduce the following derived rules of inference:

$$\frac{A \xrightarrow{f} B \quad C \xrightarrow{g} D}{A \Leftarrow D \xrightarrow{f \Leftarrow g} B \Leftarrow C}; \quad \frac{A \xrightarrow{f} B \quad C \xrightarrow{g} D}{A \vee C \xrightarrow{f \vee g} B \vee D}.$$

3. Show how $\zeta^C_{A,B}$ may be defined in terms of the rule R′6c.

4. Show that, in the presence of R1 to R6, the classical axiom R7 may be replaced by

   $\mathrm{T} \to A \vee (\bot \Leftarrow A).$

2 The deduction theorem

The usual deduction theorem asserts:

if $A \wedge B \vdash C$ then $A \vdash C \Leftarrow B$.

This result is here incorporated into R4, with the deduction symbol $\vdash$ replaced by actual arrows in the appropriate deductive system $\mathscr{L}$:

$$\frac{h: A \wedge B \to C}{h^*: A \to C \Leftarrow B}.$$

However, at a higher level, the horizontal bar functions as a deduction symbol, and we obtain a new form of the deduction theorem. It deals with proofs from an *assumption* $x: \mathrm{T} \to A$. In other words, we form a new deductive system $\mathscr{L}(x)$ by adjoining a new arrow $x: \mathrm{T} \to A$ and talk about proofs $\varphi(x): B \to C$ in this new system. More precisely, $\mathscr{L}(x)$ has the same formulas (= objects) as $\mathscr{L}$ and its proofs (= arrows) $\varphi(x)$ are freely generated from those of $\mathscr{L}$ and the new arrow x by the appropriate rules of inference (= operations). Clearly, if $\mathscr{L}$ is a conjunction calculus (positive calculus, intuitionistic calculus, classical calculus), so is the new deductive system $\mathscr{L}(x)$.

Proposition 2.1. (Deduction theorem). In a conjunction, positive, intuitionistic or classical calculus, with every proof $\varphi(x): B \to C$ from the assumption $x: \mathrm{T} \to A$ there is associated a proof $f: A \wedge B \to C$ in $\mathscr{L}$ not depending on x.

We write $f = \kappa_{x \in A}\varphi(x)$, where the subscript '$x \in A$' indicates that x is of type A.

Proof. We shall give the proof for a positive calculus. The same proof is valid for a conjunction calculus, if * is ignored. The proof goes through for an intuitionistic or classical calculus, as the additional structure is presented in the form of arrows rather than rules of inference.

We note that every proof $\varphi(x): B \to C$ from the assumption $x: \mathrm{T} \to A$ must have one of the five forms:

(i) $k: B \to C$, a proof in $\mathscr{L}$;
(ii) $x: \mathrm{T} \to A$, with $B = \mathrm{T}$ and $C = A$;
(iii) $\langle \psi(x), \chi(x) \rangle$, where $\psi(x): B \to C'$, $\chi(x): B \to C''$, $C = C' \wedge C''$;
(iv) $\chi(x)\psi(x)$, where $\psi(x): B \to D$, $\chi(x): D \to C$;
(v) $\psi(x)^*$, where $\psi(x): B \wedge C' \to C''$, $C = C'' \Leftarrow C'$.

In all cases, $\psi(x)$ and $\chi(x)$ are 'shorter' proofs than $\varphi(x)$, and we define inductively:

(i) $\kappa_{x \in A} k = k\pi'_{A,B}$;
(ii) $\kappa_{x \in A} x = \pi_{A,\mathrm{T}}$,
(iii) $\kappa_{x \in A}\langle \psi(x), \chi(x) \rangle = \langle \kappa_{x \in A}\psi(x), \kappa_{x \in A}\chi(x) \rangle$;
(iv) $\kappa_{x \in A}(\chi(x)\psi(x)) = \kappa_{x \in A}\chi(x)\langle \pi_{A,B}, \kappa_{x \in A}\psi(x) \rangle$;
(v) $\kappa_{x \in A}(\psi(x)^*) = (\kappa_{x \in A}\psi(x)\alpha_{A,B,C'})^*$;

where $\alpha_{A,B,C'}:(A \wedge B) \wedge C' \to A \wedge (B \wedge C')$ is the proof of associativity discussed in Section 1.

The above argument was by induction on the *length* of the proof $\varphi(x)$. Formally, this may be defined as 0 in cases (i) and (ii), as the sum of the lengths of $\chi(x)$ and $\psi(x)$ plus 1 in cases (iii) and (iv) and as the length of $\psi(x)$ plus 1 in case (v).

Remark 2.1. Logicians don't usually talk of an assumption $x: \mathrm{T} \to A$ if there is a known proof $a: \mathrm{T} \to A$ or another assumption $y: \mathrm{T} \to A$; but from our algebraic viewpoint, this does not matter.

The reader is warned that we do not distinguish notationally between composition of proofs gf in $\mathscr{L}$ and in $\mathscr{L}(x)$. In $\mathscr{L}$, $\kappa_{x \in A}gf = gf\pi'_{A,B}$ and in $\mathscr{L}(x)$ it is $g\pi'_{A,B}\langle \pi_{A,B}, f\pi'_{A,B}\rangle$.

Exercise

Prove the following general form of the deduction theorem for the positive intuitionistic propositional calculus: with every proof $\varphi(x): B \to C$ from the assumption $x: D \to A$ there is associated a proof $f:(A \Leftarrow D) \wedge B \to C$. Hint: writing $f = \rho_x\varphi(x)$, put

(i) $\rho_x k = k\pi'_{A \Leftarrow D,B}$, (ii) $\rho_x x = \varepsilon_{A,B}$,

(iii) $\rho_x\langle \psi(x), \chi(x)\rangle = \langle \rho_x\psi(x), \rho_x\chi(x)\rangle$,

(iv) $\rho_x(\chi(x)\psi(x)) = \rho_x\chi(x)\langle \pi_{A \Leftarrow D,B}, \rho_x\psi(x)\rangle$,

(v) $\rho_x(\psi(x)^*) = (\rho_x\psi(x)\alpha_{A \Leftarrow D,B',B''})^*$, where $\psi(x): B' \wedge B'' \to C$.

3 Cartesian closed categories equationally presented

A *category* is a deductive system in which the following equations hold between proofs:

E1. $f1_A = f, \quad 1_B f = f, \quad (hg)f = h(gf)$,
for all $f: A \to B, \quad g: B \to C, \quad h: C \to D$.

Thus, from any deductive system one may obtain a category by imposing a suitable equivalence relation between proofs.

A *cartesian category* is both a category and a conjunction calculus satisfying the additional equations:

E2. $f = \bigcirc_A$, for all $f: A \to \mathrm{T}$;

E3a. $\pi_{A,B}\langle f, g\rangle = f$,

E3b. $\pi'_{A,B}\langle f, g\rangle = g$,

E3c. $\langle \pi_{A,B}h, \pi'_{A,B}h\rangle = h$,
for all $f: C \to A, \quad g: C \to B, \quad h: C \to A \wedge B$.

E2 asserts T is a terminal object. One usually writes $\mathrm{T} \equiv 1$, and we shall do so from now on. An equivalent formulation of E2 is

E′2: $1_1 = \bigcirc_1, \bigcirc_B f = \bigcirc_A$ for all $f: A \to B$.

E3 asserts that $A \wedge B$ is a product of A and B with projections $\pi_{A,B}$ and $\pi'_{A,B}$. We shall adopt the usual notation $A \wedge B \equiv A \times B$.

As a consequence of E3, let us record the *distributive law*:

$$\langle f, g\rangle h = \langle fh, gh\rangle \tag{3.1}$$

for all $f: C \to A, \quad g: C \to B, \quad h: D \to C$.

Proof. We show this as follows, omitting subscripts:

$$\begin{aligned}\langle f, g\rangle h &= \langle \pi(\langle f, g\rangle h), \pi'(\langle f, g\rangle h)\rangle && \text{// E3c } \eta\\ &= \langle (\pi\langle f, g\rangle)h, (\pi'\langle f, g\rangle)h\rangle && \text{// E1 assoc.}\\ &= \langle fh, gh\rangle. && \text{// E3a,b } \beta \text{ rule}\end{aligned}$$

We shall also write

$$f \times g \equiv f \wedge g = \langle f\pi_{A,C}, g\pi'_{A,C}\rangle,$$

whenever $f: A \to B$ and $g: C \to D$, and note that $\times: \mathscr{A} \times \mathscr{A} \to \mathscr{A}$ is a functor (see Part 0, Definition 1.3). Indeed, we have

$$\begin{aligned}1_A \times 1_C &= \langle 1_A\pi_{A,C}, 1_C\pi'_{A,C}\rangle\\ &= \langle \pi_{A,C}, \pi'_{A,C}\rangle\\ &= \langle \pi_{A,C}1_{A\times C}, \pi'_{A,C}1_{A\times C}\rangle\\ &= 1_{A\times C}\end{aligned}$$

and, omitting subscripts, by the distributive law,

$$\begin{aligned}(f \times g)(f' \times g') &= \langle f\pi, g\pi'\rangle\langle f'\pi, g'\pi'\rangle\\ &= \langle f\pi\langle f'\pi, g'\pi'\rangle, g\pi'\langle f'\pi, g'\pi'\rangle\rangle && \text{distr}\\ &= \langle ff'\pi, gg'\pi'\rangle && \beta \text{ rule}\\ &= ff' \times gg'. && \eta\end{aligned}$$

A *cartesian closed category* is a cartesian category $\mathscr{A}$ with additional structure R4 satisfying the additional equations

E4a. $\varepsilon_{A,B}\langle h^*\pi_{C,B}, \pi'_{C,B}\rangle = h$,

E4b. $(\varepsilon_{A,B}\langle k\pi_{C,B}, \pi'_{C,B}\rangle)^* = k$,
for all $h: C \wedge B \to A$ and $k: C \to A \Leftarrow B$.

Thus, a cartesian closed category is a positive intuitionistic propositional calculus satisfying the equations E1 to E4. This illustrates the general principle that one may obtain interesting categories from deductive systems by imposing an appropriate equivalence relation on proofs.

Inasmuch as we have decided to write $C \wedge B \equiv C \times B$, we shall also write $A \Leftarrow B \equiv A^B$. The equations E4 assure that the mapping

$$\mathrm{Hom}(C \times B, A) \xrightarrow{*} \mathrm{Hom}(C, A^B)$$

is a one-to-one correspondence. In fact, one has the following universal property of the arrow $\varepsilon_{A,B}: A^B \times B \to A$:

given any arrow $h: C \times B \to A$, there is a unique arrow $h^*: C \to A^B$ such that

$$\varepsilon_{A,B}(h^* \times 1_B) = h.$$

The reader who recalls the notion of adjoint functor will recognize that therefore $U_B = (-)^B$ is right adjoint to the functor $F_B = (-) \times B: \mathscr{A} \to \mathscr{A}$ with coadjunction $\varepsilon_B: F_B U_B \to 1_{\mathscr{A}}$ defined by $\varepsilon_B(A) = \varepsilon_{A,B}$. Thus, an equivalent description of cartesian closed categories makes use of the adjunction $\eta_B: 1_{\mathscr{A}} \to U_B F_B$ in place of $*$, where $\eta_B(C) = \eta_{C,B}: C \to (C \times B)^B$, and stipulates equations expressing the functoriality of U_B and the naturality of ε_B and η_B as well as the two adjunction equations. Here

$$U_B(f) = f^B \equiv f \Leftarrow 1_B = (f \varepsilon_{A,B})^*,$$

for all $f: A \to A'$. (For η_B see R′4b in Section 1.)

We shall state another useful equation, which may also be regarded as a kind of distributive law.

$$h^*k = (h\langle k\pi_{D,B}, \pi'_{D,B}\rangle)^*, \tag{3.2}$$

where $h: A \times B \to C$ and $k: D \to A$.

Proof. We show this as follows, omitting subscripts:

$$\begin{aligned} h^*k &= (\varepsilon\langle h^*k\pi, \pi'\rangle)^* \\ &= (\varepsilon\langle h^*\pi, \pi'\rangle\langle k\pi, \pi'\rangle)^* \\ &= (h\langle k\pi, \pi'\rangle)^*. \end{aligned}$$

Quite important is the following bijection, which holds in any cartesian closed category.

$$\mathrm{Hom}(A, B) \cong \mathrm{Hom}(1, B^A). \tag{3.3}$$

Proof. As in Section 1, with any $f: A \to B$ we associate $\ulcorner f \urcorner: 1 \to B^A$, called the *name* of f by Lawvere, given by

$$\ulcorner f \urcorner \equiv (f\pi'_{1,A})^*,$$

and with any $g: 1 \to B^A$ we associate $g^{\int}: A \to B$, read 'g of', given by

$$g^{\int} \equiv \varepsilon_{B,A}\langle g \bigcirc_A, 1_A\rangle.$$

We then calculate

$$\ulcorner f \urcorner^{\int} = f, \quad \ulcorner g^{\int} \urcorner = g.$$

Exercises

1. Show that in any cartesian category

$$A \times 1 \cong A, \quad A \times B \cong B \times A, \quad (A \times B) \times C \cong A \times (B \times C).$$

2. Show that in any cartesian closed category

$$A^1 \cong A, \quad 1^A \cong 1, \quad (A \times B)^C \cong A^C \times B^C, \quad A^{B \times C} \cong (A^C)^B.$$

3. Write down the equivalent definition of a cartesian closed category in terms of U_B, F_B, η_B and ε_B.
4. Prove the last two equations of Section 3.

4 Free cartesian closed categories generated by graphs

Given a graph $\mathscr{X}$, we may construct the positive intuitionistic calculus $\mathscr{D}(\mathscr{X})$ and the cartesian closed category $\mathscr{F}(\mathscr{X})$ *freely generated* by $\mathscr{X}$.

Informally speaking, $\mathscr{D}(\mathscr{X})$ is the smallest positive intuitionistic calculus whose formulas include the vertices of $\mathscr{X}$ and whose proofs include the arrows of $\mathscr{X}$. (Logicians may think of the latter as 'postulates', although there may be more than one way of postulating $X \to Y$, as there may be more than one arrow $X \to Y$ in $\mathscr{X}$.) More precisely, the formulas and proofs of $\mathscr{D}(\mathscr{X})$ are defined inductively as follows: all vertices of $\mathscr{X}$ are formulas, $\mathrm{T}(\equiv 1)$ is a formula, if A and B are formulas so are $A \wedge B (\equiv A \times B)$ and $B \Leftarrow A (\equiv B^A)$; the arrows of $\mathscr{X}$ and the arrows 1_A, $\bigcirc_A$, $\pi_{A,B}$, $\pi'_{A,B}$ and $\varepsilon_{A,B}$ are proofs, for all formulas A and B, and proofs are closed under the rules of inference–composition, $\langle -, - \rangle$ and $(-)^*$.

We construct $\mathscr{F}(\mathscr{X})$ from $\mathscr{D}(\mathscr{X})$ by imposing all equations between proofs which have to hold in any cartesian closed category. Another way of saying this is that we pick the smallest equivalence relation between proofs satisfying the appropriate substitution laws and respecting the equations of a cartesian closed category. The equivalence classes of proofs are then the arrows of $\mathscr{F}(\mathscr{X})$; but, as usual, we will not distinguish notationally between proofs and their equivalence classes.

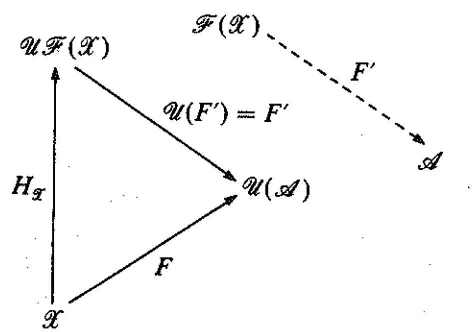
Let **Grph** be the category of graphs, whose objects are graphs and whose morphisms $F: \mathscr{X} \to \mathscr{Y}$ are pairs of mappings F: Objects($\mathscr{X}$) $\to$ Objects($\mathscr{Y}$) and F: Arrows($\mathscr{X}$) $\to$ Arrows($\mathscr{Y}$) such that $f: X \to X'$ implies $F(f): F(X) \to F(X')$.

Let **Cart** be the category of cartesian closed categories, whose objects are cartesian closed categories and whose arrows are functors $F: \mathscr{A} \to \mathscr{B}$ which preserve the cartesian closed structure on the nose, that is,

$$F(1)=1, \quad F(A\times B)=F(A)\times F(B), \quad F(A^B)=F(A)^{F(B)};$$
$$F(\bigcirc_A)=\bigcirc_{F(A)}, \quad F(\pi_{A,B})=\pi_{F(A),F(B)}, \text{ etc.};$$
$$F(\langle f,g\rangle)=\langle F(f),F(g)\rangle \text{ etc.}$$

Let $\mathscr{U}$ be the obvious *forgetful* functor **Cart** → **Grph**. With any graph $\mathscr{X}$ we associate a morphism of graphs $H_{\mathscr{X}}:\mathscr{X}\to\mathscr{U}\mathscr{F}(\mathscr{X})$ as follows: $H_{\mathscr{X}}(X)=X$ and, if $f:X\to Y$ in $\mathscr{X}$, then $H_{\mathscr{X}}(f)=f$ (the equivalence class of f regarded as a proof in $\mathscr{D}(\mathscr{X})$). We then have the following universal property:

Proposition 4.1. Given any cartesian closed category $\mathscr{A}$ and any morphism $F:\mathscr{X}\to\mathscr{U}(\mathscr{A})$ of graphs, there is a unique arrow $F':\mathscr{F}(\mathscr{X})\to\mathscr{A}$ in **Cart** such that $\mathscr{U}(F')H_{\mathscr{X}}=F$.



Proof. Indeed, the construction of F' is forced upon us:

$$F'(X)=F(X), \quad F'(\mathrm{T})=1, \quad F'(A\wedge B)=F'(A)\times F'(B), \text{ etc.};$$
$$F'(f)=F(f) \quad \text{for all } f:X\to Y, \quad F'(\bigcirc_A)=\bigcirc_{F'(A)}, \text{ etc.};$$
$$F'(\langle f,g\rangle)=\langle F'(f),F'(g)\rangle, \text{ etc.}$$

We must check that F' is well defined, that is, for all $f,g:A\to B$ in $\mathscr{F}(\mathscr{X})$, $f=g$ implies $F'(f)=F'(g)$. This easily follows because no equations hold in $\mathscr{F}(\mathscr{X})$ except those that have to hold.

The above universal property means that $\mathscr{F}$ is a functor **Grph** → **Cart** which is left adjoint to $\mathscr{U}$ with adjunction $H_{(\,)}:\mathrm{Id}\to\mathscr{U}\mathscr{F}$.

The reader will have noticed that the objects of the category **Grph** and **Cart** introduced here are classes. These may have to be regarded as sets in an appropriate universe.

Exercise

Show that the deductive system $\mathscr{L}(x)$ in Section 2 is $\mathscr{D}(\mathscr{L}_x)$, where $\mathscr{L}_x$ is the graph obtained from $\mathscr{L}$ by adjoining a new edge x between the old vertices T and A.

5 Polynomial categories

Given objects A_0 and A of a (cartesian, cartesian closed) category $\mathscr{A}$, how does one adjoin an indeterminate arrow $x:A_0\to A$ to $\mathscr{A}$? One method is to adjoin an arrow $x:A_0\to A$ to the underlying graph of $\mathscr{A}$ and then to form the (cartesian, cartesian closed) category freely generated by the new graph, as was done in Section 4 for cartesian closed categories. Equivalently, one could first form the deductive system (conjunction calculus, positive intuitionistic calculus $\mathscr{A}[x]$ based on the 'assumption' x, as was done in Section 2 in the special case $A_0=\mathrm{T}$. The formulas of $\mathscr{A}[x]$ are the objects of $\mathscr{A}$ and the proofs of $\mathscr{A}[x]$ are formed from the arrows of $\mathscr{A}$ and the new arrow $x:A_0\to A$ by the appropriate rules of inference.

To assure that $\mathscr{A}[x]$ becomes a category and that the inclusion of $\mathscr{A}$ into $\mathscr{A}[x]$ becomes a functor, one then imposes the appropriate equations between proofs. If equality of proofs is denoted by $\underset{x}{=}$, we may also regard $\underset{x}{=}$ as the smallest equivalence relation $\equiv$ between proofs such that

$$gf=h \text{ in } \mathscr{A} \text{ implies } gf\equiv h,$$
$$\psi(x)\equiv\psi'(x) \text{ and } \chi(x)\equiv\chi'(x) \text{ implies } \chi(x)\psi(x)\equiv\chi'(x)\psi'(x),$$
$$\varphi(x)1_B\equiv\varphi(x)\equiv 1_C\varphi(x),$$
$$(\chi(x)\psi(x))\varphi(x)\equiv\chi(x)(\psi(x)\varphi(x)),$$

for all $\varphi(x):B\to C$, $\psi(x),\psi'(x):C\to D$, $\chi(x),\chi'(x):D\to E$.

Note that, in view of the reflexive law, $\equiv$ and $\underset{x}{=}$ extend equality in $\mathscr{A}$. Arrows in the category $\mathscr{A}[x]$ are proofs on the assumption x modulo $\underset{x}{=}$, they may be regarded as *polynomials* in x.

The same construction works for cartesian categories or cartesian closed categories, only then $\underset{x}{=}$ must be such that $\mathscr{A}[x]$ becomes a cartesian or cartesian closed category and that the functor $\mathscr{A}\to\mathscr{A}[x]$ preserves the cartesian (closed) structure. That is, the equivalence relations $\equiv$ between proofs considered above must also satisfy:

$$\text{if } \langle f,g\rangle=h \text{ in } \mathscr{A} \text{ then } \langle f,g\rangle\equiv h, \text{ etc.};$$
$$\text{if } \psi(x)\equiv\psi'(x) \text{ and } \chi(x)\equiv\chi'(x) \text{ then } \langle\psi(x),\chi(x)\rangle\equiv\langle\psi'(x),\chi'(x)\rangle,$$
$$\pi_{B,C}\langle\psi(x),\chi(x)\rangle\equiv\psi(x), \text{ etc.},$$

for all $\psi(x)$, $\psi'(x):D\to B$ and $\chi(x)$, $\chi'(x):D\to C$.