CHAPTER 2

WHAT CATEGORIES ARE

"... *understanding consists in reducing one type of reality to another.*"

Claude Levi-Strauss

2.1. Functions are sets?

A good illustration of the way in which set theory formalises an intuitive mathematical idea is provided by an examination of the notion of a *function.* A function is an association between objects, a correspondence that assigns to a given object one and only one other object. It may be thought of as a rule, or operation, which is applied to something to obtain its associated thing. A useful way of envisaging a function is as an input–output process, a kind of "black box" (see figure). For a given input the function produces a uniquely determined output. For example, the instruction "multiply by 6" determines a function which for input 2 gives output $6 \times 2 = 12$, which associates with the number 1 the number 6, which assigns 24 to 4, and so on. The inputs are called *arguments* of the function and the outputs *values,* or *images* of the inputs that they are produced by. If f denotes a function, and x an input, then the corresponding output, the image of x under f, is denoted $f(x)$. The above example may then be displayed as that function f given by the rule $f(x) = 6x$.

If A is the set of all appropriate inputs to function f (in our example A will include the number 2, but not the Eiffel Tower), and B is a set that includes all the f-images of the members of A (and possibly the Eiffel Tower as well), then we say that f is a function *from* A *to* B. This is symbolised as $f: A \to B$ or $A \xrightarrow{f} B$. A is called the *domain* or *source* of f and B is the *codomain* or *target.*

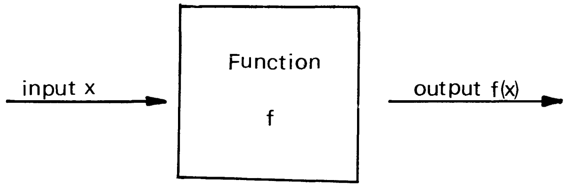


Fig. 2.1.

How does set theory deal with this notion? To begin with we introduce the notion of an *ordered* pair, as consisting of two objects with one designated as first, and the other as second. The notation $\langle x, y\rangle$ is used for the ordered pair having x as first element and y as second. The essential property of this notion is that $\langle x, y\rangle = \langle z, w\rangle$ if and only if $x = z$ and $y = w$.

We now define a (binary) *relation* as being a set whose elements are all ordered pairs. This formalises the intuitive idea of an association referred to earlier. If R is a relation (set of ordered pairs) and $\langle x, y\rangle \in R$ (sometimes written xRy) then we think of x being assigned to y by the association that R represents. For example the expression "is less than" establishes an association between numbers and determines the set

$$\{\langle x, y\rangle: x \text{ is less than } y\}.$$

Note that the pairs $\langle 1, 2\rangle$ and $\langle 1, 3\rangle$ both belong to this set, i.e. a relation may associate several objects to a given one.

From a function we obtain the relation

$$\hat{f} = \{\langle x, y\rangle: y \text{ is the } f\text{-image of } x\}.$$

To distinguish those relations that represent functions we have to incorporate the central feature of functions, namely that a given input produces one uniquely corresponding output. This means that each x can be the first element of only one of the ordered pairs in $\hat{f}$. That is

$$(*) \qquad \text{if} \quad \langle x, y\rangle \in \hat{f} \quad \text{and} \quad \langle x, z\rangle \in \hat{f}, \quad \text{then} \quad y = z.$$

This then is our set-theoretical characterisation of a function; as a set of ordered pairs satisfying the condition (∗). What happens next is a ploy often used in mathematics – a formal representation becomes an actual definition. It is quite common, in books at all levels, to find near the beginning a statement to the effect that "a function *is* a set of ordered pairs such that . . .".

How successful is this set-theoretical formulation of the function concept? Technically it works very well and allows an easy development of the theory of functions. But there are a number of rejoinders that can be made on the conceptual level. Some would say that the set $\hat{f}$ is not a function at all, but is the *graph* of the function f. The word of course comes from co-ordinate geometry. If we plot in the plane the points with co-ordinates of the form $\langle x, 6x\rangle$ we obtain a straight line (see figure) which is known as the *graph* of the function $f(x) = 6x$. This usage is carried over to more general contexts, particularly in subjects like topology and analysis, where writers often explicitly distinguish the function $f: A \to B$ from the *graph of* f as the set $\{\langle x, f(x)\rangle: x \in A\}$. Conflation of the two notions can easily lead to confusion.

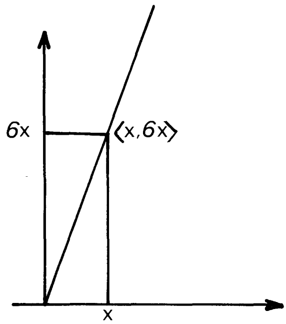


Fig. 2.2.

Another difficulty relates to the notion of codomain. Given a function f simply as a set of ordered pairs we can readily recover the domain (set of inputs) as the set

$$\operatorname{dom} f = \{x : \text{for some } y, \langle x, y\rangle \in f\}.$$

But what about the codomain of f? Recall that this can be any set that includes all the outputs of f. The outputs themselves form the so-called *range* or *image* of f, symbolically

$$\operatorname{Im} f = \{y : \text{for some } x, \langle x, y\rangle \in f\}.$$

In general f can be called a function from A to B whenever $A = \operatorname{dom} f$ and $\operatorname{Im} f \subseteq B$. Thus a function given simply as a set of ordered pairs does not have a uniquely determined codomain. This may seem a trifling point, but it leads to an interesting complication with the very important notion of *identity function.* This function is characterised by the rule $f(x) = x$, i.e. the output assigned to a given input is just that input itself. Each set A has its own identity function, called the *identity function on* A, denoted id_A, whose domain is the set A. Thus the image of id_A is also A, i.e. $\mathrm{id}_A : A \to A$. On the set-theoretic account, $\mathrm{id}_A = \{\langle x, x\rangle : x \in A\}$.

Now if A is a subset of a set B, then the rule $f(x) = x$ provides a function from A to B. In this case we talk of the *inclusion* function from A to B, for which we reserve the symbol $A \hookrightarrow B$. The use of a new word indicates a different intention. It conveys the sense of the function acting to include the elements of A amongst those of B. However even though the identity function on A and the inclusion map from A to B are conceptually quite different, as set-theoretical entities they are identical, i.e. exactly the same set of ordered pairs.

One way to cope with this point would be to modify the definition of function in the following way. Firstly for sets A and B we define the *product set* or *Cartesian product* of A and B to be the set of all ordered pairs whose first elements are in A and second elements in B. This is denoted $A \times B$, and so

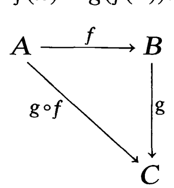
$$A \times B = \{\langle x, y\rangle : x \in A \text{ and } y \in B\}.$$

A function is now defined as a triple $f = \langle A, B, R\rangle$, where $R \subseteq A \times B$ is a relation from A to B (the *graph* of f), such that for each $x \in A$ there is one and only one $y \in B$ for which $\langle x, y\rangle \in R$. Thus the domain ($A$) and codomain ($B$) are incorporated in the definition of a function from the outset.

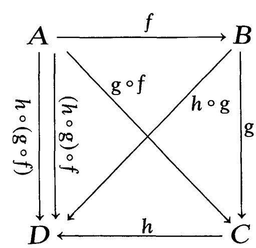
Although the modified definition does tidy things up a little it still presents a function as being basically a set of some kind – a fixed, static object. It fails to convey the "operational" or "transitional" aspect of the concept. One talks of "applying" a function to an argument, of a function "acting" on a domain. There is a definite impression of action, even of motion, as evidenced by the use of the arrow symbol, the source-target terminology, and commonly used synonyms for "function" like "transformation" and "mapping". The impression is analogous to that of a physical force acting on an object to move it somewhere, or replace it by another object. Indeed in geometry, transformations (rotations, reflections, dilations etc.) are functions that quite literally describe motion, while in applied mathematics forces are actually modelled as functions. This dynamical quality that we have been describing is an essential part of the meaning of the word "function" as it is used in mathematics. The "ordered-pairs" definition does not convey this. It is a formal set-theoretic *model* of the intuitive idea of a function, a model that captures an aspect of the idea, but not its full significance.

2.2. Composition of functions

Given two functions $f: A \to B$ and $g: B \to C$, with the target of one being the source of the other, we can obtain a new function by the rule "apply f and then g". For $x \in A$, the output $f(x)$ is an element of B, and hence an input to g. Applying g gives the element $g(f(x))$ of C. The passage from x to $g(f(x))$ establishes a function with domain A and codomain C. It is called the *composite of f and g*, denoted $g \circ f$, and symbolically defined by the rule $g \circ f(x) = g(f(x))$.



Now suppose we have three functions $f:A\to B$, $g:B\to C$, and $h:C\to D$ whose domains and codomains are so related that we can apply the three in succession to get a function from A to D. There are actually two ways to do this, since we can first form the composites $g\circ f:A\to C$ and $h\circ g:B\to D$. Then we follow either the rule "do f and then $h\circ g$", giving the function $(h\circ g)\circ f$, or the rule "do $g\circ f$ and then h", giving the composite $h\circ(g\circ f)$.



In fact these two functions are the same. When we examine their outputs we find that

$$[h\circ(g\circ f)](x)=h(g\circ f(x))=h(g(f(x))),$$

while

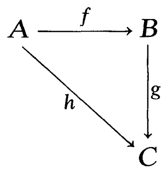
$$[(h\circ g)\circ f](x)=h\circ g(f(x))=h(g(f(x))).$$

Thus the two functions have the same domain and codomain, and they give the same output for the same input. They each amount to the rule "do f, and then g, and then h." In other words, they are the same function, and we have established the following.

Associative Law for Functional Composition. $h\circ(g\circ f)=(h\circ g)\circ f$.

This law allows us to drop brackets and simply write $h\circ g\circ f$ without ambiguity. Note that the law does not apply to *any* three functions – the equation only makes sense when they "follow a path", i.e. their sources and targets are arranged as described above.

The last figure is an example of the notion of *commutative diagram*, a very important aid to understanding used in category theory. By a diagram we simply mean a display of some objects, together with some arrows (here representing functions) linking the objects. The "triangle" of arrows f, g, h as shown is another diagram.



It will be said to *commute* if $h = g \circ f$. The point is that the diagram offers two paths from A to C, either by composing to follow f and then g, or by following h directly. Commutativity means that the two paths amount to the same thing. A more complex diagram, like the previous one, is said to be commutative when all possible triangles that are parts of the diagram are themselves commutative. This means that any two paths of arrows in the diagram that start at the same object and end at the same object compose to give the same overall function.

Composing with identities

What happens when we compose a function with an identity function? Given $f: A \to B$ we can follow f by id_B. Computing outputs we find, for $x \in A$, that

$$\mathrm{id}_B \circ f(x) = \mathrm{id}_B(f(x)) = f(x).$$

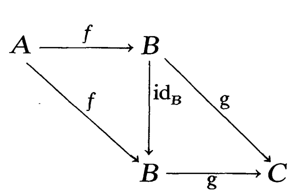
Similarly, given $g: B \to C$ we can precede g by id_B, in which case, for $x \in B$,

$$g \circ \mathrm{id}_B(x) = g(\mathrm{id}_B(x)) = g(x).$$

Since $\mathrm{id}_B \circ f$ and f have the same source and target, as do $g \circ \mathrm{id}_B$ and g, we have established the following.

IDENTITY LAW FOR FUNCTIONAL COMPOSITION. *For any* $f: A \to B$, $g: B \to C$, $\mathrm{id}_B \circ f = f$, *and* $g \circ \mathrm{id}_B = g$.

The Identity Law amounts to the assertion of the commutativity of the following diagram



2.3. Categories: First examples

We have already stated that a category can initially be conceived as a universe of mathematical discourse, and that such a universe is determined by specifying a certain kind of object and a certain kind of "function" between objects. The less suggestive word "arrow" is used in place of "function" in the general theory of categories (the word "morphism" is also used). The following table lists some categories by specifying their objects and arrows.

CATEGORY	OBJECTS	ARROWS
Set	all sets	all functions between sets
Finset	all finite sets	all functions between finite sets
Nonset	all nonempty sets	all functions between nonempty sets
Top	all topological spaces	all continuous functions between topological spaces
Vect	vector spaces	linear transformations
Grp	groups	group homomorphisms
Mon	monoids	monoid homomorphisms
Met	metric spaces	contraction maps
Man	manifolds	smooth maps
Top Grp	topological groups	continuous homomorphisms
Pos	partially ordered sets	monotone functions

In each of these examples the objects are sets with, apart from the first three cases, some additional structure. The arrows are all set functions which in each appropriate case satisfy conditions relating to this structure. It is not in fact vital that the reader be familiar with all of these examples. What is important is that she or he understands what they all have in common – what it is that makes each of them a category. The key lies, not in the particular nature of the objects or arrows, but in the way the arrows behave. In each case the following things occur;

(a) each arrow has associated with it two special objects, its *domain* and its *codomain*,

(b) there is an operation of *composition* that can be performed on certain pairs $\langle g, f\rangle$ of arrows in the category (when domain of $g =$ codomain of f) to obtain a new arrow $g \circ f$, *which is also in the category.*

(A composite of group homomorphisms is a group homomorphism, a composite of continuous functions between topological spaces is itself continuous etc.) This operation of composition always obeys the Associative Law described in the last section,

(c) each object has associated with it a special arrow in the category, the *identity arrow* on that object. (The identity function on a group is a group homomorphism, on a topological space is continuous etc). Within the category the identity arrows satisfy the Identity Law described in §2.2.

There are other features common to our list of examples. But as categories it is the two properties of associative composition and existence of identities that we single out for particular attention in the

Axiomatic Definition of a Category. A *category* $\mathscr{C}$ comprises

(1) a collection of things called *$\mathscr{C}$-objects*;

(2) a collection of things called *$\mathscr{C}$-arrows*;

(3) operations assigning to each $\mathscr{C}$-arrow f a $\mathscr{C}$-object $\operatorname{dom} f$ (the "domain" of f) and a $\mathscr{C}$-object $\operatorname{cod} f$ (the "codomain" of f). If $a = \operatorname{dom} f$ and $b = \operatorname{cod} f$ we display this as

$$f : a \to b \quad \text{or} \quad a \xrightarrow{f} b;$$

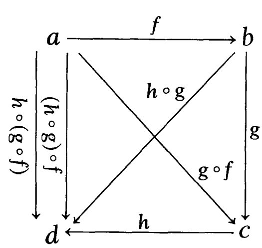
(4) an operation assigning to each pair $\langle g, f\rangle$ of $\mathscr{C}$-arrows with $\operatorname{dom} g = \operatorname{cod} f$, a $\mathscr{C}$-arrow $g \circ f$, *the composite of f and g*, having $\operatorname{dom}(g \circ f) = \operatorname{dom} f$ and $\operatorname{cod}(g \circ f) = \operatorname{cod} g$, i.e. $g \circ f : \operatorname{dom} f \to \operatorname{cod} g$, and such that the following condition obtains:

Associative Law: Given the configuration

$$a \xrightarrow{f} b \xrightarrow{g} c \xrightarrow{h} d$$

of $\mathscr{C}$-objects and $\mathscr{C}$-arrows then $h \circ (g \circ f) = (h \circ g) \circ f$.

The associative law asserts that a diagram having the form



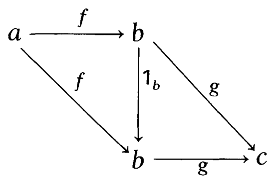
always commutes;

(5) an assignment to each $\mathscr{C}$-object b of a $\mathscr{C}$-arrow $1_b : b \to b$, called *the identity arrow on b*, such that

Identity Law: For any $\mathscr{C}$-arrows $f : a \to b$ and $g : b \to c$

$$1_b \circ f = f, \quad \text{and} \quad g \circ 1_b = g$$

i.e. the diagram



commutes.

2.4. The pathology of abstraction

The process we have just been through in identifying the notion of a category is one of the basic modi operandi of pure mathematics. It is called *abstraction*. It begins with the recognition, through experience and examination of a number of specific situations, that certain phenomena occur repeatedly, that there are a number of common features, that there are formal analogies in the behaviour of different entities. Then comes the actual *process* of abstraction, wherein these common features are singled out and presented in isolation; an axiomatic description of an "abstract" concept. This is precisely how we obtained our general definition of a category from an inspection of a list of particular categories. It is the same process by which all of the abstract structures that mathematics investigates (group, vector space, topological space etc) were arrived at.

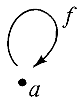
Having obtained our abstract concept we then develop its general theory, and seek further instances of it. These instances are called *examples* of the concept or *models* of the axioms that define the concept. Any statement that belongs to the general theory of the concept (i.e. is derivable from the axioms) will hold true in all models. The search for new models is a process of specialisation, the reverse of abstraction. Progress in understanding comes as much from the recognition that a particular new structure is an instance of a more general phenomenon, as from the recognition that several different structures have a common core. Our knowledge of mathematical reality advances through the interplay of these two processes, through movement from the particular to the general and back again. The procedure is well illustrated, as we shall see, by the development of topos theory.

An important aspect of specialisation concerns so-called *representation* theorems. These are propositions to the effect that any model of the axioms for a certain abstract structure must be (equivalent to) one of a particular list of concrete models. They "measure" the extent to which the original motivating examples encompass the possible models of the general notion. Thus we know (Cayley's Theorem) that any group can be thought of as being a group of permutations of some set, while any Boolean algebra is essentially an algebra of subsets of some set. Roughly speaking, the stronger the abstraction, i.e. the more we put into the abstract concept, the fewer will be the possible examples. The extreme case is where there is only one model. A classic example of this is the axiomatically presented concept of a complete ordered field. There is in fact only one such field, viz the real number system.

The category axioms represent a very weak abstraction. There is no representation theorem in terms of our original list. We talked at the outset of "general universes of mathematical discourse". However we have picked out only the bare bones of our initial examples, and so little of the flesh that the axioms admit of all sorts of "pathological" cases that differ wildly in appearance from **Set**, **Top**, **Vect** etc. One readily finds categories that are not universes of discourse at all, in which the objects are not sets, the arrows look nothing like functions, and the operation $\circ$ has nothing to do with functional composition. The following list includes a number of such categories. The reader is urged to examine these closely, to fill out the details of their definition, and to check that in each case the Associative and Identity axioms are satisfied.

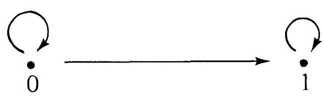
2.5. Basic examples

EXAMPLE 1. **1**: This category is to have only one object, and one arrow. Having said that, we find that its structure is completely determined. Suppose we call the object a, and the arrow f. Then we must put $\operatorname{dom} f = \operatorname{cod} f = a$, as a is the only available object. Since f is the only arrow, we have to take it as the identity arrow on a, i.e. we put $1_a = f$. The only composable pair of arrows is $\langle f, f \rangle$, and we put $f \circ f = f$. This gives the identity law, as $1_a \circ f = f \circ 1_a = f \circ f = f$, and the associative law holds as $f \circ (f \circ f) = (f \circ f) \circ f = f$. Thus we have a category, which we display diagramatically as



We did not actually say what a and f are. The point is that they can be anything you like. a might be a set, with f its identity function. But f might be a number, or a pair of numbers, or a banana, or the Eiffel tower, or Richard Nixon. Likewise for a. Just take any two things, call them a and f, make the above *definitions* of $\operatorname{dom} f$, $\operatorname{cod} f$, $\mathbf{1}_a$, and $f \circ f$, and you have produced a structure that satisfies the axioms for a category. Whatever a and f are, the category will look like the above diagram. In this sense there is "really" only one category that has one object and one arrow. We give it the name **1**. As a paradigm description of it we might as well take the object to be the number 0, and the arrow to be the ordered pair $\langle 0, 0\rangle$.

Example 2. **2**: This category has two objects, three arrows, and looks like



We take the two objects to be the numbers 0 and 1. For the three arrows we take the pairs $\langle 0, 0\rangle$, $\langle 0, 1\rangle$, and $\langle 1, 1\rangle$, putting

$$\langle 0, 0\rangle : 0 \to 0$$

$$\langle 0, 1\rangle : 0 \to 1$$

$$\langle 1, 1\rangle : 1 \to 1$$

Thus we must have

$$\langle 0, 0\rangle = \mathbf{1}_0 \quad \text{(the identity on 0)}$$

and

$$\langle 1, 1\rangle = \mathbf{1}_1.$$

There is only one way to define composition for this set up:

$$\mathbf{1}_0 \circ \mathbf{1}_0 = \mathbf{1}_0$$

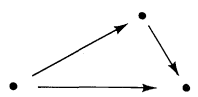
$$\langle 0, 1\rangle \circ \mathbf{1}_0 = \langle 0, 1\rangle$$

$$\mathbf{1}_1 \circ \langle 0, 1\rangle = \langle 0, 1\rangle$$

and

$$1_1 \circ 1_1 = 1_1$$

EXAMPLE 3. **3**: This category has three objects and six arrows, the three non-identity arrows being arranged in a triangle thus:



Again there is only one possible way to define composites.

EXAMPLE 4. *Preorders in general.* In each of our first three examples there is only one way that composites can be defined. This is because between any two objects there is never more than one arrow, so once the dom and cod are known, there is no choice about what the arrow is to be. In general a category with this property, that between any two objects p and q there is *at most one* arrow $p \to q$, is called a *pre-order.* If P is the collection of objects of a pre-order category then we may define a binary relation R on P (i.e. a set $R \subseteq P \times P$) by putting

> $\langle p, q \rangle \in R$ iff there is an arrow $p \to q$ in the pre-order category.

The relation R then has the following properties (writing "pRq" in place of "$\langle p, q \rangle \in R$"); it is

(i) *reflexive,* i.e. for each p we have pRp, and

(ii) *transitive,* i.e. whenever pRq and qRs, we have pRs.

(Condition (i) holds as there is always the identity arrow $p \to p$, for any p. For (ii), observe that an arrow from p to q composes with one from q to s to give an arrow from p to s).

A binary relation that is reflexive and transitive is commonly known as a *pre-ordering.* We have just seen that a pre-order category has a natural pre-ordering relation on its collection of objects (hence its name of course). Conversely if we start simply with a set P that is pre-ordered by a relation R (i.e. $R \subseteq P \times P$ is reflexive and transitive) then we can obtain a pre-order category as follows. The objects are the members p of P. The arrows are the pairs $\langle p, q \rangle$ for which pRq. $\langle p, q \rangle$ is to be an arrow from p to q. Given a composable pair

$$p \xrightarrow{\langle p,q \rangle} q \xrightarrow{\langle q,s \rangle} s,$$

we put

$$\langle q, s\rangle \circ \langle p, q\rangle = \langle p, s\rangle.$$

Note that if $\langle p, q\rangle$ and $\langle q, s\rangle$ are arrows then pRq and qRs, so pRs (transitivity) and hence $\langle p, s\rangle$ is an arrow. There is at most one arrow from p to q, depending on whether or not pRq, and by transitivity there is only one way to compose arrows. By reflexivity, $\langle p, p\rangle$ is always an arrow, for any p, and indeed $\langle p, p\rangle = 1_p$.

Examples 1–3 are pre-orders whose associated pre-ordering relation R satisfies a further condition, viz it is

(iii) *antisymmetric*, i.e. whenever pRq and qRp, we have

$$p = q.$$

An antisymmetric pre-ordering is called a *partial ordering*. The symbol "$\sqsubseteq$" will generally be used for this type of relation, i.e. we write $p \sqsubseteq q$ in place of pRq. A *poset* is a pair $\mathbf{P} = \langle P, \sqsubseteq\rangle$, where P is a set and $\sqsubseteq$ is a partial ordering on P. These structures will play a central role in our study of topoi.

The set $\{0\}$ becomes a poset when we put $0 \sqsubseteq 0$. The corresponding pre-order category is **1** (Example 1). The pre-order **2** corresponds to the partial ordering on the set $\{0, 1\}$ that has $0 \sqsubseteq 1$ (and of course $0 \sqsubseteq 0$ and $1 \sqsubseteq 1$). This is the usual numerical ordering, $\leqslant$, of the numbers 0 and 1 (where "$\leqslant$" means "less than or equal to"). The category **3** corresponds to the usual ordering on the three element set $\{0, 1, 2\}$. We could continue this process indefinitely, constructing a pre-order **4** from the usual ordering on $\{0, 1, 2, 3\}$, and in general for each natural number n, a pre-order $\mathbf{n}$ from the usual ordering on $\{0, 1, 2, \ldots, n-1\}$. Continuing even further we can consider the infinite collection

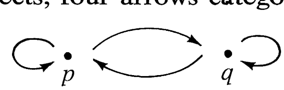
$$\omega = \{0, 1, 2, 3, \ldots\}$$

of all natural numbers under the usual ordering, to obtain a pre-order category which has the diagram

$$0 \to 1 \to 2 \to 3 \to \ldots$$

(composites and identities not shown).

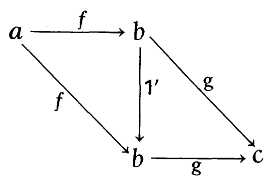
A simple example of a pre-order that is not partially ordered would be a two-objects, four-arrows category



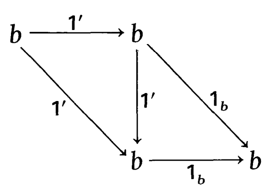
which has pRq and qRp, but $p \neq q$.

A categorial expression of the antisymmetry condition will be given in the next chapter, while the above numerical examples will be reconsidered in Example 9.

EXAMPLE 5. *Discrete categories.* If b is an object of a category $\mathscr{C}$, then the $\mathscr{C}$-arrow 1_b is uniquely determined by the property expressed in the Identity Law. For if $1': b \to b$ has the property that



commutes for any $\mathscr{C}$-arrows f and g as shown, then in the particular case of $f = 1'$ and $g = 1_b$,



commutes giving $1_b = 1_b \circ 1'$ (right triangle). But by the Identity Law (with $f = 1'$), $1_b \circ 1' = 1'$, and so $1_b = 1'$.

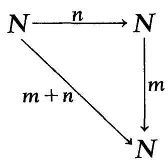
Since 1_b is thus uniquely determined, the practice is sometimes adopted of identifying the object b with the arrow 1_b and writing $b : b \to b$, $b \circ f$ etc. Now the category axioms require that the $\mathscr{C}$-arrows include, at a minimum, an identity arrow for each $\mathscr{C}$-object (why must distinct objects have distinct identity arrows?). $\mathscr{C}$ is a *discrete* category if these are the only arrows, i.e. every arrow is the identity on some object. A discrete category is a pre-order since, as we have just seen, there can only be one identity arrow on a given object. Equating objects with identity arrows, we see that a discrete category is really nothing more than a collection of objects. Indeed, any set X can be made into a discrete category by adding an identity arrow $x \to x$ for each $x \in X$, i.e. X becomes the pre-order corresponding to the relation $R \subseteq X \times X$ that has

$$xRy \quad \text{iff} \quad x = y.$$

EXAMPLE 6. **N**: It is time we looked at some categories that have more than one arrow between given objects. The present example has only one object, which we shall call N, but an infinite collection of arrows from N to N. The arrows are, by definition, the natural numbers 0, 1, 2, 3, ... Each arrow has the same dom and cod, viz the unique object N. This means that all pairs of arrows are composable. The composite of two arrows (numbers) m and n is to be another number. The definition is

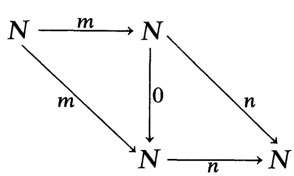
$$m \circ n = m + n.$$

Thus the diagram



commutes by definition. The associative law is satisfied, since addition of numbers is an associative operation, i.e., $m+(n+k)=(m+n)+k$ is true for any numbers m, n and k.

The identity arrow 1_N on the object N is defined to be the number 0. The diagram



commutes because $0+m=m$ and $n+0=n$.

EXAMPLE 7. *Monoids.* The category **N** of the last example *is* a category because the structure $(N, +, 0)$ is an example of the abstract algebraic concept of a *monoid.*

A monoid is a triple $\mathbf{M}=(M, *, e)$ where

(i) M is a set

(ii) $*$ is a binary operation on M, i.e. a function from $M \times M$ to M assigning to each pair $\langle x, y\rangle \in M \times M$ an element $x * y$ of M, that is *associative*, i.e. satisfies $x * (y * z)=(x * y) * z$ for all $x, y, z \in M$.

(iii) e is a member of M, the monoid *identity*, that satisfies $e * x = x * e = x$, for all $x \in M$.

Any monoid **M** gives rise to a category with one object, exactly as in Example 6. We take the object to be M the arrows $M \to M$ to be the members of M, and put $e = 1_M$. Composition of arrows $x, y \in M$ is given by

$$x \circ y = x * y.$$

Conversely, if $\mathscr{C}$ is a category with only one object a, and M is its collection of arrows, then $(M, \circ, 1_a)$ is a monoid. All arrows have the same dom and cod and so all pairs are composable. Hence composition $\circ$ is a function from $M \times M$ to M, i.e. a binary operation on M, that is associative by the Associative Law for categories. 1_a is an identity for the monoid by the Identity Law for categories.

EXAMPLE 8. **Matr(K)** (for linear algebraists). If **K** is a commutative ring then the matrices over **K** yield a category **Matr(K)**. The objects are the positive integers 1, 2, 3, ... An arrow $m \to n$ is an $n \times m$ matrix with entries in **K**. Given composable arrows

$$m \xrightarrow{B} n \xrightarrow{A} p,$$

i.e. A a $p \times n$ matrix and B $n \times m$, *we define* $A \circ B$ to be the matrix product AB *of* A and B (which is $p \times m$ and hence an arrow $m \to p$). The Associative Law is given by the associativity of matrix multiplication. 1_m is the identity matrix of order m.

In the remainder of this chapter we consider ways of forming new categories from given ones.

EXAMPLE 9. *Subcategories.* If $\mathscr{C}$ is a category, and a and b are $\mathscr{C}$-objects, we introduce the symbol $\mathscr{C}(a, b)$ to denote the collection of all $\mathscr{C}$-arrows with dom $= a$ and cod $= b$, i.e.

$$\mathscr{C}(a, b) = \left\{ f : f \text{ is a } \mathscr{C}\text{-arrow and } a \xrightarrow{f} b \right\}.$$

$\mathscr{C}$ is said to be a *subcategory* of category $\mathscr{D}$, denoted $\mathscr{C} \subseteq \mathscr{D}$, if

(i) every $\mathscr{C}$-object is a $\mathscr{D}$-object, and

(ii) if a and b are any two $\mathscr{C}$-objects, then $\mathscr{C}(a, b) \subseteq \mathscr{D}(a, b)$, i.e. all the $\mathscr{C}$-arrows $a \to b$ are present in $\mathscr{D}$.

For example, we have **Finset** $\subseteq$ **Set**, and **Nonset** $\subseteq$ **Set**, although neither of **Finset** and **Nonset** are subcategories of each other.

$\mathscr{C}$ is a *full* subcategory of $\mathscr{D}$ if $\mathscr{C} \subseteq \mathscr{D}$, and

(iii) for any $\mathscr{C}$-objects a and b, $\mathscr{C}(a, b) = \mathscr{D}(a, b)$, i.e. $\mathscr{D}$ has no arrows $a \to b$ other than the ones already in $\mathscr{C}$.

If $\mathscr{D}$ is a category and C is any collection of $\mathscr{D}$-objects we obtain a full subcategory $\mathscr{C}$ of $\mathscr{D}$ by taking as $\mathscr{C}$-arrows all the $\mathscr{D}$-arrows between members of C. Thus we see that **Finset** and **Nonset** are each full subcategories of **Set**.

An important full subcategory of **Finset** (and hence of **Set**) is the category **Finord** of all finite *ordinals*. The finite ordinals are sets that are used in set-theoretic foundations as representations of the natural numbers. We use the natural numbers as names for these sets and put

0 for $\emptyset$ (the empty set)
1 for $\{0\}$ $(=\{\emptyset\})$
2 for $\{0, 1\}$ $(=\{\emptyset, \{\emptyset\}\})$
3 for $\{0, 1, 2\}$ $(=\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\})$
4 for $\{0, 1, 2, 3\}$

and so on.

Proceeding "inductively", where n is a natural number, we put

$$n \quad \text{for} \quad \{0, 1, 2, \ldots, n-1\}.$$

The sequence of finite sets thus generated are the finite ordinals. They form the objects of the category **Finord**, whose arrows are all the set functions between finite ordinals.

Of course it is ridiculous to suggest that the number 1 *is* the set $\{0\}$ whose only member is the null set. The point is that in axiomatic set theory, where we seek an explicit and precise account of mathematical entities and their intuitively understood properties, the finite ordinals provide such a paradigmatic representation of the natural numbers. They have an intricate and elegant structure that exhibits all the arithmetic and algebraic properties of the natural number system. They are related by set inclusion and set membership as follows:

$$0 \subseteq 1 \subseteq 2 \subseteq 3 \subseteq \ldots$$

$$0 \in 1 \in 2 \in 3 \in \ldots$$

In fact the following three statements are equivalent

(a) $n < m$ (the number n is numerically less than the number m)
(b) $n \subset m$ (the set n is a proper subset of set m)
(c) $n \in m$ (n is a member of set m)

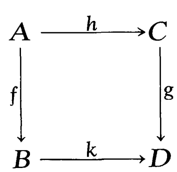
Thus $n \leqslant m$ iff $n \subseteq m$.

So the ordinal (set) $n = \{0, 1, \ldots, n-1\}$ has the ordering $\leqslant$ built into its structure in a natural set-theoretic way. The corresponding pre-order category is none other than **n** of Example 4. Notice that if $n \leqslant m$, the pre-order **n** is a full subcategory of **m**.

Example 10. **Product categories.** The category $\mathbf{Set}^2$ of pairs of sets has as objects all pairs $\langle A, B\rangle$ of sets. An arrow in $\mathbf{Set}^2$ from $\langle A, B\rangle$ to $\langle C, D\rangle$ is a pair $\langle f, g\rangle$ of set functions such that $f: A \to C$ and $g: B \to D$. Composition is defined by $\langle f, g\rangle \circ \langle f', g'\rangle = \langle f \circ f', g \circ g'\rangle$, where $f \circ f'$ and $g \circ g'$ are the functional compositions. The identity arrow on $\langle A, B\rangle$ is $\langle \mathrm{id}_A, \mathrm{id}_B\rangle$.

This construction generalises: given any two categories $\mathscr{C}$ and $\mathscr{D}$, the product category $\mathscr{C} \times \mathscr{D}$ has objects the pairs $\langle a, b\rangle$ where a is a $\mathscr{C}$-object and b a $\mathscr{D}$-object. A $\mathscr{C} \times \mathscr{D}$-arrow $\langle a, b\rangle \to \langle c, d\rangle$ is a pair $\langle f, g\rangle$ where $f: a \to c$ is a $\mathscr{C}$-arrow and $g: b \to d$ a $\mathscr{D}$-arrow. Composition is defined "componentwise" with respect to composition in $\mathscr{C}$, and composition in $\mathscr{D}$.

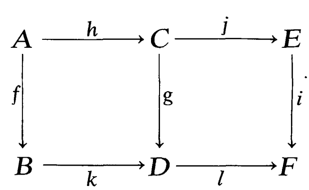
Example 11. **Arrow categories**. The category $\mathbf{Set}^{\to}$ of functions has as objects the set functions $f: A \to B$. An arrow in $\mathbf{Set}^{\to}$ from the $\mathbf{Set}^{\to}$-object $f: A \to B$ to the $\mathbf{Set}^{\to}$-object $g: C \to D$ is a pair of functions $\langle h, k\rangle$ such that



commutes, i.e. $g \circ h = k \circ f$.

For composition we put

$$\langle j, l\rangle \circ \langle h, k\rangle = \langle j \circ h, l \circ k\rangle$$

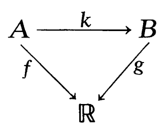


The identity arrow for the $\mathbf{Set}^{\to}$-object $f: A \to B$ is the function pair $\langle \mathrm{id}_A, \mathrm{id}_B\rangle$.

This construction can also be generalised to form, from any category $\mathscr{C}$, the *arrow category* $\mathscr{C}^{\to}$ whose objects are all the $\mathscr{C}$-arrows.

EXAMPLE 12. **Comma categories.** These can be thought of as specialisations of arrow categories, where we restrict attention to arrows with fixed domain or codomain.

Thus if $\mathbb{R}$ is the set of real numbers, we obtain the category **Set** $\downarrow \mathbb{R}$ of *real valued functions*. The objects are all functions $f: A \to \mathbb{R}$ that have codomain $\mathbb{R}$. An arrow from $f: A \to \mathbb{R}$ to $g: B \to \mathbb{R}$ is a function $k: A \to B$ that makes the triangle

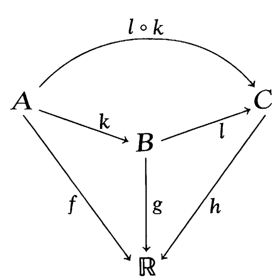


commute, i.e. has $g \circ k = f$.

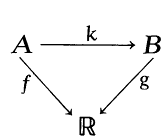
It is sometimes convenient to think of **Set** $\downarrow \mathbb{R}$-objects as pairs (A, f), where $f: A \to \mathbb{R}$. Then the **Set** $\downarrow \mathbb{R}$ composite of

$$(A, f) \xrightarrow{k} (B, g) \xrightarrow{l} (C, h)$$

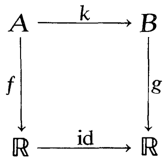
is defined as $l \circ k: (A, f) \to (C, h)$



The identity arrow on the object $f: A \to \mathbb{R}$ is $\mathrm{id}_A: (A, f) \to (A, f)$. **Set** $\downarrow \mathbb{R}$ is not as it stands a subcategory of $\mathbf{Set}^{\to}$ as the two have different sorts of arrows. However, we could equate the **Set** $\downarrow \mathbb{R}$ arrow $k: (A, f) \to (B, g)$ with the $\mathbf{Set}^{\to}$ arrow $\langle k, \mathrm{id}_{\mathbb{R}} \rangle$, as



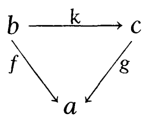
commutes iff



does.

In this way **Set** $\downarrow \mathbb{R}$ can be "construed" as a (not full) subcategory of **Set**$^{\rightarrow}$.

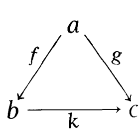
Similarly for any set X we obtain the category **Set** $\downarrow X$ of "X-valued functions". More generally if $\mathscr{C}$ is any category, and a any $\mathscr{C}$-object then the category $\mathscr{C} \downarrow a$ of *objects over a* has the $\mathscr{C}$-arrows with codomain a as objects, and as arrows from $f: b \rightarrow a$, to $g: c \rightarrow a$ the $\mathscr{C}$-arrows $k: b \rightarrow c$ such that



commutes, i.e. $g \circ k = f$.

Categories of this type are going to play an important role both in the provision of examples of topoi, and in the development of the general theory.

Turning our attention to domains, we define the category $\mathscr{C} \uparrow a$ of *objects under a* to have as objects the $\mathscr{C}$-arrows with dom $= a$ and as arrows from $f: a \rightarrow b$ to $g: a \rightarrow c$ the $\mathscr{C}$-arrows $k: b \rightarrow c$ such that



commutes, i.e. $k \circ f = g$.

Categories of the type $\mathscr{C} \downarrow a$ and $\mathscr{C} \uparrow a$ are known as *comma* categories.