Generalised Abstract Nonsense

A Short Introduction to the Theory of Categories

—

MA247 Mathematical Excursions

Jonathan Elliott
0511028

1 Introduction

It is often the case that similar concepts and theorems appear in different forms in different branches of mathematics. For instance, in the context of group theory, two groups are called isomorphic if there is a bijection between them that preserves the group structure; a similar idea appears in topology, where two topological spaces are said to be homeomorphic if there is a homeomorphism between them, i.e. a continuous bijection whose inverse is also continuous. As another example, the Cartesian product of two sets and the direct product of two groups share many of the same properties.

Category theory aims to describe these similarities in a more general setting, by allowing us to remove ourselves from the actual mathematical entities involved and concentrate on the way they interrelate with each other. We ignore the internal structure of the objects (as they are known) and instead focus on the "functions" between them, which we call arrows (or morphisms). At its highest level, the degree of abstraction can sometimes become extreme, earning category theory the nickname "generalised abstract nonsense", even among its initiators [**15**]; hence the title of this essay.

We first define a category:

Definition 1.1. A *category* $\mathbf{C}$ is a collection $\operatorname{Ob}\mathbf{C}$ of *objects* $A, B, C, \ldots$, together with a collection $\operatorname{Mor}_{\mathbf{C}}(A, B)$ of maps $f, g, h, \ldots$ for each pair A, B of objects in $\operatorname{Ob}\mathbf{C}$, which we call *arrows* (or *morphisms*), such that the following conditions are satisfied:

- For every arrow f in $\mathbf{C}$ we can assign an object $\operatorname{dom} f$, its *domain*, and an object $\operatorname{cod} f$, its *codomain*. If $f \in \operatorname{Mor}_{\mathbf{C}}(A, B)$ we say $\operatorname{dom} f = A$ and $\operatorname{cod} f = B$ and we write $f\colon A \to B$.
- For any pair of arrows f and g with $\operatorname{cod} f = \operatorname{dom} g$ there is a *composite* arrow $g \circ f\colon \operatorname{dom} f \to \operatorname{cod} g$ such that the following associative law holds for all arrows $f\colon A \to B$, $g\colon B \to C$ and $h\colon C \to D$, where A, B, C, D are (not necessarily distinct) objects in $\mathbf{C}$:

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

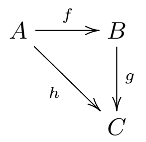
- For every object A in $\mathbf{C}$ there is an *identity* arrow $\mathrm{id}_A\colon A \to A$ such that for all objects A, B in $\mathbf{C}$ and all arrows $f\colon A \to B$ we have $f \circ \mathrm{id}_A = f = \mathrm{id}_B \circ f$.

(If $A \in \operatorname{Ob}\mathbf{C}$ then we may sometimes refer loosely to A as an object "in $\mathbf{C}$". In some cases it is useful to call A a $\mathbf{C}$-*object* to avoid confusion when more than one category is being discussed. Similarly, if $f \in \operatorname{Mor}_{\mathbf{C}}(A, B)$ for some $A, B \in \operatorname{Ob}\mathbf{C}$ then we call f a $\mathbf{C}$-*arrow*.)

1.1 Diagrams

Category theory makes substantial use of so-called "commutative diagrams" and many of the concepts relating to categories are expressed most clearly using them; in fact, the technique of "diagram chasing" can make many category theoretical proofs appear trivial. A diagram is a collection of nodes and edges which we label with objects and arrows (in a consistent way, so that, for instance, if $f\colon A \to B$ is an arrow from an object A to an object B then we may label the edge between two nodes A and B as "f".) A diagram may have many different edges between any two nodes, or none at all.

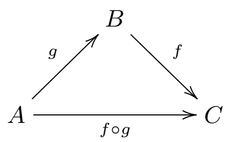
A diagram is said to *commute* if, for every pair of vertices A and B, the same result is achieved regardless of which path from A to B we choose to take. For example, the diagram



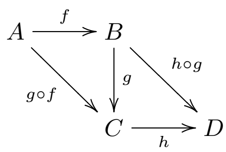
commutes if and only if h gives the same result in general as f followed by g.

The axioms of definition 1.1 can be stated succinctly by saying that the following diagrams commute for any choice of objects A, B, C, D in **C** and any arrows f, g, h between them.

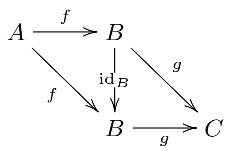
- Composability:



- Associativity:



- Identity:



1.2 Examples of Categories

Now that we have defined a category, let us look at some examples. (Further examples can be found in almost any textbook on category theory, for instance Mac Lane [**8**], Pierce [**13**], Pareigis [**12**], Mitchell [**11**], Arbib and Manes [**2**], Herrlich and Strecker [**7**], Bucur and Deleanu [**4**] and many others.)

- Perhaps the most obvious candidate for a category is the collection **Set** with sets as objects and set functions as arrows. The category axioms are easily satisfied, as shown below.
  - An object in **Set** is a set and an arrow $f\colon A \to B$ in **Set** is a function f from A to B.
  - For a function $f\colon A \to B$ we have $\operatorname{dom} f = A$ and $\operatorname{cod} f = B$ and write $f \in \mathrm{Mor}_{\mathbf{C}}(A, B)$.
  - The composition of two arrows f, g in **Set** with $\operatorname{cod} f = \operatorname{dom} g$ is just the normal set-theoretic composition of functions, defined by $(f \circ g)(x) = f(g(x))$ for every $x \in \operatorname{dom} g$, and composition is associative.

– For each set A the identity function id_A has $\mathrm{dom}\,\mathrm{id}_A = \mathrm{cod}\,\mathrm{id}_A = A$ and clearly $f \circ \mathrm{id}_A = f$ and $g = \mathrm{id}_A \circ g$ for any functions $f\colon A \to B$ and $g\colon C \to A$.

(There is an important subtlety to be aware of in our definition of arrows in **Set**: namely, codomains matter! The arrow $x \mapsto x^2\colon \mathbb{R} \to \mathbb{R}$ in **Set** is distinct from the arrow $x \mapsto x^2\colon \mathbb{R} \to \mathbb{R}^+$, where $\mathbb{R}^+$ is the set of non-negative real numbers. In other words, each function on sets correponds to many different arrows in **Set**.)

The astute reader may have noticed an apparent conflict between this example and our original definition of a category, which talked about such things as a collections of objects. However, we are not defining sets in terms of categories here, so there is no circularity in our definitions and the set-theoretic definition of a category given above is mainly a convenience. Mac Lane [**8**] actually first defines a *metacategory* as anything satisfying the axioms of definition 1.1 and a *category* as any interpretation of this within set theory. Categories can in fact be defined in a purely algebraic way, as in the original paper on category theory [**6**]. (For a discussion of the various alternative definitions of a category see Marquis [**9**].)

It is worth noting that a naive set-theoritic formulation of category theory can, however, give rise to certain difficulties relating to the apparent problem of talking about arbitrary collections of objects like sets. Russell's Paradox[1] shows us that we cannot just talk about notions such as "the sets of all sets" without leading to contradictions. To avoid such paradoxes we adopt the approach of axiomatic set theory, wherein sets are an example of a more general object called *classes*. If a class X is an element of some other class then X is called a *set*. Otherwise X is called a *proper class*. The "set of all sets" is actually a proper class.[2]

Here is another example of a category based on a familiar mathematical concept.

- The collection of vector spaces over a field K and the linear maps between these vector spaces form respectively the objects and arrows a category $\mathbf{Vect}_K$. Verification of the axioms requires a little more work than for **Set**, namely showing that the composition of linear maps is actually another linear map, but is not difficult.

There is any number of similar categories, whose objects are some kinds of "sets with structure" and whose arrows are the structure-preserving maps between these objects. We list some possibilities:

- Groups and group homomorphisms between groups form a category **Grp**.
- Rings and ring homomorphisms between rings form a category **Rng**.
- Topological spaces and continuous maps between topological spaces form a category **Top**.

There are many more examples, some of which require more mathematics than this essay assumes[3], as well as many which can be formed from more basic categories by imposing more

[1] In 1901 Bertrand Russell announced the following contradiction that arises in naive set theory. Consider the set of all sets that do not contain themselves:

$$M = \{A \mid A \notin A\}$$

Is $M \in M$? If yes, then it must satisfy the property that it does not contain itself, contradicting the fact that it does. If no, then $M \notin M$, which because of the definition of M implies that $M \in M$, another contradiction.

[2] We say a category $\mathbf{C}$ is *small* if both its collections of objects and arrows are sets. If for any two objects A and B the collection $\mathrm{Mor}_{\mathbf{C}}(A, B)$ of arrows from A to B is a set, then $\mathbf{C}$ is *locally small*. This is different from asking that the collection of *all* arrows in $\mathbf{C}$ be a set. Otherwise we say $\mathbf{C}$ is *large*. It turns out that some theorems from category theory only apply to small categories, but the foundational issues can become quite complex — for more information see McLarty [**10**], Ch. 12.

[3] For instance, $\mathbf{Mod}_R$, the category of (right-)modules over a ring R, **Diff**, the category of differentiable manifolds and smooth maps and **HoTop**, the category of topological spaces with equivalence classes of homotopic functions as arrows. The last of these is an important example of a category whose arrows are not just structure-preserving maps [**9**].

restrictions. For example...

- **AbGrp**, the category of abelian groups and group homomorphisms.[4]
- **Met**, the category of metric spaces and continuous maps.
- **Haus**, the category of topological spaces satisfying the Hausdorff property[5] and continuous maps.
- **Mon**, the category of monoids and monoid homomorphisms (see definition 1.3).
- **Poset**, the category of partially ordered sets (posets) and order-preserving (monotone) functions (see definition 1.2).

1.3 More abstract examples

So far all our examples have been cases of what are known as "concrete categories", based on collections of familiar mathematical objects and the appropriate structure-preserving maps between them.

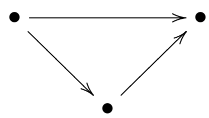
However, as mentioned earlier, it is entirely possible to talk about completely abstract categories where the objects and arrows do not correspond to any particular mathematical objects; we simply require that the axioms of definition 1.1 are satisfied.

Examples.

1. The category **0** is the empty category with no objects and no arrows.
2. The category **1** has one object. The only arrow is the identity arrow, which by convention we do not draw:

3. The category **2** has two objects and one non-identity arrow between these objects:

4. The category **3** has three objects with the non-identity arrows arranged in a triangle as shown below:



Note that there is only one way in which the arrows could be arranged for the axioms of definition 1.1 to be satisfied.

There is a third type of category that we will consider, where we think of single mathematical objects as categories in their own right. This is perfectly valid, as long as we define our objects, arrows and composition in such a way that the axioms of definition 1.1 are satisfied. This abstractness is typical of the category theoretic approach; the more we generalise, the more of the "bigger picture" we see.

[4]When it is obvious from context, "the category of Xs and Ys" is used to refer to the category whose objects are Xs and whose arrows are Ys between Xs.

[5]A topological space T has the Hausdorff property if any two distinct points of T can be contained inside disjoint open sets. In particular, every metric space has the Hausdorff property: simply take open balls around each point with radius equal to half the distance between them.

Examples.

1. We give the following definition:

   Definition 1.2. A *preorder* is an ordered pair $(P, \leqslant_P)$ consisting of a set P and a binary relation $\leqslant_P$ defined on P satisfying

   - Reflexivity: For all $p \in P$, $p \leqslant_P p$;
   - Transitivity: For all $p, q, r \in P$, $p \leqslant_P q$ and $q \leqslant_P r \implies p \leqslant_P r$.

   If $(P, \leqslant_P)$ satisfies a third condition

   - Antisymmetry: For all $p, q \in P$, $p \leqslant_P q$ and $q \leqslant_P p \implies p = q$.

   then we call $(P, \leqslant_P)$ a *partially ordered set* (or *poset* for short).

   We can show that the collection of all partially ordered sets together with order-preserving mappings[6] form a category **Poset**, as mentioned earlier, but we can also do something else interesting:

   An individual poset can be considered as a category whose objects are the elements of the poset. We draw an arrow between two objects p and q if $p \leqslant_P q$. If this is not the case, we draw no arrow between them.

   Composition of arrows is guaranteed by transitivity and is associative; the existence of identity arrows for each object corresponds to the axiom of reflexivity. Note that antisymmetry is not actually needed, so in fact any preorder can be considered as a category.

2. We give another definition:

   Definition 1.3. A *monoid* is an ordered triple $(M, *, e)$ consisting of a set M, a binary operation $*\colon M \times M \to M$ and a distinguished element $e \in M$ satisfying

   - Associativity: For all $x, y, z \in M$, $x * (y * z) = (x * y) * z$;
   - Identity: For all $x \in M$, $x * e = e * x = x$.

   Notice that a monoid is a group where elements do not necessarily have an inverse. Alternatively, a monoid can be thought of as a semigroup[7] that has an identity element; in particular a semigroup may be empty, while a monoid may not [**14**].

   It is clear that there is a category **Mon** where the objects are monoids and the arrows are monoid homomorphisms[8], but this is not all.

   We can consider a monoid as a category with one object. The arrows in this category represent the elements of the monoid; composition of arrows corresponds to the binary operation $*$ and is associative, as required, and the identity arrow corresponds to the identity element e.

[6] A map $f\colon P_1 \to P_2$ is *order-preserving* or *monotone* if $p \leqslant_{P_1} q \implies f(p) \leqslant_{P_2} f(q)$ for all $p, q \in P_1$, where $\leqslant_{P_1}$ and $\leqslant_{P_2}$ are the binary relations on P_1 and P_2, respectively. For example, if S is a set and $\mathscr{P}(S) = \{A \mid A \subseteq S\}$ is its power set, then the map $f\colon \mathscr{P}(S) \to \mathbb{N}$, where $f(A) = |A|$, is order-preserving, since $A_1 \subseteq A_2 \implies |A_1| \leqslant |A_2|$. (This example is taken from Arbib and Manes [**2**].)

[7] A *semigroup* $(G, *)$ is defined here to be a set G together with a binary operation $*\colon G \times G \to G$ that is associative, i.e. $a * (b * c) = (a * b) * c$ for all $a, b, c \in G$.

[8] A monoid homomorphism is a map $\theta\colon M_1 \to M_2$ such that for every $a, b \in M_1$ we have $\theta(a * b) = \theta(a) \star \theta(b)$, where $*$ and $\star$ are the binary operations in M_1 and M_2, respectively.

3. As an example of the notion of a category used in a more philosophical setting, we note that we can consider any deductive logical system T as a category whose objects are formulae (or theorems) and whose arrows are proofs, provided we have some notion of equivalence of proofs. The categorical operations correspond to the allowed logical deductions. See Marquis [**9**], for more information.

1.4 Categories from categories

Given a category $\mathbf{C}$ or a pair of categories $\mathbf{C}$ and $\mathbf{D}$ one may wonder if there are ways to derive other categories. There are in fact many, and we give some abstract examples, which will become more relevant later.

Definition 1.4. A category $\mathbf{B}$ is a *subcategory* of a category $\mathbf{C}$ if $\operatorname{Ob}\mathbf{B} \subset \operatorname{Ob}\mathbf{C}$, for each pair of objects A and B $\operatorname{Mor}_{\mathbf{B}}(A,B) \subset \operatorname{Mor}_{\mathbf{C}}(A,B)$ and composite and identity arrows in $\mathbf{B}$ and $\mathbf{C}$ coincide.

Definition 1.5. The *dual* (or *opposite*) *category* of a category $\mathbf{C}$, which we write as $\mathbf{C}^{\mathrm{op}}$ is the category whose objects are the objects of $\mathbf{C}$, but whose arrows are the arrows of $\mathbf{C}$ with domain and codomain reversed. That is, an arrow $f\colon A \to B$ of $\mathbf{C}$ appears as an arrow $f^{\mathrm{op}}\colon B \to A$ in $\mathbf{C}^{\mathrm{op}}$. This definition will be very important later.

Definition 1.6. Given a pair of categories $\mathbf{C}$ and $\mathbf{D}$, the *product category* $\mathbf{C} \times \mathbf{D}$ has as objects ordered pairs (C, D) where C is a $\mathbf{C}$-object and D is a $\mathbf{D}$-object, and, as arrows, ordered pairs (f, g), where f is a $\mathbf{C}$-arrow and g is a $\mathbf{D}$-arrow. We define

$$(f,g) \circ (h,i) = (f \circ h, g \circ i) \quad \text{and} \quad \mathrm{id}_{(C,D)} = (\mathrm{id}_C, \mathrm{id}_D).$$

Definition 1.7. For a category $\mathbf{C}$ the *category of arrows over* $\mathbf{C}$, written $\mathbf{C}^{\to}$, has $\mathbf{C}$-arrows as objects. That is, every arrow $f\colon A \to B$ in $\mathbf{C}$ is an *object* in $\mathbf{C}^{\to}$, hence each arrow in $\mathbf{C}^{\to}$ must have $\mathbf{C}$-arrows as its domain and codomain. We define a $\mathbf{C}^{\to}$-arrow from $f\colon A \to B$ to $f'\colon A' \to B'$ to be an ordered pair (a, b) of $\mathbf{C}$-arrows $a\colon A \to A'$ and $b\colon B \to B'$ such that $f' \circ a = b \circ f$, illustrated in the commutative diagram below:

$$\begin{array}{ccc} A & \xrightarrow{a} & A' \\ {\scriptstyle f}\big\downarrow & & \big\downarrow{\scriptstyle f'} \\ B & \xrightarrow[b]{} & B' \end{array}$$

2 Constructions in categories

One of the main aims of category theory is to describe systems of mathematical objects not by their internal structure but by the relationships between them. For example, consider the empty set $\emptyset$ as an object of **Set**. This set has the unique property that for any other set A there is precisely one function from $\emptyset$ to A, namely the empty function; in fact, the empty set is an example of a more general categorical construction called an *initial object*.

There are many possible constructions that we could identify, and we will not list them all, but we introduce some of the more fundamental ones below.

2.1 Terminal and initial objects

Definition 2.1. An object A in a category $\mathbf{C}$ is called a *terminal object* if for every object B in $\mathbf{C}$ there is exactly one arrow from B to A.

Example. In **Set** the terminal objects are the singleton sets $\{x\}$. For any set B the only function from B to $\{x\}$ is the function that sends b to x for every $b \in B$.

Example. A deductive system considered as a category has no terminal objects, as no theorem can be proved in only one way. If a theorem T can be proved assuming T', then it can be proved assuming T' *and* any other theorem U, provided that U does not contradict T.

Definition 2.2. An object A in a category $\mathbf{C}$ is called an *initial object* if for every object B in $\mathbf{C}$ there is exactly one arrow from A to B.

Example. In **Set** the only initial object is the empty set, as we have already seen. For each set B the empty function is the only function from $\emptyset$ to B.

Example. Consider the vector space $\{\mathbf{0}\}$ of the category **Vect** (where we omit the subscript indicating the field). This is a terminal object, since it contains only one element, and so for any vector space V in **Vect** the constant map $\mathbf{v} \mapsto \mathbf{0}$ for every $\mathbf{v} \in V$ is the only arrow from V to $\{\mathbf{0}\}$. But $\{\mathbf{0}\}$ is also an initial object, since every linear map from $\{\mathbf{0}\}$ must fix $\mathbf{0}$ and hence there is only one linear map from $\{\mathbf{0}\}$ to V for any vector space V in **Vect**.

We call any object that is both initial and terminal a *zero object*, because of this example. Many familiar categories contain zero objects, including **Grp**.

(Sometimes arrows to a terminal object or from an initial object are marked with an exclamation mark to indicate their uniqueness, as shown below.)

$$A \xrightarrow{!} B$$

Note that the definition of an initial object is exactly the same as the definition of a terminal object, but with the arrows reversed. This is an example of an important concept in category theory called *duality*.

2.2 Duality

The principle of duality is a very important concept in category theory. Recall definition 1.5 in which we defined the dual category $\mathbf{C}^{\mathrm{op}}$ of a given category $\mathbf{C}$ and observe that every category is the dual of its dual: $\mathbf{C} = (\mathbf{C}^{\mathrm{op}})^{\mathrm{op}}$.

Definition 2.3 (Duality Principle)**.** Let W be any construct defined for some category $\mathbf{C}$. Then the dual of W, usually called co-W, is the construct defined for the category $\mathbf{C}$ by defining W in $\mathbf{C}^{\mathrm{op}}$ and reversing the arrows.

The point of this is that it allows us to form dual statements to any definitions and theorems we want to talk about, simply by reversing the arrows in our diagrams. So the concept of an initial object is actually dual to that of a terminal object; every terminal object in a category $\mathbf{C}$ is an initial object in the dual category $\mathbf{C}^{\mathrm{op}}$. Furthermore, if a theorem holds in $\mathbf{C}$ then its dual holds in $\mathbf{C}^{\mathrm{op}}$. An important corollary of this is that if a theorem holds in every category then so does its dual.

2.3 Monomorphisms, epimorphisms and isomorphisms

In the following definitions we assume A, B and C are objects of a category $\mathbf{C}$.

Definition 2.4. A arrow $f\colon B \to C$ is a *monomorphism* if for any pair of arrows $g, h\colon A \to B$ we have

$$f \circ g = f \circ h \implies g = h$$

In other words, f is a monomorphism (or is *monic*) if $g = h$ whenever the following diagram commutes:

$$A \underset{h}{\overset{g}{\rightrightarrows}} B \xrightarrow{f} C$$

(We adopt the widely accepted convention that a diagram commutes if any two paths between two objects given the same result, *provided at least one path consists of more than one arrow.* So the diagram above on its own means that $f \circ g = f \circ h$, but **not** that $g = h$.)

Definition 2.5. An arrow $f\colon A \to B$ is an *epimorphism* if for any pair of arrows $g, h\colon B \to C$ we have

$$g \circ f = h \circ f \implies g = h$$

In other words, f is an epimorphism (or is *epic*) if $g = h$ whenever the following diagram commutes:

$$A \xrightarrow{f} B \underset{h}{\overset{g}{\rightrightarrows}} C$$

In the same way that initial objects are the categorical dual to terminal objects, epimorphisms are dual to monomorphisms. An epimorphism is a co-monomorphism; in other words, an epimorphism in $\mathbf{C}$ is a monomorphism in $\mathbf{C}^{\mathrm{op}}$ and since $(\mathbf{C}^{\mathrm{op}})^{\mathrm{op}} = \mathbf{C}$ a monomorphism is also a co-epimorphism.

Definition 2.6. An arrow $f\colon A \to B$ is an *isomorphism* if there exists an arrow $k\colon B \to A$ such that

$$k \circ f = \mathrm{id}_A \quad \text{and} \quad f \circ k = \mathrm{id}_B$$

If there is an isomorphism between two objects A and B we write $A \cong B$.

Example. As hinted at in the introduction, isomorphisms in **Set** are bijections (which will be proved in section 2.5), in **Grp** they are group isomorphisms and in **Top** they are homeomorphisms.

Often two or more categorical constructions (in the widest sense) can be considered to be essentially the same. We say two objects are "unique up to isomorphism" if there is unique isomorphism between them. For example

Lemma 2.7. *Terminal objects are unique up to isomorphism. That is, for any two terminal objects there is a unique isomorphism between them.*

Proof. Let T and T' be two terminal objects in a category $\mathbf{C}$. Since T is terminal there is a unique arrow $u\colon T' \to T$ and similarly there is a unique $u'\colon T \to T'$, since T' is terminal. But this uniqueness implies that $u' \circ u = \mathrm{id}_T$ and $u \circ u' = \mathrm{id}_{T'}$, so u is a unique isomorphism with inverse u' (and vice versa). □

Applying the duality principle to lemma 2.7 gives us the following for free, since an initial object in $\mathbf{C}$ is just a terminal object in $\mathbf{C}^{\mathrm{op}}$.

Lemma 2.8. *Initial objects are unique up to isomorphism.*

Note that an isomorphism f (with inverse k) is always a monomorphism and an epimorphism:

$$f \circ g = f \circ h \implies k \circ f \circ g = k \circ f \circ h \implies g = h$$
and
$$g \circ f = h \circ f \implies g \circ f \circ k = h \circ f \circ k \implies g = h.$$

The dual of an isomorphism is both the dual of a monomorphism and the dual of an epimorphism; that is, a co-isomorphism is both an epimorphism and a monomorphism, i.e. an isomorphism. We say that isomorphism is a self-dual concept [**2**].

One might expect the converse to be true that an arrow that is both a monomorphism and an epimorphism is necessarily an isomorphism, but this is not generally the case, as illustrated by the following example:

Example. Consider the category **2** mentioned in section 1.2:

$$\bullet \xrightarrow{f} \bullet$$

The single non-identity arrow f is a monomorphism, since there is only one arrow, the identity of its domain, that can be composed with it on the left. Similarly f is an epimorphism, since only the identity arrow of its codomain can compose with it on the right. But f is not an isomorphism, for there is no arrow k such that $f \circ k$ and $k \circ f$ are the identity arrows of f's domain and codomain, respectively.

In the next section we will take a quick detour into so-called "generalised elements" in order to describe necessary and sufficient conditions for an arrow to be an isomorphism. In some categories all epic monomorphisms are indeed isomorphisms; such categories are called *balanced*. In fact, many familiar categories are balanced, including **Set** (see section 2.5), but notably the "category of categories", which we will discuss later, is not balanced.

2.4 Generalised elements

(This section is based heavily on [**10**], Ch. 1. The notation may be non-standard.)

Definition 2.9. (McLarty [**10**], p. 17) We call an arrow $x\colon T \to A$ a *generalised element* of A, at stage of definition T, and write $x \in_T A$. Furthermore, if there is an arrow $f\colon A \to B$ we can express the fact that $f \circ x\colon T \to B$ by writing $f(x) \in_T B$.

If A is terminal then there is a one-to-one correspondence between the elements of an object B and the arrows $x\colon A \to B$. In this case $x \in_A B$ is called a *global element* of B. While this is not immediately obvious in generality, it can easily be seen to be true for the category **Set**.

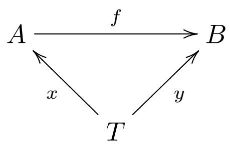
Generalised elements allow us to formulate the concept of a monomorphism in a familiar way:

Lemma 2.10. *Let $f\colon A \to B$ and suppose $x, y \in_T A$. Then f is a monomorphism if and only if $f(x) = f(y) \implies x = y$. That is, monomorphisms are injective on generalised elements and the converse.*

Proof. $x, y \in_T A$ means precisely that there are arrows $x, y\colon T \to A$. As f is monic, $f \circ x = f \circ y \implies x = y$. The converse is similar. □

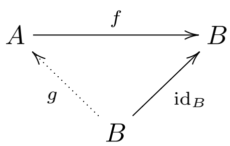
Unfortunately, there is no similar description of epimorphisms in terms of their actions of generalised elements ([**10**], p. 17), so we proceed with a new definition.

Definition 2.11. An arrow $f\colon A \to B$ is *surjective on generalised elements* if for each stage of definition T and each $y \in_T B$ there is some $x \in_T A$ with $f(x) = y$, i.e. such that the following diagram commutes.

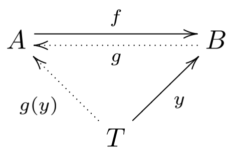


Lemma 2.12. *An arrow $f\colon A \to B$ is surjective on generalised elements if and only if there is some $g\colon B \to A$ with $f \circ g = \mathrm{id}_B$.*

Proof. ($\Longrightarrow$) Suppose that f is surjective on generalised elements. Then considering the element $\mathrm{id}_B \in_B B$, there must exist some $g \in_B A$ with $f(g) = \mathrm{id}_B$.



($\Longleftarrow$) Conversely, suppose that such a g exists. Then for any $y \in_T B$ the element $g(y) \in_T A$ has $f(g(y)) = y$, i.e. $f \circ g = \mathrm{id}_B$.



□

When $f \circ g = \mathrm{id}_B$ as above, we say that g is a *right inverse* for f and that f is a *left inverse* for g. Right inverses are not unique. Every arrow f with a right inverse g is an epimorphism since

$$h \circ f = k \circ f \implies h \circ f \circ g = k \circ f \circ g \implies h = k$$

Theorem 2.13. *An arrow is an isomorphism if and only if it is both injective and surjective on generalised elements; that is, if and only if it is both a monomorphism and a split epimorphism.*

Proof. ($\Longrightarrow$) Since every isomorphism f has a both-sided inverse f^{-1}, all isomorphisms are split epic. Every isomorphism is also monic since

$$f \circ h = f \circ k \implies f^{-1} \circ f \circ h = f^{-1} \circ f \circ k \implies h = k.$$

($\Longleftarrow$) Suppose f is monic and split epic and consider a right inverse g of f. Then

$$f \circ \mathrm{id}_A = f = (f \circ g) \circ f = f \circ (g \circ f) \implies g \circ f = \mathrm{id}_A,$$

as f is monic. So f is an isomorphism with inverse g. □

2.5 Monomorphisms and epimorphisms in Set

We follow the exposition of Arbib and Manes [**2**].

Proposition 2.14. *In* **Set***, an arrow is injective if and only if it is a monomorphism*

Proof. ($\Longrightarrow$) Suppose $f\colon B \to C$ is injective and for arrows $g, h\colon A \to B$ we have $f \circ g(a) = f \circ h(a)$ for every $a \in A$. By the definition of injectivity this implies that $g(a) = h(a)$ for every $a \in A$, i.e. that $g = h$.

($\Longleftarrow$) Suppose f is not injective; then there exist distinct elements $b_1, b_2 \in B$ for which $f(b_1) = f(b_2)$. Define $g, h\colon B \to B$ by

$$g(b) = b \qquad h(b) = \begin{cases} b & \text{if } b \neq b_1 \\ b_2 & \text{if } b = b_1 \end{cases}$$

Then clearly $f \circ g(b) = f(b)$ for every $b \in B$. Also, $f \circ h(b) = f(b)$ for $b \neq b_1$ and $f \circ h(b_1) = f(b_2) = f(b_1)$, so $f \circ g(b) = f \circ h(b)$ for every $b \in B$, i.e. $f \circ g = f \circ h$ even though $g \neq h$. Hence f is not a monomorphism. ☐

Proposition 2.15. *In* **Set***, an arrow is surjective if and only if it is an epimorphism*

Proof. ($\Longrightarrow$) Suppose $f\colon A \to B$ is surjective and for arrows $g, h\colon B \to C$ we have $g \circ f(a) = h \circ f(a)$ for every $a \in A$. Since f is surjective, every $b \in B$ can be written as $f(a)$ for some $a \in A$ and this implies that $g(b) = h(b)$ for every $b \in B$. In other words, f is an epimorphism.

($\Longleftarrow$) Again we use a contrapositive argument. Assume f is not surjective; then there exists at least one element $b_1 \in B$ which is not equal to $f(a)$ for any $a \in A$. Pick any other element $b_2 \in B$ an define $g, h\colon B \to B$ by

$$g(b) = b \qquad h(b) = \begin{cases} b & \text{if } b \neq b_1 \\ b_2 & \text{if } b = b_1 \end{cases}$$

Then $g \circ f(a) = f(a)$ for every $a \in A$ and also, since no element of A is mapped to b_1 we have $h \circ f(a) = f(a)$ for every $a \in A$, so $g \circ f = h \circ f$. But it is clear $g \neq h$, since $g(b_1) \neq h(b_1)$. Hence f is not an epimorphism. ☐

Corollary 2.16. *In* **Set** *an arrow is a bijection if and only if it is an isomorphism.*

Proof. The result is immediate upon realising that an arrow is a bijection if and only if it is both injective and surjective and that all isomorphisms are both monic and epic. ☐

3 The category of categories

Having seen many examples of categories in section 1.2 consisting of a certain type of mathematical domain as objects with relevant maps (or classes of maps) between these domains as arrows, it is natural to ask if we can form a "category of categories", whose objects are categories and whose arrows are the appropriate structure-preserving map between categories.

Foundational issues stop us from forming a category of *all* categories (see e.g. Mac Lane [**8**] or Cameron [**5**]), but it is possible to form a category **Cat** of all *small* categories (recall subsection 1.2). The arrows in **Cat** are called *functors*, and are defined below.

Definition 3.1. Given two categories $\mathbf{C}$ and $\mathbf{D}$, a (covariant[9]) *functor* $\mathscr{F}: \mathbf{C} \to \mathbf{D}$ is a map takes every $\mathbf{C}$-object A to a $\mathbf{D}$-object $\mathscr{F}(A)$ and every $\mathbf{C}$-arrow $f: A \to B$ to a $\mathbf{D}$-arrow $\mathscr{F}(f): \mathscr{F}(A) \to \mathscr{F}(B)$, in such a way that compositions and identities are preserved. That is, $\mathscr{F}(f \circ g) = \mathscr{F}(f) \circ \mathscr{F}(g)$ and $\mathscr{F}(\mathrm{id}_A) = \mathrm{id}_{\mathscr{F}(A)}$.

3.1 Examples of functors

Example. For each category $\mathbf{C}$ there is an identity functor $\mathbf{id_C}: \mathbf{C} \to \mathbf{C}$ that maps each object to itself each arrow to itself.

Example. A relatively straightforward example of a functor in **Set** is the power set operator $\mathscr{P}: \mathbf{Set} \to \mathbf{Set}$, taking each set X to its power set, the set of all its subsets. $\mathscr{P}(X)$ and each arrow $f: X \to Y$ to the arrow $\mathscr{P}(f): \mathscr{P}(X) \to \mathscr{P}(Y)$ which sends a subset $A \subseteq X$ to $B = f(A)$, the image of f restricted to A.

Example. Consider a monoid $(M, *, e)$ in the category **Mon**. If, for a moment, we ignore the fact that this monoid includes an associative binary operation and an identity element then we can just associate $(M, *, e)$ with the underlying set M; similarly, we can ignore the fact that the arrows in **Mon** are monoid homomorphisms and consider them simply as set functions between the underlying sets of our monoids.

The functor $\mathscr{U}: \mathbf{Mon} \to \mathbf{Set}$ which takes each monoid $(M, *, e)$ to its underlying set M and each monoid homomorphism to the corresponding set function is called the *forgetful functor* from **Mon** to **Set** as it simply "forgets" the structure of the objects and arrows in its domain. Similarly there are forgetful functors from **Grp** to **Set**, from **Top** to **Set**, from **AbGrp** to **Grp**, from **Met** to **Haus**, and so on.

Example. If $\mathbf{D}$ is a subcategory of $\mathbf{C}$ then there is an *inclusion functor* $\mathbf{i}: \mathbf{D} \to \mathbf{C}$ which takes every $\mathbf{D}$-object to itself considered as an object of $\mathbf{C}$, and every $\mathbf{D}$-arrow $f: A \to B$ to the same arrow considered as a $\mathbf{C}$-arrow. For example, there are inclusion functors from **Mon** to **Grp**, from **Top** to **Met**, and from **Vect** to **NLinSp**, the category of normed linear spaces and bounded linear transformations.

Example. The functor $\Delta: \mathbf{C} \to \mathbf{C} \times \mathbf{C}$ (where $\mathbf{C} \times \mathbf{C}$ is the product category of definition 1.6) which takes each $\mathbf{C}$-object C to the $\mathbf{C} \times \mathbf{C}$-object (C, C) and each $\mathbf{C}$-arrow $f: A \to B$ to the arrow $(f, f): (A, A) \to (B, B)$ is called the *diagonal functor*.

The above examples illustrate just a few possible functors. In general there are many different functors between any two given categories.

3.2 Natural transformations

There is a natural relationship between the identity functor $\mathbf{id_{Set}}: \mathbf{Set} \to \mathbf{Set}$ and the power set functor $\mathscr{P}: \mathbf{Set} \to \mathbf{Set}$. For each set X consider the set function $\eta_X: X \to \mathscr{P}(X)$ sending each $x \in X$ to the singleton set $\{x\}$, which is a subset of X, i.e. an element of $\mathscr{P}(X)$. For any sets X and Y and any function $f: X \to Y$ we get the same result if we consider a set X, map each element to the corresponding singleton set and then apply a restriction of f to each of these subsets of X as if we first apply f to the set X and then map each each of the elements in

[9] A *contravariant functor* $\mathscr{F}: \mathbf{C} \to \mathbf{D}$ that takes every $\mathbf{C}$-object A to a $\mathbf{D}$-object $\mathscr{F}(A)$ and every $\mathbf{C}$-arrow $f: A \to B$ to a $\mathbf{D}$-arrow $\mathscr{F}(f): \mathscr{F}(B) \to \mathscr{F}(A)$. That is, it swaps the domain and codomain of the arrows it maps. In other words, a contravariant functor $\mathscr{F}: \mathbf{C} \to \mathbf{D}$ is just a covariant functor $\mathscr{F}: \mathbf{C}^{\mathrm{op}} \to \mathbf{D}$. We will not be concerned with the distinction between covariant and contravariant functors in this essay.

the image $f(X)$ to their corresponding singletons; in other words, $\mathscr{P}(f) \circ \eta_X = \eta_Y \circ \mathbf{id_{Set}}(f)$, illustrated in the following commutative diagram:

$$\begin{array}{ccc} \mathbf{id_{Set}}(X) & \xrightarrow{\mathbf{id_{Set}}(f)} & \mathbf{id_{Set}}(Y) \\ {\scriptstyle \eta_X}\downarrow & & \downarrow{\scriptstyle \eta_Y} \\ \mathscr{P}(X) & \xrightarrow[\mathscr{P}(f)]{} & \mathscr{P}(Y) \end{array}$$

This can be simplified by remembering that the image of any set X or any set function $f\colon X \to Y$ under the identity functor $\mathbf{id_{Set}}$ is simply itself.

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle \eta_X}\downarrow & & \downarrow{\scriptstyle \eta_Y} \\ \mathscr{P}(X) & \xrightarrow[\mathscr{P}(f)]{} & \mathscr{P}(Y) \end{array}$$

The above diagrams commute for any pair of sets X and Y, and we say that the collection of maps $\eta_{X_i}\colon X \to \mathscr{P}(X)$, indexed over the objects X_i of **Set**, is a *natural transformation.* We write $\eta\colon \mathbf{id_{Set}} \stackrel{.}{\to} \mathscr{P}$.

Now we give the general definition.

Definition 3.2. A *natural transformation* between two functors $\mathscr{F}, \mathscr{G}\colon \mathbf{C} \to \mathbf{D}$ is a family of maps, indexed over the collection of objects of $\mathbf{C}$, such that

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

We write $\eta\colon \mathscr{F} \stackrel{.}{\to} \mathscr{G}$ to show that η is a natural transformation from between the functors $\mathscr{F}$ and $\mathscr{G}$. If η is an isomorphism we call it a *natural isomorphism.*

We have just defined what seems to be a kind of morphism between functors. As is customary in category theory, we abstract out one level further and consider natural transformations as the arrows in a category of functors, which we now define. If it seems strange that an arrow between two functors is actually more than one arrow, recall definition 1.7 where the arrow category $\mathbf{C}^{\to}$ of a category $\mathbf{C}$ was defined, where $\mathbf{C}^{\to}$-arrows are actually collections of $\mathbf{C}$-arrows.

Definition 3.3. Given two categories $\mathbf{C}$ and $\mathbf{D}$ the *functor category* $\mathbf{D}^{\mathbf{C}}$ is the category whose objects are functors from $\mathbf{C}$ to $\mathbf{D}$ and whose arrows are natural transformations between such functors.

Functors and natural transformations are extremely important in category theory as they allow us to the describe the relationships between different mathematical domains and then the relationships between these relationships. In fact category theory arose as a means to formalise such relationships in algebraic topology, in which topological spaces are associated with groups and other algebraic structures. Theories such as homology, cohomology, homotopy and K-theory are all examples of functors between categories of topological spaces and categories of various algebraic structures [**9**]. Eilenberg and Mac Lane [**6**] published the original paper on category theory in order to describe the natural transformations between such functors; the notions of a functor and of a category were effectively auxiliary ones.

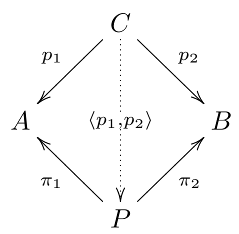
An advanced concept that is central to category theory is that of an *adjoint functor*, introduced by Daniel Kan over ten years after the initial publication of the original paper by Eilenberg and Mac Lane [**6**]. In some sense, an adjoint is like a conceptual inverse to a functor. For example (taken from [**9**]), the functor from **Set** to **Grp** that takes a set to the so-called "free group" generated from it (very loosely, the group created from a set with no constraints other than the fact that the result must be a group) is an adjoint functor to the forgetful functor $\mathscr{U}\colon \mathbf{Grp} \to \mathbf{Set}$. The amazing thing about adjoint functors is that they appear in so many different ways. In particular, a remarkable number of constructions in mathematics arise as adjoints to certain forgetful functors, which despite appearing on the surface to be very straightforward, turn out to be particularly useful [**9**]. The apparently fundamental notions of limits and colimits in section 4 are also actually specific instances of adjoint functors.

4 Universal constructions

The constructions we have seen so far — terminal and initial objects, monomorphisms and epimorphisms — share a similar property. This section introduces three more important categorical constructions (products, equalisers and pullbacks) as well as their duals constructions (coproducts, coequalisers and pushouts). These constructions all consist of an object X together with arrows — maybe only one — from (or to, in the case of the dual constructions) X satisfying certain conditions, and with the property that if any other object X' and arrows satisfy these conditions then there is a unique arrow from X' to X (from X to X'). Such a construction is said to be *universal*.

4.1 Products and coproducts

Definition 4.1. A *product* of two objects A and B consists of an object P together with arrows $\pi_1\colon P \to A$ and $\pi_2\colon P \to B$ (called *projections*) such that for any other object C and arrows $p_1\colon C \to A$ and $p_2\colon C \to B$ there is a unique arrow $\langle p_1, p_2\rangle\colon C \to P$, as shown in the following diagram:

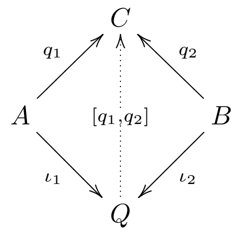


There may be many different product objects for a given pair A and B, or none at all. If we pick a particular representative object as our product, it is often denoted $A \times B$ by analogy with the Cartesian product in **Set**, which is the manifestation of the categorical product in that category (see following example). It is important to remember that this choice is not unique: for instance, in **Set**, the set $B \times A$ with the projections $\pi_2\colon B \times A \to A$ and $\pi_1\colon B \times A \to B$ is just as good as product for A and B. However, in proposition 4.3 we will show that any two product objects satisfying definition 4.1 are isomorphic, i.e. they are essentially the same in categorical terms. Note also that we do not really need to state what the product object P

is; it is determined by the arrows π_1 and π_2. So we may think of a product as an ordered pair $(\pi_1 \colon P \to A,\ \pi_2 \colon P \to B)$.

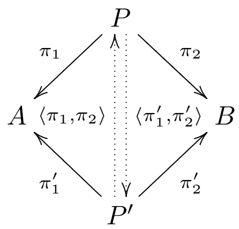
Reversing the arrows we obtain the definition of a coproduct, the dual of the categorical product:

Definition 4.2. A *coproduct* of two objects A and B is a pair of arrows $(\iota_1 \colon A \to Q,\ \iota_2 \colon B \to Q)$, called *injections*, such that for any other object C and arrows $q_1 \colon A \to C$ and $q_2 \colon B \to C$ there is a unique arrow $[q_1, q_2] \colon Q \to C$, as shown in the following diagram:



Proposition 4.3. *Product objects are isomorphic.*

Proof. Suppose $(\pi_1 \colon P \to A, \pi_2 \colon P \to B)$ and $(\pi_1' \colon P' \to A, \pi_2' \colon P' \to B)$ are products for A and B. Then, since P is a product object there is a unique arrow $\langle \pi_1, \pi_2 \rangle \colon P' \to P$ such that $\pi_1 \circ \langle \pi_1, \pi_2 \rangle = \pi_1'$ and $\pi_2 \circ \langle \pi_1, \pi_2 \rangle = \pi_2'$. Similarly, there is a unique arrow $\langle \pi_1', \pi_2' \rangle \colon P \to P'$ such that $\pi_1' \circ \langle \pi_1', \pi_2' \rangle = \pi_1$ and $\pi_2' \circ \langle \pi_1', \pi_2' \rangle = \pi_2$.



Since these arrows are unique, we must have $\langle \pi_1, \pi_2 \rangle \circ \langle \pi_1', \pi_2' \rangle = \mathrm{id}_P$ and $\langle \pi_1', \pi_2' \rangle \circ \langle \pi_1, \pi_2 \rangle = \mathrm{id}_{P'}$, so $P \cong P'$. $\square$

Example. Consider two sets A and B in the category **Set**. The Cartesian product $A \times B$ is also their categorical product with $\pi_1 \colon A \times B \to A$ and $\pi_2 \colon A \times B \to B$ the usual projections:

$$\pi_1 \colon (a, b) \mapsto a \qquad \pi_2 \colon (a, b) \mapsto b$$

Given any other set C and functions $p_1 \colon C \to A$ and $p_2 \colon C \to B$, there is a unique function $\langle h, k \rangle \colon C \to A \times B$ defined by

$$\langle h, k \rangle \colon c \mapsto (h(c), k(c))$$

whose first an second components are the values of h and k applied to c, respectively[10]. Then $\pi_i \circ \langle p_1, p_2 \rangle = p_i$ for $i = 1, 2$ as required.

Example. The coproduct of two sets A and B corresponds to their disjoint union

$$A \uplus B = (A \times \{0\}) \cup (B \times \{1\})$$

[10]The notation $\langle h, k \rangle$ is likely to have arisen from the convention, which I have chosen not to use, of writing ordered pairs in angle brackets, so that $\langle h, k \rangle \colon c \mapsto \langle h(c), k(c) \rangle$.)

with the injections $\iota_1 \colon A \to A \uplus B$ and $\iota_2 \colon B \to A \uplus B$ defined by

$$\iota_1 \colon a \mapsto (a, 0) \quad \text{and} \quad \iota_2 \colon b \mapsto (b, 1)$$

Given any other set C and functions $q_1 \colon A \to C$, $q_2 \colon B \to C$, the unique function $[q_1, q_2] \colon A \uplus B \to C$ that makes the diagram commute is defined

$$[q_1, q_2] \colon (a, 0) \mapsto q_1(a) \qquad [q_1, q_2] \colon (b, 1) \mapsto q_2(b)$$

Then $[q_1, q_2] \circ \iota_i = q_i$ for $i = 1, 2$, as required.

Example. Consider two groups $(G, \cdot)$ and $(H, *)$ in the category **Grp**. Define the *direct product* of $(G, \cdot)$ and $(H, *)$ as $(G \otimes H, \star)$, where

$$G \otimes H = \{(g, h) \mid g \in G, h \in H\}$$

and

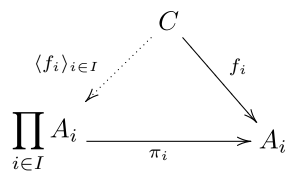
$$(g, h) \star (g', h') = (g \cdot g', h * h')$$

It is easy to check that the group axioms are satisfied[11], making $(G \otimes H, \star)$ an object in **Grp**. It is also clear that the projection arrows of the previous example are group homomorphisms[12] so that $(G \otimes H, \star)$ is a categorical product of $(G, \cdot)$ and $(H, *)$.

Example. In the category of categories **Cat** a product of two categories **C** and **D** is the product category $\mathbf{C} \times \mathbf{D}$. Their coproduct is what is known as a sum category, which we will not define here.

The definitions of categorical products and coproducts can be extended to any collection of objects in an obvious way. The definition of a general product is given below; that of a general coproduct can be discovered simply by reversing the arrows.

Definition 4.4. A *product* of a collection $(A_i)_{i \in I}$ of objects indexed by a set I consists of an object $\prod_{i \in I} A_i$ and a collection of projection arrows $(\pi_i \colon (\prod_{i \in I} A_i) \to A_j)_{j \in I}$, such that for each object C and collection of arrows $(f_i \colon C \to A_i)_{i \in I}$ there is a unique arrow $\langle f_i \rangle_{i \in I} \colon C \to (\prod_{i \in I} A_i)$ such that the the following diagram commutes for all $i \in I$:



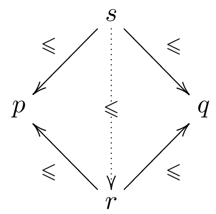
It is clear to see that definition 4.1 is just the specific case of this definition when $I = \{1, 2\}$.

Example. Consider a poset $(P, \leqslant)$ as a category. Recall that we draw an arrow from p to q if $p \leqslant q$. The product of two objects p and q is an object r such that $r \leqslant p$, $r \leqslant q$ and, whenever

[11] Associativity of $\star$ follows from the associativity of $\cdot$ and $*$; if e_G and e_H are the identities of $(G, \cdot)$ and $(H, *)$ take $e_{G \otimes H} = (e_G, e_H)$ as the identity of $(G \otimes H, \star)$; the inverse of $(g, h) \in (G \otimes H, \star)$ is (g^{-1}, h^{-1}), where g^{-1} and h^{-1} are the inverses of g and h in $(G, \cdot)$ and $(H, *)$, respectively.

[12] $\pi_1((g, h) \star (g', h')) = \pi_1(g \cdot g', h * h') = g \cdot g' = \pi_1(g, h) \cdot \pi_1(g', h')$ and similarly for π_2.

we have s with $s \leqslant p$ and $s \leqslant q$, we have $s \leqslant r$:

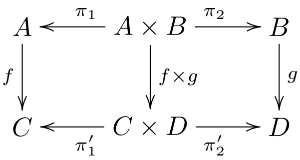


In other words, r is a lower bound for p and q, and furthermore it is the largest possible lower bound: all other lower bounds are smaller (or equal). So in a poset considered as a category the product of two objects is their minimum. The categorical product of an arbitrary collection of objects in a poset considered as a category is their greatest lower bound.

Similarly, the coproduct of two objects in a poset considered as a category is their maximum, an the coproduct of an arbitrary collection is their least upper bound.

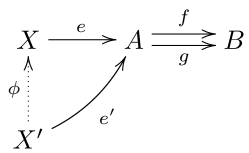
We can also define product arrows between product objects:

Definition 4.5. If $A \times B$ and $C \times D$ are two product objects then for each pair of arrows $f \colon A \to C$ and $g \colon B \to D$ the *product arrow* $f \times g \colon A \times B \to C \times D$ is the arrow $\langle f \circ \pi_1, g \circ \pi_2 \rangle$.



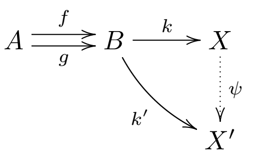
4.2 Equalisers and coequalisers

Definition 4.6. An arrow $e \colon X \to A$ is an *equaliser* of two arrows $f, g \colon A \to B$, written $e = \mathrm{eq}(f, g)$, if $f \circ e = g \circ e$ and whenever $f \circ e' = g \circ e'$ for some other object X' and arrow $e' \colon X' \to A$ there exists a unique arrow $\phi \colon X' \to X$ such that $e' = e \circ \phi$, i.e. such that the following diagram commutes:



The definition of the dual notion of an equaliser, known as a *coequaliser*, is discovered by reversing the direction of the arrows, since a coequaliser in a category $\mathbf{C}$ is just an equaliser in $\mathbf{C}^{\mathrm{op}}$:

Definition 4.7. An arrow $k \colon B \to X$ is a *coequaliser* of two arrows $f, g \colon A \to B$, written $k = \mathrm{coeq}(f, g)$, if $k \circ f = k \circ g$ and whenever $k' \circ f = k' \circ g$ for some other object X' and arrow $k' \colon B \to X'$ there exists a unique arrow $\psi \colon X \to X'$ such that $k' = \psi \circ k$, i.e. such that the following diagram commutes:



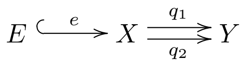
The following examples show how equalisers and coequalisers manifest themselves in **Set**.

Example. (Adapted from Arbib and Manes [**2**], pp. 21–22.) Equalisers in **Set** give us an idea of where the term "equaliser" comes from. Suppose that X is a set and we have two functions $q_1, q_2 \colon X \to Y$.

Proposition 4.8. *The equaliser $e\colon E \to X$ of q_1 and q_2 is the inclusion map $e\colon E \to X$, sending each $x \in E$ to itself considered as an element of X, where*

$$E = \{x \in X \mid q_1(x) = q_2(x)\}$$

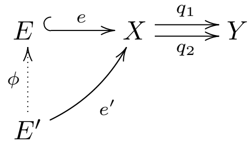
Proof. By definition of E we have that $q_1 \circ e(x) = q_2 \circ e(x)$ for every $x \in E$, hence



Suppose there is some other arrow $e'\colon E' \to X$ such that $q_1 \circ e' = q_2 \circ e'$. Define a map

$$\phi\colon E' \to E \qquad \phi\colon x \mapsto e'(x)$$

which works because our supposition tells that every element $x \in E'$ satisfies $q_1 \circ e'(x) = q_2 \circ e'(x)$, and so $e'(x) \in E$.



Any other map in place of ϕ would differ on some elements $x \in E'$, but $e(x) = x$ for all $x \in E$ so ϕ must be unique. Hence $e = \mathrm{eq}(q_1, q_2)$. □

Example. (Adapted from Arbib and Manes [**2**], pp. 16–20.) In **Set** the concept of a coequaliser is related to that of an equivalence relation. Before proposing exactly how they are related we briefly run through some terminology.

Recall that an equivalence relation is a binary relation that is reflexive, symmetric and transitive. That is, for all x, y, z, we have

$$x \sim x, \qquad x \sim y \Longleftrightarrow y \sim x \quad \text{and} \quad x \sim y,\ y \sim z \implies x \sim z$$

If $\sim$ is an equivalence relation on a set X we define the set $E \subseteq X \times X$, which we will also refer to as the equivalence relation, by

$$E = \{(x, y) \in X \times X \mid x \sim y\}$$

and for each $x \in X$ we call the set

$$[x]_E = \{y \in X \mid (x, y) \in E\}$$

the *equivalence class* of x (with respect to X). The set of distinct equivalence classes is called the *quotient set* of X (with respect to E) and is written X/E.

We associate with each element $x \in X$ the equivalence class containing x by the function

$$\eta_E \colon X \to X/E \qquad \eta_E \colon x \mapsto [x]_E$$

which we call the *canonical onto map*. Since the equivalence classes partition X/E, this function is clearly surjective (hence the name).

Notice that given any relation R on a set X (i.e. any subset of $X \times X$, not necessarily an equivalence relation), we can define projections

$$p_i \colon R \to X \qquad p_i \colon (x_1, x_2) \mapsto x_i$$

for $i = 1, 2$. Conversely, any pair of maps $p_1, p_2 \colon A \to X$ give rise to a relation, defined as

$$R_A = \{(p_1(a), p_2(a)) \mid a \in A\}$$

The maps p_1 an p_2 define which elements are equivalent. Two elements $x, y \in X$ are equivalent if and only if there is some $a \in A$ with $p_1(a) = x$ and $p_2(a) = y$. Note that if p_1 and p_2 are the usual projections then the relation they define is the whole of their domain.

It is reasonable to expect that any such relation can be extended to an equivalence relation on X by including more pairs, and this is indeed the case[13]. We denote the smallest equivalence relation containing R_A by $\overline{R}$. Then

$$\eta_{\overline{R}} \circ p_1(a) = \eta_{\overline{R}} \circ p_2(a)$$

for every $a \in A$, since $(p_1(a), p_2(a)) \in R_A$ and hence $p_1(a)$ and $p_2(a)$ are in the same equivalence class with respect to $\overline{R}$. Hence

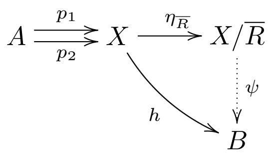
$$A \underset{p_2}{\overset{p_1}{\rightrightarrows}} X \xrightarrow{\eta_{\overline{R}}} X/\overline{R}$$

Proposition 4.9. *Given maps* $p_1, p_2 \colon A \to X$, *then* $\eta_{\overline{R}} = \mathrm{coeq}(p_1, p_2)$.

Proof. Suppose there is some other arrow $h \colon X \to B$ such that $h \circ p_1 = h \circ p_2$. Writing $[x]$ for $[x]_{\overline{R}}$, let us define a map

$$\psi \colon X/\overline{R} \to B \qquad \psi \colon [x] \mapsto h(x)$$

which is well-defined, since if $[x] = [y]$ then $x \sim y$ (with respect to $\overline{R}$) and therefore $h(x) = h \circ p_1(a) = h \circ p_2(a) = h(y)$ by our assumption on h.



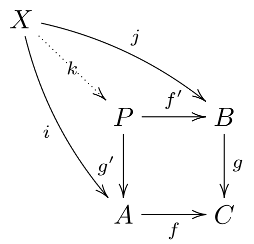
Since $\eta_{\overline{R}}$ is surjective ψ must be unique, for if it were not then any distinct arrow would have to differ from ψ on some elements of $X/\overline{R}$, but this is not possible. Hence $\eta_{\overline{R}}$ is a coequaliser of p_1 and p_2. □

4.3 Pullbacks and pushouts

Definition 4.10. A *pullback* of a pair of arrows $f \colon A \to C$ and $g \colon B \to C$ is an object P together with pair of arrows $f' \colon P \to B$ and $g' \colon P \to A$ such that $f \circ g' = g \circ f'$ and for any

[13] For proof, see Arbib and Manes [**2**], p. 19.

other object X and arrows $i\colon X \to A$ and $j\colon X \to B$ with $f \circ i = g \circ j$ there is a unique arrow $k\colon X \to P$ such that the following diagram commutes:



The definition of the dual notion, a *pushout*, can be found by reversing the arrows, since a co-pullback, i.e. a pushout, in a category **C** is simply a pullback in $\mathbf{C}^{\mathrm{op}}$.

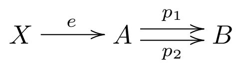
A pullback can be thought of as a sort of constrained product, where the additional constraint is the fact that we must have $f \circ g' = g \circ f'$ [**13**]. Pullbacks are actually extremely useful constructions in many fields including algebraic geometry, but we will not go into details here.

4.4 Theorems on universal constructions

In this section we give a small number of examples of theorems involving some of the constructions introduced in the previous sections to give the reader an idea of how they interrelate with one another.

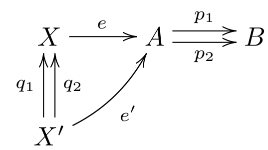
Theorem 4.11. *Every equaliser is a monomorphism.*

Proof. Let $p_1, p_2\colon A \to B$ and let $e\colon X \to A$ be an equaliser of p_1 and p_2, so that $p_1 \circ e = p_2 \circ e$:



Now suppose $q_1, q_2\colon X' \to X$ are such that $e \circ q_1 = e \circ q_2$. We must prove that we necessarily have $q_1 = q_2$, i.e. that e is a monomorphism.

Let $e' = e \circ q_1$ $(= e \circ q_2)$. Then the following diagram commutes, since $p_1 \circ e' = (p_1 \circ e) \circ q_1 = (p_2 \circ e) \circ q_1 = p_2 \circ e'$:



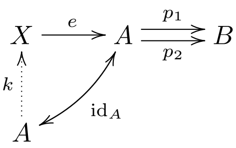
Since e is an equaliser, there is a unique arrow $k\colon X' \to X$ such that $e' = e \circ k$. But $e' = e \circ q_1 = e \circ q_2$ by assumption, so we must have $q_1 = k = q_2$. Hence e is a monomorphism. ☐

By applying the duality principle we immediately see that every coequaliser is an epimorphism. Furthermore we have the following proposition.

Theorem 4.12. *Every epic equaliser is an isomorphism.*

Proof. We draw the same diagram as in the previous proposition. Since e is epic, we have $p_1 \circ e = p_2 \circ e \implies p_1 = p_2$. Then as $p_1 \circ \mathrm{id}_A = p_2 \circ \mathrm{id}_A$, we may replace X' by A and e' by id_A

to get



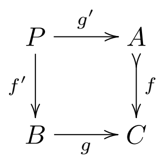
As e is an equaliser of p_1 and p_2 there is a unique arrow $k\colon A \to X$ such that $\mathrm{id}_A = e \circ k$. Looking at the arrow $\mathrm{id}_A\colon A \to A$ in the opposite direction we see that $\mathrm{id}_X = k \circ (\mathrm{id}_A \circ e) = k \circ e$. Hence e is an isomorphism, with inverse k. □

The dual of this statement tells us that every monic coequaliser is also an isomorphism.

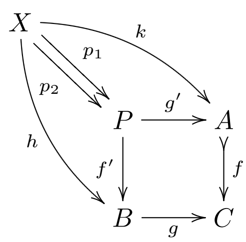
The two theorems above show that the notion of an equaliser is stronger than that of a monomorphisms (and similarly for the duals). As we saw in section 2.3, for an arrow to be an isomorphism, there are cases where being a monomorphism and an epimorphism is not enough; however, being an equaliser and an epimorphism is always sufficient.

Theorem 4.13. *The pullback of a monomorphism is also a monomorphism.*

Proof. Suppose there are arrows f, g, f' and g' such that $f \circ g' = g \circ f'$ and that f is monic:



Let $p_1, p_2\colon X \to P$ be any two arrows such that $f' \circ p_1 = f' \circ p_2$. Then $g \circ f' \circ p_1 = g \circ f' \circ p_2$, which by commutativity of the diagram implies that $f \circ g' \circ p_1 = f \circ g' \circ p_2$. But since f is monic this means that we must have $g' \circ p_1 = g' \circ p_2$. Hence we can write $h = f' \circ p_1$ and $k = g' \circ p_1$, and the following diagram commutes:



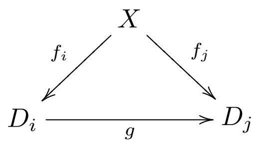
But this diagram is a pullback, and so there is a *unique* arrow $\psi\colon X \to P$ which makes everything commute. Hence $p_1 = p_2 = \psi$, i.e. f' is monic. □

4.5 Limits and colimits

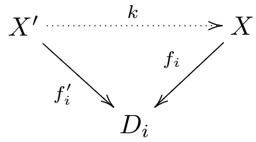
Many of the above constructions — terminal and initial objects, products and coproducts, equalisers and coequalisers, pullbacks and pushouts — turn out to be specific cases of the more general ideas of limits and colimits. In order to define these we first give a preliminary definition.

Definition 4.14. A *cone* for a diagram $\mathbf{D}$ in a category $\mathbf{C}$ is a $\mathbf{C}$-object X and a family of arrows $\{f_i\colon X \to D_i\}$ indexed over all the objects in $\mathbf{D}$ such that for each arrow $g\colon D_i \to D_j$ in

D the following diagram commutes:



Definition 4.15. A *limit* for a diagram **D** in a category **C** is a cone $\{f_i \colon X \to D_i\}$ such that whenever $\{f'_i \colon X' \to D_i\}$ is another cone for **D** then there is a unique arrow $k \colon X' \to X$ making the following diagram commute for each object D_i in **D**:



The dual notions of a *cocone* and a *colimit* are found by reversing the arrows.

In other words, if we consider the cones for a diagram as objects of a category, then the limits are the terminals objects in this category and are unique (up to isomorphism) because terminal objects are (see lemma 2.7). Similarly, colimits are the initial objects in the category of cocones.

Examples.

1. Let **C** be any category and let **D** be the diagram

$$A \qquad B$$

with two objects and no arrows. A cone for this diagram is an object X and two arrows f and g such that

$$A \xleftarrow{f} X \xrightarrow{g} B$$

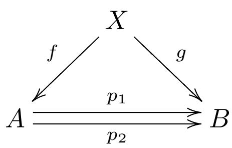
If a limit exists, then it is a product of A and B.

2. Let **D** be the diagram with no objects and no arrows. A cone for this diagram is just any object, since there are no objects to draw arrows to and no arrows to require commutativity. If a limit exists, it is an object X such that for any other object X' (i.e. for any cone for **D**) there is exactly one arrow from X' to X. That is, X is a terminal object.

3. Let **D** be the diagram

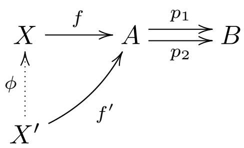
$$A \underset{p_2}{\overset{p_1}{\rightrightarrows}} B$$

A cone for this diagram is an object X together with arrows f and g such that



The commutativity of the diagram completely determines g; we must have $g = p_1 \circ f = p_2 \circ f$, so we may omit it from the diagram. A limiting cone is an object X and an arrow

$f\colon X \to A$ with $p_1 \circ f = p_2 \circ f$, such that whenever X' is another object and $f'\colon X' \to A$ has $p_1 \circ f' = p_2 \circ f'$ then there is a unique arrow $\phi\colon X' \to X$:



In other words, if a limit exists it is an equaliser of p_1 and p_2.

Dually, the colimits of the above examples are a coproduct of A an B, an initial object C, and a coequaliser of p_1 and p_2, respectively.

It is important to realise that limits (or colimits) may or may not always exist. In any given category limits may exist for some diagrams but not for others: for example, a category may have products but no equalisers. Categories in which limits exist for all finite diagrams are said to have *all finite limits*.

Howevere, if limits exist for basic diagrams then we can use this to find limits for more complicated ones. In fact it turns out that a limit exists for every diagram if two basic types of diagram have a limit. We finish by stating the following theorem without proof. The wording is taken from Pierce [**13**], p. 29.

Theorem 4.16 (Limit Theorem)**.** *Let* **D** *be a diagram in a category* **C***, with sets V of vertices and E of edges. If every V-indexed and every E-indexed family of objects in* **C** *has a product and every pair of arrows in* **C** *has an equaliser, then* **D** *has a limit.*

Proof. Omitted. See Pierce [**13**], pp. 29–32, Arbib and Manes [**2**], p. 47–48 or Mac Lane [**8**], p. 109. □

5 Concluding remarks

Category theory is an important subject both mathematically and philosophically. In both respects, a great deal of its importance is due to its unifying nature.

Due to the central notions of functors and natural transformations that allow the relationships between different categories, e.g. between categories of topological spaces and categories of particular algebraic structures, category theory has been key to the development of algebraic geometry and algebraic topology in the second half of the twentieth century, so much so that it is hard to imagine to what degree the subjects would have progressed without a categorical framework in which to work [**9**]. Category theory has also had a huge impact on mathematical logic, particularly since the development of the concept of a topos[14] with the field of categorical logic remaining an active area of research. Additionally, categories and functors have found applications further afield, ranging from cognitive science to theoretical computer science (e.g. in the design of functional programming languages) to mathematical physics, where so-called "higher dimensional categories" are used to study types of knots arising in quantum field theory.

The unifying nature of categories has also led to attempts to develop category theory as a foundation for mathematics, perhaps as a replacement for axiomatic set theory. However, it is difficult to say where category theory stands in this regard. See [**9**] and the articles referenced

[14]Very roughly speaking, a topos is a category with enough properties to be able to do normal maths in — for example **Set**. See McLarty [**10**].

therein for the philosophical implications of employing category theory, where objects and arrows are defined entirely abstractly, as a foundation for mathematics.

In any case, category theory provides an incredibly versatile framework in which complex mathematical ideas can be more easily expressed and understood, as well as a theory which is fascinating as a field in its own right. I aim to have shown that by abstracting the important properties of similar mathematical constructions in different contexts, general categorical ideas allow the underlying structure of many different ideas to been seen more clearly.

Jonathan Elliott, February 2007

Note The following list includes all the works I have consulted while writing this essay. Not all of them are cited directly, but I have included all those I feel have influenced my understanding on the ideas presented here.

References

1. PlanetMath, 2007. URL `http://planetmath.org`.

2. Michael A. Arbib and Ernest G. Manes. *Arrows, Structures and Functors – The Categorical Imperative.* Academic Press, Inc., 1975.

3. Steve Awodey. *Category Theory.* Oxford Logic Guides. OUP, 2006.

4. Ion Bucur and Aristide Deleanu. *Introduction to the Theory of Categories and Functors.* John Wiley & Sons, Ltd., 1968.

5. Peter J. Cameron. *Sets, Logic and Categories.* Springer, 1999.

6. Samuel Eilenberg and Saunders Mac Lane. General Theory of Natural Equivalences. *Transactions of the American Mathematical Society*, 58:231–294, 1945.

7. Horst Herrlich and George E. Strecker. *Category Theory – An Introduction.* Allyn and Bacon, Inc., Boston, 1973.

8. Saunders Mac Lane. *Categories for the Working Mathematician.* Springer, 1971.

9. Jean-Pierre Marquis. "Category Theory", The Stanford Encyclopedia of Philosophy, 2006. URL `http://plato.stanford.edu/entries/category-theory`.

10. Colin McLarty. *Elementary Categories, Elementary Toposes.* Clarendon Press, Oxford, 1995.

11. Barry Mitchell. *Theory of Categories.* Academic Press, Inc., 1965.

12. Bodo Pareigis. *Categories and Functors.* Academic Press, Inc., 1970.

13. Benjamin C. Pierce. *Basic Category Theory for Computer Scientists.* The MIT Press, 1991.

14. Eric W. Weisstein. Mathworld – A Wolfram Web Resource, 2007. URL `http://mathworld.wolfram.com`.

15. Wikipedia. Wikipedia, The Free Encyclopedia, 2007. URL `http://en.wikipedia.org`.