Category Theory: I

Notes towards a gentle introduction

Peter Smith

LOGIC MATTERS

This is very much work in progress, but I hope you find it helpful and interesting. Two friendly requests:

1. Please do send any comments, suggestions and corrections – however small! – to this address. They are always welcome. It can help a lot if you give the date of the PDF you are commenting on: this version is dated January 24, 2023.
2. Please do spread the word about this gentle intro. But please don't directly share this pdf or place it on a website as it could be superseded in a few days! Instead, share the url where the latest version can always be found, namely logicmatters.net/categories.

Contents

Preface to Part I

The project Some dozen years ago, I started putting together a set of notes on category theory, just as an exercise in getting things a bit clearer in my own mind. The notes eventually became quite substantial, and began to take on something of the shape of a book, a gently-moving beginner's guide of the kind that I myself would earlier have welcomed. So in 2015 I posted this approximation-to-a-book online, in the hope that some others might perhaps find my efforts helpful.

I revised those notes a little in 2018. But they remained very much work in sporadic progress; so there were chapters at different levels of development and with different degrees of integration with what was around them (and no doubt different levels of reliability). Despite their half-baked character, the notes have been steadily downloaded a lot, sometimes over a thousand times in a month. Which has been pretty surprising – pleasing in its way, but also increasingly embarrassing, as I of course knew all along how very flawed the draft notes were.

Of late, I have mostly had to let categories simmer on the back burner as I got on with other projects. But those distractions are at last finished, or at least can be set aside for a while in their turn, and I can from time to time rework these notes. The aim is at least to improve some of the technical exposition to make it as accessible as I can, and to add just a bit more discursive commentary.

For the moment, I am dividing the notes into two Parts – there is a rather natural division point, which I explain in §1.3. The second Part will have to stay unrevised for a while. And this first Part is still work in progress: revised or re-revised chapters get added when I think that they are at least significantly better than what they replace, not when I think they are as good as they could be! Needless to say, all comments and corrections will continue to be very gratefully received.

Who are these notes for? What do you need to bring to the party? I'm still writing up this material as much as anything for my own satisfaction. Whether my angle of approach and level of exposition will satisfy *you* will of course depend entirely on your interests, background, and preferences in matters of mathematical style. I can only suggest that you just dip in and see whether they might work for you.

One thing is pretty obvious from the outset: these notes are long given their coverage, because I *do* go at a gentle pace. I don't apologize at all for this: there are plenty of faster-track alternative introductions available for those who want one instead. However, it is not just my own experience which suggests that getting a secure understanding of category-theoretic ways of thinking by initially taking things slowly can make later adventures exploring beyond the basics *very* much more manageable.

I imagine one reader, then, to be a mathematics student who wants a first introduction to some categorial[1] ideas, perhaps as a preliminary warm-up before taking on an industrial-strength graduate-level course.

Another reader might be a philosopher interested in the foundations of mathematics who wants an accessible introduction to give them an initial sense of what the categorial fuss is about, so that they can tell if they then need to find out more. (But note, I only touch on a few conceptual issues here. Think of this as a review of the elementary mathematics of categories; it certainly is not intended as much of a contribution to the sometimes heated debates about the status of category theory e.g. as a supposed rival to set theory as a 'foundation' for mathematics.)

Now, category theory gives us a story about the ways in which different parts of modern abstract mathematics hang together. Obviously, you can't be in a good position to appreciate this if you really know *nothing* beforehand about modern mathematics! But I do try to presuppose relatively little. Suppose you know a few basic facts about groups (there's some revision in Chapter 2), know a little about different kinds of orderings, are acquainted with some elementary topological ideas, and know a few more bits and pieces; then you should in fact be able to cope fairly easily with the introductory discussions here. And if some later illustrative examples pass you by, don't panic. I usually try to give multiple illustrations of important concepts and constructs; so feel free simply to skip those examples that happen not to work so well for you.

How far do we aim to get? In these introductory notes, I only explore the beginnings of category theory. But what count as 'beginnings'? I'll be guided in part by the coverage of some avowedly introductory books. But I also note, for example, that the famous introduction to topos theory by Mac Lane and Moerdijk (1992) starts with a fourteen page chapter of 'Categorical Preliminaries'. This isn't supposed to be a stand-alone exposition so much as a checklist of assumed basics. That checklist turns out to correspond pretty closely to the overall coverage of the two Parts of these notes, which suggests that my menu of topics is sensible enough.

Theorems as exercises Almost all the proofs of the theorems you meet as you begin category theory are *very* straightforward. Almost always, you just have to 'do the obvious thing': there's little ingenious trickery needed at the outset. So you can think of the statement of a theorem as in fact presenting you with an exercise which you could, and even should, attempt to work through for yourself in order to fix ideas. The ensuing proof which I spell out is then the answer (or

[1] Logicians already have a quite different use for 'categorical'. So when talking about categories, I much prefer the adjectival form 'categorial', even though it is the minority usage.

at least, *an* answer) to the exercise. Sometimes theorems are explicitly stated as challenges for you to prove.

Notation and terminology I try to keep to settled notation and terminology, and where there are standard alternatives often mention them too: what I say here should therefore be easy to relate to other discussions of the same material.

'Iff', as usual, abbreviates 'if and only if'. In addition to using the familiar '□' as an end-of-proof marker or to conclude the statement of a theorem whose proof needn't be spelt out, I also use '△' as an end-of-definition marker.

And from now on, I mostly follow the usual mathematicians' practice of omitting quotation marks when mentioning symbolic expressions, if no confusion is likely to result. Logicians can get irritatingly fussy about this sort of thing, and I try to avoid that.

Thanks! Andrew Bacon, Malcolm F. Lowe and Mariusz Stopa very kindly sent long lists of corrections to an early ancestor of these notes. I had then further corrections to a revised version from Malcolm F. Lowe, David Ozonoff, Jan Thiemann, Zoltán Tóth, and Adrian Yee.

And, as I start work on revising the notes again, I've now had more corrections and suggestions on chapters in Part I, in particular from Sam Butchart and Rowsety Moid, together with Matthew Bjerknes. Very warm thanks to everyone!

1 Introduction

1.1 The categorial imperative

Modern pure mathematics explores abstract structures. And these mathematical structures cluster in families.

Take a family of structures together with a good helping of the structure-respecting maps between them. Then we can think of this inter-related family as forming a further structure – a structure-of-structures, if you like – something else to explore mathematically.

(1) Here's a basic example. A particular *group* is a structure which comprises some objects equipped with a binary operation defined on them, where the operation obeys the well-known requirements. But we can also think of a whole family of groups, together with appropriate maps between them – i.e. homomorphisms which respect group structure – as forming a further structure-of-structures.

(2) Another example: any particular *topological space* is a structure, classically conceived as comprising some objects, 'points', which are equipped with a topology. But again, a whole family of these spaces, together with appropriate maps between them – this time, the continuous functions which respect topological structure – forms another structure-of-structures.

(3) And so it goes. Perhaps what interests you are some *well-ordered objects*: these constitute another mathematical structure. In fact, there is a whole family of such well-ordered structures together with order-respecting maps between them. We are interested in the structure of this family (perhaps in the guise of the theory of ordinals, the theory of order-types of well-orderings). We want to know too about other kinds of families of ordered objects and the relations between them.

In each of these various cases, then, we not only investigate *individual* structures (the particular groups, particular topological spaces, particular collections of ordered objects), but we can also explore *families* of such structures (families of groups, families of topological spaces, families of ordered pluralities), with the family itself structured by the maps or morphisms between its members.

As a further step, we can next go on to consider the interrelations between these structures-of-structures. This will involve looking at an additional level of structure-respecting maps, the so-called *functors*, this time linking structures-of-structures (as when, for example, we map a family of topological spaces with base points to their corresponding fundamental groups). And even this is not the end of it. Going up yet another level of abstraction, we will find ourselves wanting to consider operations which map one functor to another while preserving their functorial character (in ways we will later explain).

So here is *one* central imperative of modern mathematics: to explore these upper levels of increasing abstract structure.

Let's agree straight away that this project certainly doesn't appeal to all – or even most – mathematicians. A vast amount of pure mathematics is of course carried on at much less exalted levels. Still, the hyper-abstracting project can resonate with a certain systematizing cast of mind. And evidently, if we *are* going to set out on such an enquiry, we will want a framework for dealing with these upper layers of abstraction in a disciplined and illuminating way.

This is where category theory comes into play for us (at least in these notes): it provides exactly what we need as we first set out to explore the territory, because suitably structured families of structures are prime examples of categories. Category theory's basic ideas and constructions will provide a general toolkit for systematically probing structures-of-structures and even structures-of-structures-of-structures. And it is the theory in *this* role that will be our main concern in this beginners' guide.

1.2 From a bird's eye view

But what do we really gain by ascending through those levels of abstraction and by developing tools for imposing some order on what we find?

For a start, we should get a richer conceptual understanding of how various parts of mathematics relate to each other. And I suppose we might reasonably say that, in *one* sense of that contested label, this will be a 'philosophical' gain. After all, many philosophers, pressed for a crisp characterization of their discipline, like to quote a famous remark by Wilfrid Sellars:

> The aim of philosophy, abstractly formulated, is to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term. (Sellars 1963, p. 1)

Category theory indeed provides us with a suitable unifying framework for exploring in depth some of the ways in which a lot of mathematics hangs together. That's why it should be of considerable interest to philosophers of mathematics as well as to mathematicians interested in the conceptual shape of their own discipline.

But category theory does much more than give us a helpful way of relating some aspects of structures that we already know about. As Tom Leinster so very nicely puts it, the theory

> . . . takes a bird's eye view of mathematics. From high in the sky, details become invisible, but we can spot patterns that were impossible to detect from ground level. (Leinster 2014, p. 1)

So category theory crucially enables us to reveal *new* connections we hadn't made before. What are called 'adjunctions' are a prime example, as we will eventually find.

Seeing recurrent patterns in different families of structures and making new connections between them in turn enables new mathematical discoveries. And it was because of the depth and richness of the resulting discoveries in e.g. algebraic topology that category theory first came to prominence. But it would be distracting to investigate those roots in these notes. I will stick to very much more elementary concerns, with an emphasis on unification and conceptual clarification. This will, it has to be said, give us a limited and partial picture: but we will still have more than enough to explore. And this way, I at least can hope to keep everything relatively accessible.

1.3 A slow ascent

The gadgets of basic category theory do fit together rather beautifully in multiple ways. These interconnections mean that there certainly isn't a single best route into the theory. Different treatments can take topics in significantly divergent orders, all illuminating in their various ways.

I will follow the simplest plan, however, and make a slow ascent to the categorial heights. We begin at that first new level of abstraction, one step up from talking about particular structures. In other words, we start by talking about *categories*. For, as I said, many paradigm cases of categories are in fact structured-families-of-structures. And we go on to develop ways of describing what happens inside a category. In this new setting, we revisit many familiar ideas about maps between structures, and about ways of forming new structures by e.g. taking products or taking quotients. Which gives us our topics for Part I of these notes.

Only after extended exploration of categories taken singly do we move up another level to consider *functors*, maps between categories (typically, maps between families of structures). And only after we have spent a number of chapters thinking about how particular functors work (and how they interact with products, quotients and the like) do we next move up a further level to define operations sending one functor to another – these are the so-called *natural transformations* and *natural isomorphisms*. We then explore these notions, and the related idea of one functor being a *representation* of another, at some length before we at last start exploring the key notion of *adjunctions*. All this will be covered in Part II, which ends with some brief pointers forward to other topics that come into view.

In summary, my chosen route here into the basics of category theory steadily ascends through the increasing levels of abstraction in a rather natural way. This route has logical appeal; but it does mean that we don't meet the *really* exciting and most novel categorial ideas until Part II. However, this disadvantage will (I hope) be counterbalanced – at least for some – by the gains in understanding which come from taking our gently sloping path. I will just have to do my best to make the ideas we encounter in Part I already seem pretty interesting and fruitful!

2 Groups, and categories of groups

Category theory gives us a framework in which we can think systematically about structured families of mathematical structures. One paradigm case of such a family, I said, comprises some groups organized by homomorphisms between them. But that's vague. By the end of this chapter, I will have explained more carefully what it takes to form a category of groups.

It might be helpful, though, to begin by reminding ourselves about some *very* elementary facts about groups. The point of this exercise is to highlight a few themes which are already there in pre-categorial mathematics. For example, we'll make some general points about products and quotients, and – more abstractly – note some ideas about what might count as 'objects' in structures and about 'identity up to isomorphism'. We will soon enough meet these themes again in a categorial guise: but first, here they are in a more familiar context.

Depending on your background, you may *very* well be able to skim-read this quick review at pace. Though you will probably spot straight away that the definitions I give in this chapter are in a key respect not *quite* the usual ones. I'll explain the reason for their mildly deviant character in Chapter 3; so indulge me for now!

2.1 Groups revisited

(a) I will use the likes of 'G' as a plural variable, to pick our some objects, one or many. And '$x \varepsilon G$' will abbreviate the claim that the particular object x is among the objects G. Then, with that notation to hand, we can take this as our preferred definition:

Definition 1. The objects G with a distinguished object e, equipped with a binary operation $*$, form a *group* iff

(i) G are closed under the operation $*$, i.e. for any $x, y \varepsilon G$, $x * y \varepsilon G$;
(ii) $*$ is associative, i.e. for any $x, y, z \varepsilon G$, $(x * y) * z = x * (y * z)$;
(iii) $e \varepsilon G$, and e acts as a group identity, i.e. for any $x \varepsilon G$, $x * e = x = e * x$;
(iv) every group object has an inverse, i.e. for any $x \varepsilon G$, there is at least one object $y \varepsilon G$ such that $x * y = e = y * x$. △

Don't read too much into 'equipped'. It's a standard turn of phrase here; but it means no more than that we are dealing with some objects G *and* an operation defined over it.

If e and e' are both identities for a group formed by G equipped with $*$, then $e = e * e' = e'$; so group identities are unique. Likewise, inverses are unique.

(b) Note immediately the variety of objects and operations that can form a group. In fact, *any* item e, whatever you like, together with the only possible binary operation $*$ such that $e*e = e$, forms a trivial one-object group. Similarly, any two items e, j, whatever you like, form a group when they are equipped with the binary operation $*$ for which e is the identity and $j * j = e$.

Less trivially, there are additive groups of numbers (e.g. the integers equipped with addition, or with addition mod n, with 0 as the identity), and there are multiplicative groups of numbers (e.g. the non-zero complex numbers equipped with multiplication, with 1 as the identity). These examples are *abelian*, i.e. the binary operation is commutative.

Likewise, there are groups of functions. For a simple case, take the group of permutations of the first n naturals, with functional composition as the group operation and the do-nothing permutation as the group identity. If $n > 2$, then this permutation group is non-abelian. Or consider groups of geometrical transformations – for instance the non-abelian group of symmetries of a regular polygon (i.e. the rotation and reflection operations which map the polygon to itself).

Then there are various groups of real invertible matrices, groups of closed directed paths through a base point in a topological space (with concatenation of paths as the group operation), and so on and on it goes. Groups are very many and various! But you knew all that.

(c) We need to agree some more notation. So let's use '$(G, *, e)$' simply to abbreviate '(the objects) G equipped with the operation $*$ and with distinguished object e'.[1] Similarly, of course, for e.g. '$(H, \star, d)$' etc. And when convenient we can abbreviate such expressions further by '$\mathcal{G}$', '$\mathcal{H}$', etc.

If $(G, *, e)$ satisfy the conditions for forming a group, then let's briskly write 'the group $(G, *, e)$' (or simply 'the group $\mathcal{G}$') rather than 'the group consisting in $(G, *, e)$'.

The group operation can be significantly different from case to case (all that is required is that it satisfies Defn. 1). But it is customary to default to using multiplicative notation and to talk generically of group 'products';[2] we will correspondingly default to denoting the inverse of a group object x by x^{-1}.

2.2 A quick word about 'objects'

There is a view, introduced into modern logic by Frege, according to which there are *absolute* type-theoretic distinctions to be made between objects (individual things) and first-level functions (sending objects to objects) and second-level functions (sending first-level functions to first-level functions), etc.

[1] So the parentheses here are just helpful punctuation.

[2] Additive notation, however, is commonly used when dealing with abelian groups.

Whatever the virtues of that view, I should emphasize that when we talk about the objects of a group, the notion of object in play here is a *relative* one. A group involves a group operation (a binary function of some type, whose inputs and outputs must be at the same type-level); and then this group's 'objects' are the items (of whatever type) which are the inputs and outputs for that operation. These items can be objects-as-individuals (like numbers); but the items can equally well be first-level functions (like permutations of some numbers, i.e. bijections between those numbers); or they can be of other types too.

I stress this obvious point in this familiar context because, looking ahead, we will meet just the same relative use of the notion of object when we get round to defining the general notion of a category.

2.3 New groups from old

(a) Given one or more groups, we can form further groups from them in various natural ways. For a start, there are subgroups, in the entirely predictable sense:

Definition 2. $(G', *, e)$ form a *subgroup* of the group $(G, *, e)$ iff (i) G' are some or all of G, and (ii) G' are closed with respect to the group operation and taking inverses (meaning that all $*$-products and $*$-inverses of objects among G' are also among G'). △

Example: the even integers (still with addition as the group operation, and with zero as the group identity) form a subgroup of the additive group of integers. For another example: the complex numbers on the unit circle form a subgroup of the multiplicative group of non-zero complex numbers.

(b) Next, products of groups. And, as a preliminary, we first need the general idea of a *pairing scheme*:

Definition 3. A scheme for pairing any objects from among G with any object from among G' provides

(i) some pair-objects O (which can be any suitable objects, and may or may not be already among G and/or G');

(ii) a binary pairing function which we can notate '$\langle\ ,\ \rangle$' which sends $x \varepsilon G$ and $x' \varepsilon G'$ to a pair-object $\langle x, x'\rangle \varepsilon O$ (where every one of the objects O is some such $\langle x, x'\rangle$);

(iii) two unpairing functions which send a pair-object $\langle x, x'\rangle$ to x and x' respectively. △

Note, it is immediate from this definition that the pairing function sends distinct pairs x, x' and y, y' to distinct pair-objects $\langle x, x'\rangle$ and $\langle y, y'\rangle$.

Don't jump to over-interpreting the notation here. The angle-brackets might remind you of some standard set-theoretic construction of ordered pairs. But all we need for a pairing scheme are *some* objects to 'code' for pairs together with interlocking pairing and unpairing functions. For example, if the objects G and G' are in both cases the natural numbers, then we could perfectly well take suitable pair-objects $\langle m, n\rangle$ to be the numbers 2^m3^n, with the obvious pairing and unpairing functions. What matters about pair-objects is not their intrinsic nature but the role they serve in a pairing scheme.

With Defn. 3 to hand, we can now define the notion of a product group:

Definition 4. Suppose we have the groups $\mathcal{G}$ and $\mathcal{G}'$, i.e. $(G, *, e)$ and $(G', *', e')$, together with some pairing scheme which maps $x \in G$ and $x' \in G'$ to a pair-object $\langle x, x'\rangle$. Let H be the resulting pair-coding objects. Define $d = \langle e, e'\rangle$, and define multiplication of pairs componentwise, so $\langle x, x'\rangle \star \langle y, y'\rangle = \langle x * y, x' *' y'\rangle$. Then $(H, \star, d)$ form a group, a *product* of the groups $\mathcal{G}$ and $\mathcal{G}'$ (which we can notate $\mathcal{G} \times \mathcal{G}'$). △

It is routine to check that $(H, \star, d)$ really do form a group.

For a very simple example, suppose $\mathcal{J}$ is a group comprising just the two objects e, j. If $\mathcal{K}_1$ is to be a product of $\mathcal{J}$ with itself, it will need to comprise four distinct objects $\langle e, e\rangle$, $\langle e, j\rangle$, $\langle j, e\rangle$, $\langle j, j\rangle$, with the first of these objects being the group identity. For brevity's sake, call these four pair-objects $1, a, b, c$ respectively. $\mathcal{K}_1$'s group operation $\star$ is then defined by the following table (read the table entry as giving the value of row-object $\star$ column-object):

$\star$	1	a	b	c
1	1	a	b	c
a	a	1	c	b
b	b	c	1	a
c	c	b	a	1

The symmetry of the table reflects the fact that $\mathcal{K}_1$ is abelian.

Note that we speak here of 'a' product of $\mathcal{J}$ with itself, not 'the' product. Why? Because there are unlimitedly many alternative schemes for coding pairs of objects, and different schemes will give rise to different product groups. In this present example, *any* four distinct objects we like can play the role of the required pair-objects, as long as we have pairing and unpairing functions to match. However, the resulting different groups *will* be equivalent-as-groups: any way of forming a product group from a two-object group and itself gives us a group describable by reinterpreting the same table.

The point of course generalizes. Products $\mathcal{G} \times \mathcal{G}'$ produced by using different pairing schemes will always be equivalent, in a familiar sense we'll clarify shortly.

(c) Now for a third, rather more interesting, way of forming new groups. We start with another general idea, and define a *quotient scheme*:

Definition 5. A scheme for quotienting the objects G by the equivalence relation $\sim$ defined over G provides

(i) some quotient-objects Q (which can be any suitable objects, which may or may not be among G),
(ii) a unary function which we can notate '[ ]' which sends $x \in G$ to a quotient-object $[x] \in Q$ (with every object among Q being some such $[x]$), where
(iii) for all $x, y \in G$, $[x] = [y]$ iff $x \sim y$. △

So $[x]$ behaves in the crucial respect like an $\sim$-equivalence class containing x.

But note, just as pair-objects in pairing schemes do not have to be sets, we similarly do *not* require $[x]$ to be an equivalence class or other set. For example, take the integers Z and consider the equivalence relation $\equiv_8$, i.e. congruence mod 8. Then we can simply put $[x]$ to be the remainder when x is divided by 8, and trivially $[x] = [y]$ iff $x \equiv_8 y$. Again, what matters about quotient-objects is the role they serve in a quotienting scheme. (And as with the parallel point about pairs, this point about quotients illustrates what will turn out to be an important motif of category theory.)

We can now define the notion of a quotient group in two steps:

Definition 6. (i) Given a group $\mathcal{G}$, i.e. $(G, *, e)$, then $\sim$ is a *congruence* relation for the group iff it is an equivalence relation on the group objects which respects the group structure of $\mathcal{G}$. In other words, for any objects $x, y, z \varepsilon G$, given $x \sim y$, then $x * z \sim y * z$ and $z * x \sim z * y$ (that is to say, 'multiplying' equivalent objects by the same object yields equivalent results).

(ii) Suppose we have a group $\mathcal{G}$, i.e. $(G, *, e)$ again, with $\sim$ is a congruence relation for the group. And suppose we also have a quotient scheme for $\sim$, which sends $x \varepsilon G$ to $[x]$. Let $[G]$ be all the objects $[x]$ for $x \varepsilon G$, and put $[x] \star [y] = [x * y]$. Then $([G], \star, [e])$ also form a group, a *quotient* of the original group $\mathcal{G}$ with respect to $\sim$, which we symbolize $\mathcal{G}/\sim$. △

For this definition to work, $\star$ has to be a genuine function. So we need to show that the result of $\star$-multiplication does not depend on how we pick out the multiplicands. In other words – *without* yet assuming $\star$ is a function so we can trivially substitute identicals! – we need to show that if $[x] = [x']$ then (i) $[x] \star [y] = [x'] \star [y]$, and (ii) $[y] \star [x] = [y] \star [x']$. But for (i), just note that if $[x] = [x']$, then by definition $x \sim x'$, hence (since $\sim$ respects group structure) $x * y \sim x' * y$, hence $[x * y] = [x' * y]$, hence by definition $[x] \star [y] = [x'] \star [y]$. We derive (ii) similarly.

It remains to check that $\mathcal{G}/\sim$ form a group with $\star$ the group operation. But that's straightforward.

Let's take a quick example. Using $\mathbb{Z}$ to denote the integers, $(\mathbb{Z}, +, 0)$ form an additive group, $\mathcal{Z}$ for short: and consider again the equivalence relation of congruence mod 8. This equivalence relation respects the additive structure of the integers; for if $x \equiv_8 y$ then $x + z \equiv_8 y + z$ and $z + x \equiv_8 z + y$. As suggested before, we can take our quotient scheme for this equivalence relation simply to send x to the remainder on dividing x by 8; this gives us as quotient-objects the eight numbers from 0 to 7, call them $\overline{8}$. Then $(\overline{8}, +_8, 0)$ evidently form a group (where $+_8$ is addition mod 8), which is a quotient $\mathcal{Z}/\equiv_8$.

Two points to note. First, for the definition of $\star$ to be in good order and provide a group operation on $[G]$, it is essential that the equivalence relation $\sim$ on G which defines $[G]$ is not just any old equivalence relation, but is a congruence respecting the group structure on G.

Second, we again talk of 'a' quotient group rather than 'the' quotient group. There will be many ways of finding quotient schemes for $\sim$, hence many alternative pluralities $[G]$ from which to build a quotient group $\mathcal{G}/\sim$ (though, as with product groups, quotient groups constructed using different quotient schemes will all 'look the same').

2.4 Group homomorphisms

(a) Let's now equally briskly recall some basic facts about structure-respecting maps between the groups.

Definition 7. A *group homomorphism* from the group $(G, *, e)$ as source to the group $(H, \star, d)$ as target is a function f defined over G with values among H such that:

(i) for every $x, y \varepsilon G$, $f(x * y) = f(x) \star f(y)$,
(ii) $f(e) = d$. △

So a homomorphism sends products in the source group to corresponding products in the target group. It similarly sends the identity object in the source group to the identity in the target group.

Since $f(x) \star f(x^{-1}) = f(x * x^{-1}) = f(e) = d$, and similarly $f(x^{-1}) \star f(x) = d$, $f(x^{-1})$ is the inverse of $f(x)$. In other words, a homomorphism also sends inverses to inverses.

Thought of simply in its role of mapping objects to objects, the function $f\colon G \to H$ is said to be the underlying function of the homomorphism. When thought of in its role as a structure-respecting homomorphism we can use the notation $f\colon (G, *, e) \to (H, \star, d)$, or $f\colon \mathcal{G} \to \mathcal{H}$.

(b) Some initial trivial examples:

(1) Let $(G, *, e)$ form a group $\mathcal{G}$, and let $\mathsf{1}$ be any one-object group. Then there is a homomorphism $f\colon \mathcal{G} \to \mathsf{1}$, which sends every object among G to the sole object of the target group. And this is the unique homomorphism from $\mathcal{G}$ to $\mathsf{1}$.

(2) Likewise, there is a homomorphism $g\colon \mathsf{1} \to \mathcal{G}$ which sends the single object of $\mathsf{1}$ to the group identity of $\mathcal{G}$. And this is the unique homomorphism from $\mathsf{1}$ to $\mathcal{G}$.

(3) Relatedly, there is always a 'collapse' homomorphism $h\colon \mathcal{G} \to \mathcal{G}$ which sends every $\mathcal{G}$-object to its group identity e.

These cases remind us that, although homomorphisms are often described as *preserving* group structure, this doesn't mean replicating *all* structure. A homomorphism from $\mathcal{G}$ to $\mathcal{H}$ can compress many or most aspects of $\mathcal{G}$'s structure simply by mapping distinct $\mathcal{G}$-objects to one and the same $\mathcal{H}$-object. It really is better, then, to talk of homomophisms as *respecting* group structure.

Three more elementary examples:

(4) There is a homomorphism from $\mathcal{Z}$, the additive group of integers $(\mathbb{Z}, +, 0)$, to any two object group $\mathcal{J}$ which sends even numbers to $\mathcal{J}$'s identity, and

sends odd numbers to $\mathcal{J}$'s other object. The underlying function here is surjective but not injective.

(5) There is a homomorphism from $\mathcal{Z}$ to $\mathcal{Q}$, the additive group of rationals $(\mathbb{Q}, +, 0)$, which sends an integer n to the corresponding rational $n/1$. As a function from $\mathbb{Z}$ to $\mathbb{Q}$, this is injective but not surjective.

(6) Let $\mathbb{R}$ be the real numbers, and $\mathbb{C}^*$ the non-zero complex numbers. The reals form a group under addition, and the non-zero complex numbers form a group under multiplication. Define $j\colon (\mathbb{R}, +, 0) \to (\mathbb{C}^*, \times, 1)$ by putting $j(x) = \sin x + i \cos x$. Then we have a homomorphism whose underlying function is neither injective nor surjective.

(c) Let's pause to see what can be said about group homomorphisms in general, various though they have already proved to be.

Theorem 1. *(1) Any two homomorphisms $f\colon \mathcal{G} \to \mathcal{H}$, $g\colon \mathcal{H} \to \mathcal{J}$, with the target of the first being the source of the second, will compose to give a homomorphism $g \circ f\colon \mathcal{G} \to \mathcal{J}$.*

(2) Composition of homomorphisms is associative. In other words, if f, g, h are group homomorphisms which can compose so that one of $h \circ (g \circ f)$ and $(h \circ g) \circ f$ is defined, then the other composite is defined, and the two composites are equal.

(3) For any group $\mathcal{G}$, there is a trivial identity homomorphism $1_{\mathcal{G}}\colon \mathcal{G} \to \mathcal{G}$ which sends each object to itself. Then for any $f\colon \mathcal{G} \to \mathcal{H}$ we have $f \circ 1_{\mathcal{G}} = f = 1_{\mathcal{H}} \circ f$.

Proof sketch. For (1) we simply take $g \circ f$ ('g following f') applied to an object $x \varepsilon \mathcal{G}$ to be $g(f(x))$ and then check that $g \circ f$ so defined does satisfy the condition for being a homomorphism given that g and f do.

For (2), associativity of homomorphisms is inherited from the associativity of ordinary functional composition for the underlying functions.

(3) is also immediate. □

(d) Note the important fact that this, our very first theorem, is *not* a mere logical consequence of our definitions of groups and group homomorphisms. Our proof sketch plainly depends on invoking background assumptions about functions, such as the assumption that functional composition is associative.

And so it goes. Almost *nothing* in group theory just follows from the definitions alone!

2.5 Group isomorphisms and automorphisms

(a) Now we highlight the special case where the underlying function of a homomorphism is both injective and surjective, so it gives rise to a nice one-to-one correspondence between two groups (or between a group and itself).

Definition 8. A *group isomorphism* $f\colon \mathcal{G} \xrightarrow{\sim} \mathcal{H}$ is a homomorphism where the underlying function is a bijection between the respective objects of the two groups.

We say that the groups $\mathcal{G}$ and $\mathcal{H}$ are *isomorphic* as groups iff there is a group isomorphism $f\colon \mathcal{G} \xrightarrow{\sim} \mathcal{H}$, and then write $\mathcal{G} \simeq \mathcal{H}$

A *group automorphism* is a group isomorphism $f\colon \mathcal{G} \xrightarrow{\sim} \mathcal{G}$ whose source and target are the same. △

Again, let's have some elementary examples:

(1) Any two two-object groups are isomorphic. Take e, j equipped with the only possible group operation $*$, and e', j' equipped with $*'$. Then the map which sends the group identity e to the group identity e' and sends j to j' is obviously a group isomorphism.

(2) There are two automorphisms from the additive group $(\mathbb{Z}, +, 0)$ to itself. One is the trivial identity homomorphism; the other is the function which sends an integer j to $-j$.

(3) There are infinitely many automorphisms from the group $(\mathbb{Q}, +, 0)$ to itself. Take any non-zero rational q: then the map $x \mapsto qx$ 'stretches/compresses' the rationals, perhaps reversing their order, while still preserving additive structure.

(4) Let $\mathcal{K}_2$ be the group consisting in $1, 3, 5, 7$ equipped with multiplication mod 8 as the group operation. And let $\mathcal{K}_3$ be the group of symmetries of a non-equilateral rectangle whose four 'objects' are the operations of leaving the rectangle in place, vertical reflection, horizontal reflection and rotation through 180°, with the group operation being simply composition of geometric operations. Then $\mathcal{K}_2 \simeq \mathcal{K}_3$.

The easiest way to see this is by constructing the abstract 'multiplication table'. First, take $1, a, b, c$ to be respectively the numbers $1, 3, 5, 7$, and take $\star$ to be multiplication mod 8. Second, take $1, a, b, c$ to be the geometric operations on a rectangle in the order just listed and take $\star$ to be composition. Both times we get the same table as for $\mathcal{K}_1$ that we met in §2.3. Matching up the two new interpretations of $1, a, b, c$ and the two corresponding interpretations of $\star$ gives us the claimed isomorphism $f\colon \mathcal{K}_2 \xrightarrow{\sim} \mathcal{K}_3$. By the same reasoning, both groups are isomorphic to $\mathcal{K}_1$.

This illustrates an obvious general point. Groups that can interpret the same 'multiplication table' are isomorphic; conversely, isomorphic groups can be described by the same (possibly infinite) table.

(5) In defining a product of two groups, we were allowed to invoke any scheme for coding pairs of objects from the two groups. But whichever scheme we choose, the resulting product (I said) will 'look the same', and have the same multiplication table. We can now put it like this: suppose $\mathcal{H}_1$ and $\mathcal{H}_2$ are both products $\mathcal{G} \times \mathcal{G}'$; then $\mathcal{H}_1 \simeq \mathcal{H}_2$.

Why? Just take the bijection which sends the pair-object $\langle x, x' \rangle_1$ which pairs $x \mathrel{\varepsilon} \mathcal{G}$ and $x' \mathrel{\varepsilon} \mathcal{G}'$ according to the pairing scheme used in constructing $\mathcal{H}_1$ to the corresponding pair-object $\langle x, x' \rangle_2$ formed according to the pairing scheme used in constructing $\mathcal{H}_2$. This is trivially seen to be a group isomorphism from $\mathcal{H}_1$ to $\mathcal{H}_2$.

Likewise, suppose $\mathcal{H}_1$ and $\mathcal{H}_2$ are different quotients of a group $\mathcal{G}$ with respect to a congruence relation $\sim$, different because they rely on different quotient schemes for, in effect, representing $\sim$-equivalent classes of objects from $\mathcal{G}$. Take the bijection that sends the quotient-object $[x]_1$ according to the first quotient scheme to the corresponding object $[x]_2$ according to the second scheme. Then by a similar argument we again have $\mathcal{H}_1 \simeq \mathcal{H}_2$.

(b) Another very easy result, for future reference:

Theorem 2. *A group homomorphism $f\colon \mathcal{G} \to \mathcal{H}$ is an isomorphism iff it has a two-sided inverse, i.e. there is a homomorphism $g\colon \mathcal{H} \to \mathcal{G}$ such that $g \circ f = 1_{\mathcal{G}}$ and $f \circ g = 1_{\mathcal{H}}$.*

Proof. Suppose $f\colon (G, *, e) \to (H, \star, d)$ is an isomorphism. Then the underlying function $f\colon G \to H$ is a bijection and has a two-sided inverse $g\colon H \to G$. We now need to show that this inverse function g gives rise to a homomorphism $g\colon (H, \star, d) \to (G, *, e)$. But since f is a homomorphism, $(fgx \star fgy) = f(gx * gy)$; and so, since g is a two-sided inverse for f, we have $g(x \star y) = g(fgx \star fgy) = gf(gx * gy) = gx * gy$. In addition, as required, $gd = gfe = e$.

Conversely, suppose f is a homomorphism with a two-sided inverse. Then its underlying function must have a two-sided inverse; but it is a familiar elementary result that a function with a two-sided inverse is a bijection. □

Evidently, a group is isomorphic to itself (by the identity homomorphism) and the composition of two group isomorphisms is again an isomorphism. Given that isomorphisms are homomorphisms with two-sided inverses which are homomorphisms, it is also immediate that the inverse of an isomorphism is also an isomorphism. Therefore, just as we would want,

Theorem 3. *Being isomorphic is an equivalence relation between groups.* □

2.6 Another way of forming new groups from old

Take any group $\mathcal{G}$ and consider its automorphisms $Aut_{\mathcal{G}}$. There is of course at least one such automorphism, namely the identity map $1_{\mathcal{G}}$. Note too that any two of $\mathcal{G}$'s automorphisms f, g compose to give us a new automorphism $g \circ f$. Composition here is associative. And we've just noted that isomorphisms in general, and hence automorphisms in particular, have inverses with respect to composition. Hence:

Theorem 4. *For any group $\mathcal{G}$, $(Aut_{\mathcal{G}}, \circ, 1_{\mathcal{G}})$ form a group, the automorphism group of $\mathcal{G}$, $AUT(\mathcal{G})$.* □

For example, we've already remarked that there are exactly two automorphisms from $\mathcal{Z} = (\mathbb{Z}, +, 0)$ to itself; so $AUT(\mathcal{Z})$ is a two-object group. And what is the automorphism group of that two-object group? A trivial one-object group.

By contrast, since 'stretching/compressing by a non-zero rational' is an automorphism for the additive group $\mathcal{Q} = (\mathbb{Q}, +, 0)$, and such operations can be composed by multiplying the stretching factor, the corresponding automorphism group $AUT(\mathcal{Q})$ will be isomorphic to the multiplicative group of non-zero rationals.

For one more example, look at the 'multiplication table' for $\mathcal{K}_1$ again. We see that if we swap the three entries a, b, c around, we keep the same structure. So $AUT(\mathcal{K}_1)$ will be a group of permutations of three objects. And what does the automorphism group of *that* look like? It turns out to be the same again, a group of permutations of three elements. What fun!

2.7 Homomorphisms and constructions

In §2.3 we considered some basic ways of forming new groups from old, yielding subgroups, product groups and quotient groups. In §2.4 we introduced structure-respecting maps between groups. We now bring the two themes together, foreshadowing an absolutely central motif of category theory.

(a) For the simplest case, start by noting how homomorphisms give rise to subgroups and vice versa.

Theorem 5. *For any homomorphism $f\colon \mathcal{G} \to \mathcal{H}$, the f-image of $\mathcal{G}$ is a subgroup of $\mathcal{H}$. Conversely, for every subgroup of $\mathcal{H}$, we can pick a $\mathcal{G}$ and homomorphism $f\colon \mathcal{G} \to \mathcal{H}$ such that the relevant subgroup is the f-image of $\mathcal{G}$.*

Proof. Given a group homomorphism $f\colon (G, *, e) \to (H, \star, d)$, let $f[G]$ be the objects which are f-images of objects among G, so they include d, i.e. $f(e)$. Then $(f[G], \star, d)$ form a group, which we can denote $f(\mathcal{G})$ and which will be a subgroup of $\mathcal{H}$. Why?

(i) Suppose $y_1, y_2 \mathrel{\varepsilon} f[G]$. By assumption, these objects are f-images of some objects $x_1, x_2 \mathrel{\varepsilon} G$. So we have $y_1 \star y_2 = fx_1 \star fx_2 = f(x_1 * x_2)$, and hence $y_1 \star y_2 \mathrel{\varepsilon} f[G]$ as required.

(ii) Since $\star$ is associative and d an identity for that operation, it only remains to show that if $y \mathrel{\varepsilon} f[G]$ then its inverse is in $f[G]$ too. But y is by assumption $f(x)$ for some object $x \mathrel{\varepsilon} G$, and homomorphisms send inverses to inverses. So the inverse of y, i.e. $(fx)^{-1}$, is $f(x^{-1})$ and hence is in $f[G]$.

That establishes the first half of our theorem. For the converse half, just note that any subgroup $\mathcal{G}$ of $\mathcal{H}$ gives rise to a trivial injection map $i\colon \mathcal{G} \to \mathcal{H}$ which sends an object in $\mathcal{G}$ to the same object now considered as an object of $\mathcal{H}$. □

Hence we can characterize the subgroups of a given group $\mathcal{H}$ in terms of group-homomorphisms with the target $\mathcal{H}$. Putting it roughly, then, we can trade in claims about what goes on *inside* various groups when forming subgroups for claims about corresponding homomorphisms *between* groups.

(b) I'll quickly mention another essential link between homomorphisms and subgroups.

We start by noting another important idea (which will be very familiar if you have done even a little group theory):

Definition 9. $(N, *, e)$ is a *normal subgroup* of $(G, *, e)$ iff it is a subgroup and, for any $n \varepsilon N$ and any $g \varepsilon G$, $g * n * g^{-1} \varepsilon N$.

As a first illustration, we have the following little result:

Theorem 6. *If $\sim$ is a congruence relation on $(G, *, e)$ and N are the objects which are $\sim$-equivalent to e, then $(N, *, e)$ is a normal subgroup of $(G, *, e)$.*

Proof. We need to check that $(N, *, e)$ in fact form a group. So first, we need to show that if $n, n' \varepsilon N$, then their group product $n * n' \varepsilon N$. But by assumption $n \sim e$, and so by Defn. 6, the definition of congruence, $n * n' \sim e * n'$. But $e * n' = n'$ and by assumption $n' \sim e$. Hence $n * n' \sim e$, and so $n * n' \varepsilon N$.

A similar argument shows that if $n \varepsilon N$ then $n^{-1} \varepsilon N$. And trivially N contains the group identity. So $(N, *, e)$ do indeed form a group.

To establish normality, we need to show that if $n \varepsilon N$ then so is $g * n * g^{-1}$ for any group object g. But $\sim$ is a congruence and by definition $n \sim e$; therefore we know that $g * n \sim g * e = g$. Hence by congruence again $g * n * g^{-1} \sim g * g^{-1} = e$, as required. □

And now the result linking homorphisms to (normal) subgroups:

Theorem 7. *Suppose $f\colon (G, *, e) \to (H, \star, d)$ is a group homomorphism, and let K be the objects among G which f maps to the identity d in H. Then $(K, *, e)$ form a normal subgroup of $(G, *, e)$, the* kernel *of f.*

Proof. We need to show first that K are in fact closed under the group operation $*$. But suppose $k_1, k_2 \varepsilon K$. Then $f(k_1 * k_2) = fk_1 \star fk_2 = d \star d = d$. Hence $(k_1 * k_2) \varepsilon K$.

By the definition of a homomorphism, $f(e) = d$, so the identity $e \varepsilon K$. Then recall that homomorphisms send inverses to inverses. Therefore if $f(k) = d$, then $f(k^{-1}) = d^{-1} = d$; so the inverse of an object among K is also one of the objects K. Hence $(K, *, e)$ form a subgroup of $(G, *, e)$.

For normality, we simply note that for any $k \varepsilon K$ and $g \varepsilon G$, $f(g * k * g^{-1}) = f(g) \star f(k) \star f(g^{-1}) = f(g) \star d \star f(g)^{-1} = d$, so $g * k * g^{-1} \varepsilon K$ as required. □

There is a converse theorem too, that every normal subgroup for a group $\mathcal{G}$ is the kernel of some homomorphism with the source $\mathcal{G}$.

So again, we can trade in claims about what goes on *inside* various groups, making them normal subgroups, for claims about corresponding homomorphisms *between* groups.

(c) I'll skip past product groups for now, and next consider quotient groups arising from suitable equivalence relations. We then have the following result:

Theorem 8. *Given a group homomorphism* $f\colon \mathcal{G} \to \mathcal{H}$*, and* x, y *among* $\mathcal{G}$*'s objects, put* $x \sim y$ *iff* $fx = fy$*. Then* $\sim$ *is a congruence on* $\mathcal{G}$*, and* $f(\mathcal{G})$*, the* f*-image of the group* $\mathcal{G}$*, is a quotient group* $\mathcal{G}/{\sim}$*. Conversely, given a quotient group of* $\mathcal{G}$ *with respect to a congruence relation* $\sim$*, we can find a group* $\mathcal{H}$ *and homomorphism* $f\colon \mathcal{G} \to \mathcal{H}$*, such that* $\mathcal{G}/{\sim}$ *is* $f(\mathcal{G})$*.*

Proof. The relation $\sim$ of being equalized-by-f is trivially an equivalence relation. But we need to check that $\sim$ respects $\mathcal{G}$'s group operation $*$ so that $\mathcal{G}/{\sim}$ exists. In other words, we need to show that for any group objects x, y, z, given $x \sim y$, then (i) $x * z \sim y * z$ and (ii) $z * x \sim z * x$. But for (i), if $x \sim y$, then $fx = fy$, hence $f(x * z) = fx \star fz = fy \star fz = f(y * z)$, hence $x * z \sim y * z$ (here, $\star$ is of course $\mathcal{H}$'s group operation). Case (ii) is exactly similar.

By the definition of $\sim$, the f-images of objects among G act like quotient-objects with respect to $\sim$; so it is immediate that $f(\mathcal{G})$ is a quotient group $\mathcal{G}/{\sim}$.

For the converse result, suppose $\mathcal{G}/{\sim}$ is a quotient of $\mathcal{G}$ with respect to some equivalence relation $\sim$, with $f_\sim : x \mapsto [x]$ giving us the relevant quotient scheme. Then $f_\sim : \mathcal{G} \to \mathcal{G}/{\sim}$ is easily checked to be a homomorphism, and $f_\sim(\mathcal{G})$ is the whole of $\mathcal{G}/{\sim}$. □

So again we can trade in claims about the structure of certain groups, this time about their quotient structure, for corresponding claims about homomorphisms between groups.

And note further that this trade reveals something that was not obvious before, namely that there is a kind of duality between quotient groups and subgroups: given a homomorphism $f\colon \mathcal{G} \to \mathcal{H}$, $f(\mathcal{G})$ is a quotient of $\mathcal{G}$ and a subgroup of $\mathcal{H}$.

(d) Similarly, it turns out that claims about the structure of product groups can also be traded in for claims about corresponding homomorphisms between groups. But I'll leave the proof of this for later. For now, I'll just flag up again the key general point that these sorts of trades – i.e. trades between claims about the 'internals' of structures and claims about 'external' maps between structures – will turn out to be a pivotal theme of category theory.

2.8 'Identical up to isomorphism'

Let's pause over another important point. We have met the groups $\mathcal{K}_1, \mathcal{K}_2, \mathcal{K}_3$ which are isomorphic to each other. They are also isomorphic to any other group whose four objects can be labelled $1, a, b, c$ in such a way that the same 'multiplication table' in §2.3 applies again. Call such groups *Klein four-groups*. And note, the way in which the various Klein four-groups differ from each other, namely in the internal constitution of their various *objects*, is not relevant to their core behaviour as groups, for that depends just on the *functional relations between the* *objects* induced by the group operation. In other words, despite the differences between their objects, the groups are the same at least as far as their structural properties – i.e. the properties as determined by their shared 'multiplication table' – are concerned.

A bit of care is needed in describing the situation, however. Consider, for example, the following from a rightly well-regarded algebra text:

> The groups $\mathcal{G}$ and $\mathcal{H}$ are isomorphic if there is a bijection between them which preserves the group operations. Intuitively, $\mathcal{G}$ and $\mathcal{H}$ are the same group except that the elements and the operations may be written differently in $\mathcal{G}$ and $\mathcal{H}$. (Dummit and Foote 2004, p. 37)

But that surely isn't a happy way to putting things. We have just reminded ourselves that $\mathcal{K}_1$, $\mathcal{K}_2$ and $\mathcal{K}_3$ are isomorphic groups. But, for example, $\mathcal{K}_2$ comprises four *numbers* as its objects, and $\mathcal{K}_3$ comprises four *operations* on a non-equilateral rectangle; and there is no sense in which numbers and geometric operations can be thought of as the same things 'written differently'. If anything, then, it is exactly the other way around: we have here distinct groups comprising different elements and different group operations which, however, can be 'written the same', being represented by the same table under different interpretations.

A rather happier, and widely used, way of putting things is this: our Klein groups $\mathcal{K}_1$, $\mathcal{K}_2$ and $\mathcal{K}_3$ are *identical up to isomorphism*. And for many purposes, group theory can simply ignore the differences between groups which are identical up to isomorphism. Hence the frequently encountered talk of '*the*' Klein group. We will eventually have to return to the question of quite what such talk can amount to.[3]

2.9 Categories of groups

(a) That will do by way an initial review of some basic facts about groups and homomorphisms between them. Let's now ask: given some groups and some homomorphisms between them, what does it take for these to form a structured family which counts as a *category* of groups?

We impose just two very natural conditions. First, the homomorphisms in the category should be closed under composition. And second, for each of the groups in the category, its trivial identity homomorphism needs to be included. And that's *all* it takes: it really is that simple!

[3]To fly a kite: Occasionally, we meet versions of the the idea that, as well as 'concrete' Klein groups, i.e. Klein groups whose elements have an independent nature (which could be numbers, pairs of numbers, rotations and reflections, whatever), there is also a more purely 'abstract' Klein group. This has the right multiplication table, but is supposedly built up from objects with no properties at all over and above being sent to each other by the group operation according to the given table. And then it is this group comprising these abstract de-natured elements which is then said to be, properly speaking, *the* Klein group. But does this suggestion make sense?

Still, let's say the same thing again, but this time in more laborious detail, for clarity's sake:

Definition 10. A *category of groups* comprises

(1) some groups *Grp*, and
(2) some homomorphisms *Hom*,

where these are governed by the following conditions:

Sources and targets If $f\colon \mathcal{G} \to \mathcal{H}$ is among *Hom*, then both its source $\mathcal{G}$ and its target $\mathcal{H}$ are among *Grp*.

Composition If $f\colon \mathcal{G} \to \mathcal{H}$, $g\colon \mathcal{H} \to \mathcal{I}$ are both among *Hom*, where the target of f is the source of g, then $g \circ f\colon \mathcal{G} \to \mathcal{I}$ is also among *Hom*.

Identity homomorphisms If $\mathcal{G}$ is among *Grp*, the corresponding identity homomorphism $1_{\mathcal{G}} : \mathcal{G} \to \mathcal{G}$ is among *Hom*.

Further, we have:

Associativity of composition. For any $f\colon \mathcal{G} \to \mathcal{H}$, $g\colon \mathcal{H} \to \mathcal{I}$, $h\colon \mathcal{I} \to \mathcal{J}$ among *Hom*, $h \circ (g \circ f) = (h \circ g) \circ f$.

Identity homomorphisms do behave as identities. For any $f\colon \mathcal{G} \to \mathcal{H}$ among *Hom*, $f \circ 1_{\mathcal{G}} = f = 1_{\mathcal{H}} \circ f$. △

Of course, we know the last two conditions will automatically be satisfied because of Theorem 1. But I'm (redundantly) mentioning those conditions again here so that our definition of categories of groups matches up nicely with our general definition of categories in Chapter 4.

(b) Just as groups are many and various, so too are categories of groups. For example, a single group $\mathcal{G}$ together with its identity homomorphism $1_{\mathcal{G}} : \mathcal{G} \to \mathcal{G}$ counts as a trivial category of groups. So too does any uncommunicative bunch of groups equipped only with their identity homomorphisms.

But those are *very* unexciting cases! Things can get more interesting when the groups in a category start to communicate (so to speak).

Consider next, then, the category which collects all the finite groups whose objects are some natural numbers together with all the isomorphisms between those groups. Now there is a *bit* of structure to the category, with the isomorphic groups at least connected together by the maps between them. But this is still of relatively little interest: we have different islands of isomorphic groups, and a group inhabiting one island knows nothing about groups inhabiting other islands.

So let's move on to consider the category comprising those same finite groups but this time combined with *all* the homomorphisms between them (whether isomorphisms or not). And *now* non-isomorphic groups can 'see' each other. So we have enough homomorphisms in play to be able e.g. distinguish the one-object groups in the category by saying that these are the groups which have one and only one homomorphism to and from every other group, as noted in §2.4. We can also use these homomorphisms to tell a story about e.g. subgroups and quotient groups living in the category, as indicated in our preliminary sketch in §2.7. Developing this sort of story will be a primary item of business in the coming chapters.

(c) And there will be lots more categories of groups – e.g. the category of symmetry groups of finite regular polygons and the homomorphisms between them, the category of infinite abelian groups of natural numbers and their homomorphisms, and so on and so forth.

Now those categories, those relatively small-scale families-of-structures, are all tamely unproblematic. But now let's ask: will there also be a mega-category of *all* groups and *all* the homomorphisms between them?

Good question! To get a handle on it, there are some troublesome issues we need to tangle with, at least in a preliminary way. Let's make this the business for the next chapter.

2.10 Other families of structures?

Now, we could pause here to consider at similar length another family of structures or two.

For example, we could have a chapter on the lovely example of topological spaces. There will be many obvious parallels in the story. We would again meet appropriate notions of substructures, products and quotients. We would meet the structure-respecting continuous maps between two spaces, and the continuous maps with continuous inverses which make for topological isomorphisms. We could talk too about the way that we can trade in some claims about the structure of spaces for claims about the availability of continuous maps between spaces. And so on.

However, although this would be an instructive exercise, I've decided with some regrets to skip doing this. Those who already know a bit of topology probably won't need the parallels spelt out. While for those who don't have enough background, a fair amount of motivational scene-setting would be required to make the story seem natural, and that would slow us down just when we are getting rather impatient to find out more about categories. So let me press on.

3 But where do categories of groups live?

Where do mathematical structures live? Where do groups, and the families of interconnected groups that form categories of groups, live? 'In the universe of sets', comes the speedy standard reply.

But let's not rush on too fast. It's well worth pausing to think a little about what this conventional answer buys us. And is it in fact compulsory to go set-theoretic?

3.1 Sets, virtual classes, plurals

(a) Following Cantor, I'll understand a set – strictly so called – to be a single object, a thing in itself over and above its members (so the 'set of' operator takes zero, one, or many things, and outputs a distinct new thing).

However, if this is the guiding conception, then the first thing to say is that a great deal of elementary informal talk of sets or classes is really no more than a façon de parler. Yes, it is a useful and now very familiar idiom for talking about many things at once. But in a whole range of elementary contexts informal talk of a set or class doesn't really carry any serious commitment to there being any *additional* object over and above those many things. In other words, singular talk of *the set/class of Xs* can very often be traded in without loss for plural talk of *the Xs*.

Here is Paul Finsler writing a century ago, emphasizing the key distinction we need (and adding a bit of linguistic stipulation):

> It would ... be inconvenient if one always had to speak of many things in the plural; it is much more convenient to use the singular and speak of them as a class. ... A class of things is understood as being the things themselves, while the set which contains them as its elements is a single thing, in general distinct from the things comprising it. ... Thus a set is a genuine, individual entity. By contrast, a class is singular only by virtue of linguistic usage; in actuality, it almost always signifies a plurality. (Finsler 1926, p. 106, quoted in Incurvati 2020, p. 3.)

Finsler writes 'almost always', I take it, because a class term may in fact denote just one thing, or even – perhaps by misadventure – none.

Nothing at all hangs, of course, on the stipulative choice of the particular words 'set' vs 'class' to mark the distinction. What matters is the contrast between uneliminable talk of sets in Cantor's sense of entities in their own right and, on the other hand, non-committal and eliminable talk. And here is Quine making the key point in a later and more famous passage:

> Much … of what is commonly said of classes with the help of '∈' can be accounted for as a mere manner of speaking, involving no real reference to classes nor any irreducible use of '∈'. … [T]his part of class theory … I call the virtual theory of classes. (Quine 1963, p. 16)

This same usage plays an important role in set theory itself in some treatments of so-called 'proper classes' as distinguished from sets. For example, in his standard book *Set Theory*, Kenneth Kunen writes

> Formally, proper classes do not exist, and expressions involving them must be thought of as abbreviations for expressions not involving them. (Kunen 1980, p. 24)

But, just to complicate matters, other developments of set theory do allow for proper classes (classes too big to be sets) to count as entities in their own right. So we can't reliably use 'class' and be expect to be understood in Finsler's way. For occasional later use, then, let's equally stipulatively adopt 'plurality' instead, for when we want to speak in the singular about perhaps many things at once without committing ourselves to an entity over and above those many things.

(b) Note, thought, that Finsler rather exaggerates the supposed inconvenience of plural talk. There is nothing at all unusual or forced about the use of plural terms in mathematics. Consider, for example, terms such as 'the complex fifth roots of 1', 'the real numbers between 0 and 1', 'the points where line L intersects curve C', 'the finite groups of order 8', 'the premisses' (of a certain argument), 'Hilbert's axioms for geometry', 'the symmetries of a rectangle', 'the ordinals', etc., etc. Mathematicians habitually use such terms which, taken at face value, refer plurally, to many things; and they use them without the slightest sense of strain or impropriety.

And don't be tempted by the thought that, all the same, we should really construe informal plural talk about *the X's* as disguised singular talk referring to *the set of X's*.

You in fact already know that we can't always construe talk of the Xs, plural, as being about some corresponding set of Xs. We can't, for example, trade in universally generalizing plural talk about the ordinals for singular talk about the set of ordinals because there *is* no set of ordinals (there are as many ordinals as sets – set-many, for short – and that is too many to form a set, at least according to standard set theories). And leaving aside such cases, eliminating plural talk in favour of singular talk of sets turns out to be impossible to do systematically without a lot of ad hoc and implausible contortions.

It would be far too distracting to pursue the detailed arguments here. But the headline proposition is that plural talk is in perfectly good logical order as it is, without needing to be re-interpreted as referring to sets in anything like Cantor's sense. And that is still true, even if your measure of being in good logical order is formalizability.[1]

(c) Why am I fussing about this last point? Because when we turn to talking about categories, we'll find that a lot of interesting examples are *large* in the sense that they comprise too many structures and too many maps between them to form sets. So we *can't* in general define a category as a *set* of structures suitably equipped with maps between them. One option would be to talk instead of a class of structures, a virtual class, Finsler/Quine-style. But it is less likely to lead to confusion if, at least here at the beginning of our discussion of categories, we use frankly plural idioms. That's in fact what I did at the end of the last chapter, when I defined the idea of a category of groups.

3.2 One 'generous arena' in which to pursue group theory

(a) Back to our initial presentation of some elementary group theory in the last chapter. There too I proceeded cheerfully in a plural idiom and simply avoided talk of sets.

Of course, this was the mildly deviant aspect of the mode of presentation. I used the likes of 'G' as a *plural* variable, to pick out some objects, and defined a group as comprising some objects G suitably equipped with a group operation. More conventionally we would deploy 'G' as a *singular* term, picking out a set, and say instead that a group consists in a set suitably endowed with a group operation. Instead of writing '$x \varepsilon G$' to say that x is among the objects G (plural) we would conventionally write '$x \in G$' to say that x belongs to the set G (singular). And so on. But would this invocation of sets be doing any real extra work here?

I'm with Paulo Aluffi, who explicitly acknowledges at the beginning of his fine book *Algebra, Chapter 0* that the informal set idiom which he adopts in the book is actually "little more than a system of notation and terminology" (Aluffi 2009, p. 1). That seems right. The story of elementary group theory will unfold in the same way in either system of notation and terminology, whether we habitually use (potentially eliminable) singular talk of sets or classes or pluralities, or use frankly plural talk: part of the point of the last chapter was to make this claim seem tenable.

Still, even if we grant that we can get a considerable way into informal group theory without explicitly referring to sets, some might well argue that this a merely superficial point, because there is a sense in which a theory of sets is still essentially required to be there in the background. Why so?

(b) As we noted before, even as soon as we reach our trite Theorem 1 in the last chapter, we are in fact going beyond the mere logical consequences of our

[1] For a lot more on why we shouldn't try to eliminate plurals, and for an extended formal treatment of how to argue with plural terms and plural quantifiers, taking them at face value, see e.g. Oliver and Smiley (2016).

definitions of groups and group homomorphisms. So what do we in fact need to bring to the table to get group theory going? Answer:

(i) the usual mathematical stock-in-trade of a body of assumptions about *functions*, together with

(ii) a repertoire of available *constructions*.

For example, we assume that functions always do compose when they can (i.e. when the target of the first is the source of the second), and that composition is associative. We assume that a function with a two-sided inverse is a bijection. We assume that it makes determinate sense e.g. to talk about *all* the permutations of some given objects, or *all* the automorphisms on a given group. And so on, and so forth. These, of course, look quite unproblematic assumptions – but they are needed all the same.

Again, we typically assume that we can construct what will serve as ordered pairs ad libitum; and we assume that whenever an equivalence relation partitions some objects we can somehow represent these partitions. More carefully, in our earlier terms, we assume pairing schemes and quotient schemes are available whenever we want them. And going forward, we will assume that we can not only construct pairs, triples, and finite tuples more generally, but we can form infinite sequences too. We also need to assume that we can freely construct multiple ‘copies’ of whatever structures we already have. And so on, and so forth again.

That’s all pretty vague, but quite intentionally so. I am simply gesturing at the way that standard textbook developments of group theory simply help themselves from the outset to a bunch of unproblematic background assumptions as needed as we go along. Fair enough. But what if we want to start getting more explicit and methodical about these background assumptions? Suppose we want to regiment these assumptions and organize them into a neat package – what package would suffice?

(c) Shelve that question for just a moment, and consider a different issue arising from what we earlier said about groups.

Here’s what seems a silly question. I cut out a cardboard non-equilateral rectangle: have I hereby brought into a being a new Klein four-group, the group of *this* new rectangle’s own (approximate!) rotation/reflection symmetries?

Well, we were previously entirely permissive about where we can find our groups: on our definition, we just need some new objects (in the broad sense) and a suitable operation on them and we get a new group. But on the other hand, a new physically realized Klein group is surely neither here nor there as far as the mathematics of groups is concerned. As I said before, group theory will for most purposes ignore the differences between groups which are identical up to isomorphism.

OK: suppose that there is a capacious enough fixed abstract mathematical universe in which we can implement isomorphic copies of all the different kinds of groups we will ever want. Then we just won’t care about any additional copies of these groups which are (as it were) roaming outside in the wild, or popping into existence when I cut up a new bit of cardboard (ok, we might well care about physically realized groups when doing applied mathematics; but we won't care for the purposes of pure group theory).

This last thought prompts, then, a more sensible question: where can we find a suitably rich mathematical universe in which we can construct all the groups we want?

(d) We have two related questions: what package of assumptions about available functions and about structure-building constructions will suffice for group theory? where can we can find (at least copies of) all the groups we want, neatly corralled together?

And there is of course a *very* familiar joint answer! The universe of sets provides exactly the sort of generous arena where there is a plenitude of groups along with other mathematical structures, together with all the functions and constructions we want for ordinary mathematical purposes.[2] It provides the desired foundation, in one sense, for group theory.

It might reasonably be claimed, therefore, that *that* is why it is entirely appropriate to talk about sets right from the very outset in doing group theory (for example). Once the wraps are off, once we make explicit the assumptions there in the background which we need to get our theory up and running, we will find that our theory really is set-theoretic through and through.

3.3 Alternative implementations?

Or so the story goes. A moment's reflection, however, suggests that the argument rushes on too fast at the end. Yes, a suitable set theory may provide *one* generous arena in which we can implement all the gadgets we need in developing group theory. *But why suppose that it the only option*?

(a) We are so used to being told that various mathematical widgets and whatnots are to be defined as sets of one kind or another that it can take a bit of effort to loosen the grip of that doctrine. Given our overall project, though, this is worth doing. So let's backtrack for a moment and and focus just on the simple core case of implementing one-place functions. What's the standard set-theoretic story, and is it compulsory?

Fix on some way of implementing ordered pairs as sets, e.g. as Kuratowski pairs $\langle x, y\rangle_K = \{\{x, y\}, x\}$. Then here is a familiar and entirely unproblematic definition:

Definition 11. Given a function f that maps objects X into the objects Y, $f\colon X \to Y$, the *graph* of f is the set $\hat{f}$ of ordered pairs $\langle x, y\rangle_K$ where x is among the objects X and y is among Y, and $fx = y$. △

Then an equally familiar orthodoxy, at least in its baldest and most unqualified form, *identifies* a function f with its graph $\hat{f}$.

[2]The phrase 'generous arena' is borrowed from the very helpful discussion of the idea of set-theoretic foundations in Maddy (2017).

However, we really should resist any such outright identification. For a start, to play on the set-theorists' own turf for a moment, let's consider the function which maps an object to its singleton. Then – by the set-theorists' own lights – it doesn't have a graph: the totality of pairs $\langle x, \{x\}\rangle_K$, pairing-up every set x with its singleton, is the size of the universe of sets and so is 'too big' to be a set. Likewise, the function which maps every ordinal to its successor is also 'too big' to have a graph. Therefore not all functions can be identified with their graphs.

Just one counterexample is enough to defeat a universal claim. It might be suggested, though, that the cases where a function relates too many things to be a set are in some sense rogue cases. So, in a concessive spirit, let's put such cases aside for a moment and see where that gets us.

Well, next note that treating a function as a set of ordered pairs involves arbitrary choices of implementation scheme.

(i) It is arbitrary to fix on Kuratowski's implementation of pairs. Other set-theoretic pairing schemes will work just as well.

(ii) Even relative to a choice of pairing scheme, we could equally well model a function by the set of pairs $\langle y, x\rangle$ where $f(x) = y$, rather than by the set of pairs $\langle x, y\rangle$ – and some textbooks do just this. And again other choices are possible.

However, if various permutations of choices at stages (i) and (ii) are pretty much as workable as each other, then we surely can't suppose that – when we choose to equate a function with its graph as we conventionally just defined it – we have made the uniquely *right* choice, i.e. the choice that correctly identifies which set that function 'really' is. And if there is no fact of the matter about which set a given function is, then we can't flat-out identify the function with some set such as its graph.

(b) We can dig deeper: *a function and its graph belong to different logical types* – that's fundamentally why they *can't* be identical.

Alonzo Church makes the key observation when he writes that

> it lies in the nature of any given [one-place] function to be applicable to certain things and, when applied to one of them as argument, to yield a certain value. (Church 1956, p. 15)

For example, a function such as the factorial defined over the natural numbers is, of its nature, the type of gadget which yields a numerical value for a given number as argument. By contrast a set doesn't, of its nature, take an argument or yield a value. And what applies to sets in general applies e.g. to sets of ordered pairs of numbers (graphs of numerical functions) in particular.

In insisting on a fundamental type-distinction between functions and objects, Church is here following Frege, whose metaphor of 'unsaturatedness' might be helpful. The picture is that functions of their nature are 'unsaturated', have a certain number of empty slots waiting to be filled appropriately when the function is applied to the right number of arguments. By contrast, an object like a set is already 'saturated', with no empty slots waiting to be filled.

In sum, a set of ordered pairs $\hat{f}$ can't *by itself* do the work of a function f, taking arguments and yielding output values. As the mathematician Terence Tao, who has no philosophical axe to grind, briskly puts it in his introductory book on analysis,

> functions are not sets, and sets are not functions; it does not make sense to ask whether an object x is an element of a function f, and it does not make sense to apply a set A to an input x to create an output $A(x)$. (Tao 2016, p. 51)

Which *of course* isn't to deny that we can make use of the graph of a function (a glorified input-output look-up table) in mapping an input object to an output value. But to do this, we need to deploy *another* function, namely a two-place evaluation function which takes an object x and the graph, and outputs y if and only if the pair $\langle x, y\rangle_K$ is in the graph. And unless we are planning to set off on an infinite regress, we had better not seek to again trade in this evaluation function for another set.

So a function, strictly speaking, isn't a set. But what we can do in a set-theoretic environment is *implement* functions as graphs;[3] and we can then transmute a claim about a function into a corresponding set-theoretic claim about some set of ordered pairs. (Though, to complicate the story, there is typically another step. Suppose we are considering, say, a one-place function of natural numbers. Then yes, we can implement this as a set of ordered pairs in a suitable universe of sets. But these won't be pairs of numbers – since strictly speaking numbers aren't themselves sets either[4] – but rather pairs of whatever-sets-we-choose-as-proxies-for-numbers. So if $\hat{m}$ is our preferred set-representative for the natural number m, a numerical claim $f(m) = n$ is then mirrored by a set-theoretic claim of the form $\langle \hat{m}, \hat{n}\rangle \in \hat{f}$.)

Similarly, relations strictly speaking aren't sets either. The only genuine relation to be found in the world of sets is the set-membership relation; but what we can do in a set-theoretic environment is implement other relations by their extensions. So we can then mirror a claim about a relation by a claim about its extension.

And I'll say more about the set-theoretic handling of pairs and quotients and the like in due course.

(c) What is the point of insisting that the story about functions-as-graphs doesn't tell us what functions 'really' are, but rather reports one way of implementing functions in the universe of sets? Am I just splitting hairs? I hope not! Rather, as announced, I'm trying to loosen the grip of the standard identification of functions with their graphs, and to make room for the thought that

[3] No word really seems ideal. Talk of 'proxies', 'surrogates', 'representations' has variously misleading connotations. I'll lean mostly to talk of implementation, as that is common enough and is at least relatively colourless.

[4] The locus classicus for this point is Paul Benacerraf's – very readable! – 'What numbers could not be' (1965).

there might in fact be other attractive ways of theorizing about the functions of ordinary mathematics in other foundational frameworks.

And now return to thinking about groups, and families of groups. If the groups of ordinary group-theory comprise some objects equipped with a suitable binary function, and binary functions aren't to be outright identified with sets, then what we can find in the set-theoretic universe should strictly be speaking be regarded as implementations of, or proxies for, groups. Fine! This again isn't for a moment to deny that these set-theoretic proxies can serve certain theoretical purposes brilliantly well. I am certainly *not* in the business of scorning the business of implementing mathematical structures in a set-theoretic framework. I am just emphatically highlighting that we *are* here in the implementation business. And looking at things at that way, we can more easily see that shouldn't too hastily assume that a set-theoretic framework provides the *only* general arena, the only foundational framework, in which we can find a plenitude of surrogates or proxies for implementing the mathematical structures and constructions which we want to regiment and study.

Indeed, we can't rule out that an alternative choice of general framework *might* even do the job rather better in some respects (maybe with different costs and benefits accruing to the different choices). For example, treating all mathematical widgets and whatnots as if they are sets seemingly gives rise to such daft questions as 'is the square root function for complex numbers a member of π?'. We can block such foolish questions by using some type-disciplined framework which more strongly distinguishes types of entities in our mathematical universe in the way that practicing mathematicians habitually do: so a modern type theory could be the way to go. Or perhaps an approach which is category theoretic in flavour will be illuminating. Here's the logician Dana Scott, thinking about functions in particular:

> What we are probably seeking is a 'purer' view of functions: a theory of functions in themselves, not a theory of functions derived from sets. What, then, is a pure theory of functions? Answer: category theory. (Scott 1980, p. 406)

Scott quickly goes on to remark that the general notion of a category won't give us enough. But arguably a *topos* (that's a particularly sort of category which we'll eventually meet) does provide another sort of universe in which we can regiment much of our ordinary mathematics. Such suggestions would be very puzzling if we have already jumped too quickly to assuming that ordinary mathematics is already quite fixedly set-theoretic through and through.[5]

[5]It should be noted – though we can't really pursue this interesting theme here – that different frameworks, giving us different theories of functions, can also at the margins give us different answers to questions of 'ordinary mathematics'. Indeed, even different set-theoretic frameworks will make available implementations of different functions, so can change what e.g. group homomorphisms we have available, and hence determine different answers to various group-theoretic problems. For some examples of questions which get different answers in different set universes, see tinyurl.com/groupqns.

3.4 'The' category of groups?

(a) We will eventually want to elaborate more on the options for generous arenas, foundational frameworks, in which we can pursue ordinary mathematics. But for now, as I said, we just want to leave the door ajar to the possibility of not-traditionally-set-theoretic frameworks; we do not want to foreclose options too soon.

Our more immediate concern, however, is to get back to the question we left hanging at the end of §2.9: does it make sense to talk about a mega-category of *all* groups and *all* the homomorphisms between them?

Well, we can now see that it *does* make sense, if we suppose that we do fix on *some* generous enough determinate framework in which we can implement enough of the different kinds of groups and group-theoretic gadgets we want – whether that is the entirely predictable candidate, namely a large enough universe of sets, or some alternative. Then yes, in this chosen arena, we can sensibly talk about an inclusive mega-category which comprises all the implementations of groups living in *that* universe and all the homomorphisms between them.

(b) Fine. So the thought is that, if we are going to talk about inclusive mega-categories, in the way that category theorists do, we need to assume that there is *some* generous foundational framework, at least there in the background.

But now we face something of a presentational dilemma. In developing category theory – at least when thinking of it as our abstract theory for handling structures of structures – there seems to be a rather good reason for *not* baldly assuming from the start that our mathematical structures all have their home in a universe of sets. For, given what I've just said, we should want to leave room for the possibility of other foundational frameworks, other generous arenas in which we can find enough structures for our ordinary mathematical purposes. Indeed we might eventually want to develop such an arena using categorial ideas. So we might rather like to remain neutral about where categories, structures of structures, live.

However, remaining studiously neutral would require some slightly awkward manoeuvres at key points. And much more importantly, it would be – at this stage of the game – a potentially distracting departure from the more usual way of presenting category theory. For the usual working assumption is that categories *do* live in some world of sets; and it seems sensible to remain for a good while at least in close parallel to other introductions.

Now, to assume that we *are* ultimately working in some set-theoretic framework isn't yet uniquely to pin down a chosen arena for category theory. We are very familiar with the fact that our canonical set theory first-order ZFC has multiple models (some where the continuum hypothesis holds, some where it doesn't, and so on and so forth, in a proliferating multiverse of models). How do we fix on the particular set-universe we might want to work in? Even if we go second-order, that still doesn't determine a unique universe – there are different models of second-order ZFC with different heights. But then, on further reflection, do we need such rich universes as models of full ZFC for implementing ordinary mathematical structures and structured-families-of-structures? It has been argued e.g. that the much weaker Mac Lane set theory is quite strong enough to model the standard mathematics which is not directly connected with the wilder extravagances of set theory (a salient point in the present context, Mac Lane being himself one of the founding fathers of category theory). And more radically deviant set theories like NFU arguably also provide competent generous arenas for modelling the gadgetry of ordinary mathematics.

Anyway, after the usual initial genuflection towards set theory, presentations of category theory quite often leave it open exactly how the set-theoretic details might get filled in.[6] That's surely a sensible approach: first put category theory to work as informal mathematics, issuing a promissory note to later determine the best shape of a set-theoretic universe in which we could regiment the theory (if that's what we want to do). On reflection, I too propose going along this this approach for now – except we'll try to keep at the back of our minds that eventually we might want to consider the possibility of some rather different, not-conventionally-set-theoretic, arena in which all the action can be taken to be happening. In other words, eventually we can perhaps begin to untie our developing category theory from its initial anchorage in a world of sets. But there is a great deal of ground to cover first. Let's see how all this pans out.

(c) OK, then: the headline summary is that we are going to be assuming pro tempore that the categories which concern us do live in some suitably generous universe of sets without being very specific about what that is. And given this stance – to answer that question we left hanging – it *does* make sense to define an inclusive category Grp, a category which contains all groups and group homomorphisms as implemented in that universe. So at least proxies for all the groups we might care about can be found in Grp.

And in the next chapter, we will start meeting many more inclusive categories like this. But before that, we of course first need to widen our discussion from talking about categories of groups to talking about categories much more generally. So that's our next item of business.

[6] Presentations differ on this, of course. For example, in his classic *Categories for the Working Mathematician* Mac Lane presents the axioms for category theory as in our §4.1, before saying that a category is "any interpretation of the category axioms within set theory". Then in a section titled 'Foundations' he does make a first pass at outlining the kind of set theory, an extension of ZFC, he has in mind (Mac Lane 1997, p.10, pp. 22–23). By contrast, in his much-used textbook *Categories*, Steve Awodey (2010) is rather less committal, explicitly allowing the possibility of working in other background systems than set theory. He defaults, though, to a set theoretic framework, while noting that "we sometimes run into difficulties with set theory was usually practiced" and leaving it somewhat open what is the best way to handle that fact: "we will not worry about this when it is just a matter of technical foundations" (Awodey 2010, 24–25).

4 Categories in general

We have met only one sort of category so far, namely categories comprising some groups and enough homomorphisms between them. Here, 'enough' stands in for pretty minimal requirements – essentially just that (i) compositions of homomorphisms in the category are also in the category, and (ii) the identity homomorphism for each group in the category is also present.

We now make our real start on category theory by generalizing to ...

4.1 The very idea of a category

(a) We said that many paradigm examples of categories are – as in our first illustrative case of categories of groups – families of structures with structure-respecting maps between them. But what can we say about such families at an abstract level?

One sufficiently general thought is this: if, within a family of structures including A, B, and C we have a structure-respecting map f from A to B, and another structure-respecting map g from B to C, then we should to be able to compose these maps. That is to say, the first map f followed by the second g should also count as a structure-respecting map $g \circ f$ from A to C.

What principles will govern such composition of maps? Associativity, surely. Using a natural diagrammatic notation, if we are given maps

$$A \xrightarrow{\quad f \quad} B \xrightarrow{\quad g \quad} C \xrightarrow{\quad h \quad} D$$

it really ought not matter how we carve up the journey from A to D. It ought not matter whether we apply the map f followed by the composite g-followed-by-h, or alternatively first apply the composite map f-followed-by-g and then afterwards apply h.

What else can we say at the same level of stratospheric generality about families of structures and structure-respecting maps? Very little! Except that there presumably will always in principle be the limiting case of a 'do nothing' identity map, which applied to any structure A leaves it untouched.

That apparently doesn't give us a great deal to work with. But in fact it is already enough to shape our following definition of categories. However, it is useful to abstract even further from the idea of structures with structure-respecting maps between them, and – using more neutral terminology – we'll now speak very generally of *objects* and of *arrows* between them. Then we say:

Definition 12. A category C comprises two kinds of things:

(1) C-*objects* (which we will typically notate by $A, B, C, \ldots$)

(2) C-*arrows* (which we typically notate by $f, g, h, \ldots$).

These C-objects and C-arrows are governed by the following axioms:

Sources and targets For each arrow f, there are unique associated objects $src(f)$ and $tar(f)$, respectively the *source* and *target* of f, not necessarily distinct. We write $f\colon A \to B$ or $A \xrightarrow{f} B$ to notate that f is an arrow with $src(f) = A$ and $tar(f) = B$.

Composition For any two arrows $f\colon A \to B$, $g\colon B \to C$, where $src(g) = tar(f)$, there exists an arrow $g \circ f\colon A \to C$, 'g following f', which we call the *composite* of f with g.

Identity arrows For any given any object A, there is an arrow $1_A\colon A \to A$ called the *identity arrow* on A.

We also require the arrows to satisfy the following further axioms:

Associativity of composition. For any $f\colon A \to B$, $g\colon B \to C$, $h\colon C \to D$, we have $h \circ (g \circ f) = (h \circ g) \circ f$.

Identity arrows behave as identities. For any $f\colon A \to B$ we have $f \circ 1_A = f = 1_B \circ f$. △

Evidently, a category of groups as originally defined in §2.9 will be a category in this sense. And given what we have already said, the objects which are mathematical structures of a particular kind taken together with enough arrows which are structure-respecting maps between them should also satisfy those axioms, and hence should count as forming a category too.

(b) Here are six quick remarks on terminology and notation:

(i) The objects and arrows of a category are very often called the category's *data*. That's a helpfully non-committal term if you don't read too much into it, and I will occasionally adopt this common way of speaking.

(ii) The label 'objects' for the first kind of data is quite standard. But note that, just as with the 'objects' of groups (see §2.2), the 'objects' in categories needn't be objects-as-individuals in a type-theoretic sense which contrasts objects with entities like relations or functions. There are perfectly good categories whose objects are actually relations, and other categories where they are functions.

(iii) Borrowing familiar functional notation $f\colon A \to B$ for arrows in categories is entirely natural given that arrows in many categories *are* (structure-respecting) functions: in fact, that is the motivating case. But again, as we'll soon see, not all arrows in categories are functions. Which means that not all arrows are morphisms either, in the usual sense of that term. Which is why I rather prefer the colourless 'arrow' to the equally common term 'morphism' for the second sort of data in a category. (Not that that will stop me talking of morphisms or maps when context makes it natural!)

(iv) In keeping with the functional paradigm, the source and target of an arrow are frequently called, respectively, the 'domain' and 'codomain' of the arrow (for usually, when arrows are functions, that's what the source and target are). But that usage has the potential to mislead when arrows aren't functions (or aren't functions 'in the right direction', cf. §**??**), which is why I prefer our common alternative terminology.

(v) Note again the order in which we write the components of a composite arrow, because some from computer science writing about categories do things the other way about. Our notational convention is again suggested by the functional paradigm. When $f\colon A \to B$, $g\colon B \to C$ are both functions in the ordinary sense, then $(g \circ f)(x) = g(f(x))$. Occasionally, to reduce clutter, we may write simply 'gf' rather than '$g \circ f$'.

(vi) Initially, we will explicitly indicate which object an identity arrow has as both source and target, as in '1_A'. Again to reduce clutter, we will later allow ourselves simply write '1' when context makes it clear which identity arrow is in question.

4.2 Identity arrows

The definition of a category implies our first mini-result:

Theorem 9. *Identity arrows on a given object are unique; and the identity arrows on distinct objects are distinct.*[1]

Proof. For the first part, suppose A has identity arrows 1_A and $1'_A$. Then applying the identity axioms for each, we immediately have $1_A = 1_A \circ 1'_A = 1'_A$.

For the second part, we simply note that $A \neq B$ entails $src(1_A) \neq src(1_B)$ which entails $1_A \neq 1_B$. □

So there's a one-one correlation between objects in a category and identity arrows; and we can pick out such identity arrows by the special way they interact with all the other arrows. Hence we could in principle give a variant definition of categories framed entirely in terms of arrows.[2] But I am not unusual in finding this bit of trickery rather unhelpful. As we will see, a central theme of category theory is indeed the idea that we should probe the objects in a category by considering the arrows between them; but that's no reason to write the objects out of the story altogether.

[1] As in this case, the most trivial of lemmas, as well as run-of-the-mill propositions, interesting corollaries, and the weightiest results, will continue to be labelled 'theorems' without distinction. I did initially try to mark a distinction between, as-it-were, capital-'T' theorems and unexciting lemmas and the rest, but that didn't work out well!

[2] For an account of how to do this, see Adámek et al. (2009, pp. 41–43).

4.3 Monoids and pre-ordered pluralities

Let's continue by looking at two simple but instructive types of categories, one algebraic, one order-theoretic.

(a) We have already met the example of various small-scale categories of groups and the inclusive large category Grp. But it is worth thinking now about a case where the algebraic structure is cut nearer to the bone.

Consider the finite strings of symbols from some given alphabet, including the limiting case of the empty string, together with the operation of concatenation. This operation is evidently associative, $s_1{}^\frown(s_2{}^\frown s_3) = (s_1{}^\frown s_2){}^\frown s_3$. And concatenating with the empty string leaves us where we were, so the empty string acts like an identity element for concatenation. So this structure gives us an example of a *monoid* – which is, so to speak, a group minus the requirement for inverses. And a monoid homomorphism is then a function which respects monoid structure.

More carefully, we have:

Definition 13. The objects M with a distinguished object e, equipped with a binary operation $*$, form a *monoid* $\mathcal{M} = (M, *, e)$ iff

(i) M are closed under the operation $*$, i.e. for any $x, y \varepsilon M$, $x * y \varepsilon M$;
(ii) $*$ is associative, i.e. for any $x, y, z \varepsilon M$, $(x * y) * z = x * (y * z)$;
(iii) $e \varepsilon M$, and e acts as a monoid unit or identity, i.e. for any $x \varepsilon M$, $x * e = x = e * x$.

Further, a *monoid homomorphism* from $(M, *, e)$ as source to $(N, \star, d)$ as target is a function $f\colon M \to N$ such that:

(i) for every $x, y \varepsilon M$, $f(x * y) = fx \star fy$,
(ii) $f(e) = d$. △

Just as in the case of groups, when thought of simply in its role of mapping objects to objects, the function $f\colon M \to N$ is said to be the underlying function of the homomorphism. When thought of in its role as a structure-respecting homomorphism we can use the notation $f\colon (M, *, e) \to (N, \star, d)$, or $f\colon \mathcal{M} \to \mathcal{N}$.

(b) It is evident that, as in the group case, monoid homomorphisms $f\colon \mathcal{M} \to \mathcal{N}$ and $g\colon \mathcal{N} \to \mathcal{O}$ compose to give a homomorphism $g \circ f\colon \mathcal{M} \to \mathcal{O}$. Composition of homomorphisms is associative. And the identity function on M is a homomorphism $f\colon \mathcal{M} \to \mathcal{M}$ which acts as an identity with respect to composition.

Hence, just as with groups, some monoids together with enough homomorphisms will form a category – where by 'enough' we mean as before that (i) compositions of homomorphisms in the category are also in the category, and (ii) the identity homomorphism for each monoid in the category is also present.

And assuming now that we are working in some capacious universe which contains proxies for all the monoids we want together with set-proxies for their homomorphisms, we can also sensibly say:

(C1) Mon is the category whose objects are all the monoids and whose arrows are all the monoid homomorphisms (as living in that universe).

Fine print. Yes, we could insist again that what we have in Mon will strictly speaking be implementations of monoids and their homomorphisms. But still, these implementations of monoids and their homomorphisms can count perfectly well as objects and arrows in a category. In particular, note that arrows in a category don't have to be kosher functions. So, in a slogan, Mon is a genuine category of (proxies for) monoids, and not a proxy category!

(c) Next, an example involving ordered objects; and again we'll cut structure to the bone by considering the simplest case, pre-orderings.

Definition 14. The objects P equipped with a relation $\preccurlyeq$ form a pre-ordered plurality $(P, \preccurlyeq)$ iff, for all $a, b, c \,\varepsilon\, P$,

(i) if $a \preccurlyeq b$ and $b \preccurlyeq c$, then $a \preccurlyeq c$,
(ii) $a \preccurlyeq a$.

A monotone map $f\colon (P, \preccurlyeq) \to (Q, \sqsubseteq)$ between pre-ordered pluralities is then defined to be a function $f\colon P \to Q$ which respects order, i.e. such that for any $a, b \,\varepsilon\, P$, if $a \preccurlyeq b$, then $fa \sqsubseteq fb$. △

Here, recall 'plurality' is just doing duty as a convenient singular way of talking about perhaps many things at once: see §3.1.

(d) It is obvious that monotone maps between pre-ordered pluralities will compose to give monotone maps; and the identity map on some pre-ordered objects gives rise to an identity monotone map on that plurality.

Evidently, then, we can we have categories with the following data:

(i) objects: various pre-ordered pluralities $(P, \preccurlyeq)$,

(ii) arrows: enough monotone maps between these various objects,

where (and we won't keep repeating this) 'enough' means the maps are closed under composition and each $(P, \preccurlyeq)$ gets its own identity map.

OK, now assume again that we are working in some suitably capacious set-universe. So for any objects P equipped with a pre-order $\preccurlyeq$ which we care about there is a corresponding pre-ordered *set* which represents it – by abuse of notation we'll again use $(P, \preccurlyeq)$ to denote it. Then we can sensibly give this definition:

(C2) Preord is the category whose objects are *all* pre-ordered sets $(P, \preccurlyeq)$ in that universe, and whose arrows are the monotone set-functions between *them*.

4.4 A very quick word about notation

I'm already falling into a pattern which I will try to stick to pretty systematically. I'll use sans serif font (as in 'Grp', 'Mon', etc.) for names of categories, and will also use the same font for informal variables for categories (as in 'C', 'J' etc.).

And when I want to distinguish a structure from the objects it is built from, I will typically use a script font, as in our use of '$\mathcal{G}$' vs 'G', and '$\mathcal{M}$' vs 'M'.

4.5 Some rather sparse categories

(a) So far, so very unsurprising.

However, note that monoids can get into the story in a second way. As we've seen, monoids as objects taken together with enough monoid homomorphisms as arrows can form a category. However, any single monoid taken just by itself can also be thought of giving rise to a category. Here's how:

(C3) Take any monoid $(M, *, e)$. Then define a corresponding category M whose data is as follows:

(i) M's sole object is some arbitrary entity – choose whatever you like, it *doesn't* have to be in M, and dub it '$\bullet$';

(ii) Then any object $a \varepsilon M$ counts as an M-arrow $a \colon \bullet \to \bullet$ (in other words, we put $src(a) = tar(a) = \bullet$). Composition of arrows $a \circ b$ is defined to be the monoid product $a * b$, and the identity arrow $1_\bullet$ is defined to be the monoid identity e.

It is then immediate that the category axioms are satisfied (check this!).

Note in this case, since the 'object' in the category M can be anything you like, it needn't be an object in any ordinary sense (let alone be a structure). And unless the objects of the original monoid $(M, *, e)$ happen to be functions, the arrows of the associated category M will also not be functions or morphisms or maps in any ordinary sense. So this sort of single-monoid-as-a-category won't usually be *anything* like a 'structure of structures'.

Note too that there is a sort of converse to (C3). Any one-object M category gives rises to an associated monoid built from M's arrows, with multiplication in the associated monoid being composition of arrows. Hence we can think of many-object categories as, in a sense, generalizing from the case of the one-object categories which are tantamount to monoids.

(b) Similarly, while we can put pre-ordered pluralities and the monotone maps which interrelate them together to form a category, we can also think of a single plurality equipped with a pre-order as giving us a category just by itself. Here's how:

(C4) Take any pre-ordered objects $(P, \preccurlyeq)$. Then define a corresponding category P whose data is as follows:

(i) P's objects are P again;

(ii) there is a (single) P-arrow from A to B just in case $A \preccurlyeq B$ – this arrow might as well be identified as an ordered pair $\langle A, B\rangle$ (according to some pairing scheme), which is assigned the 'source' A and 'target' B. We define composition by putting $\langle B, C\rangle \circ \langle A, B\rangle = \langle A, C\rangle$. Take the identity arrow 1_A to be $\langle A, A\rangle$.

It is trivial that, so defined, the arrows for P satisfy the identity and associatively axioms, so we do have another category here (check this!). And again, this isn't a category comprising structures and structure-respecting maps.

Conversely, if you think about it, any category with objects O and where there is at most one arrow between two objects can be regarded as some pre-ordered objects $(O, \preccurlyeq)$, where for $A, B \,\varepsilon\, O$, $A \preccurlyeq B$ just in case there is an arrow from A to B in the category. It is therefore natural to say

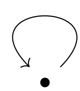
Definition 15. A *pre-order category* is a category with at most one arrow between any two objects.

Hence we can think of the unrestricted notion of a category as a generalization of the case of pre-order categories.

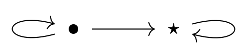
(c) Monoids-as-categories and pre-ordered-objects-as-categories can give us very small categories with few objects and/or arrows. And here are some more sparse categories.

(C5) For any objects we take, we get the *discrete category* on those by adding as few arrows as possible, i.e. just an identity arrow for each of objects we started with.

For convenience, we can allow the empty category, with zero objects and zero arrows. Otherwise, the smallest discrete category is $\mathbb{1}$ which has exactly one object and one arrow (the identity arrow on that object). Let's picture it in all its glory!



(C6) And having mentioned the one-object category $\mathbb{1}$, here's another very small category, this time with two objects, the necessary identity arrows, and one further arrow between them. We can picture it like this:



Call this category $\mathbb{2}$.

We could think this category as arising from the von Neumann ordinal 2, i.e. the set $\{\varnothing, \{\varnothing\}\}$; take the ordinal's members as objects of the category, and let there be an arrow between objects when the source is a subset of the target. Other von Neumann ordinals, finite and infinite, similarly give rise to other categories.

But hold on! Should we in fact talk about *the* category $\mathbb{1}$ (or *the* category $\mathbb{2}$, etc.)? Won't different choices of object make for different one-object categories, etc.? Well, yes and no! We can of course have, in our mathematical universe, different cases of single objects equipped with an identity arrow – *but they will be indiscernible from within category theory*. So as far as category theory is concerned, they are all 'essentially the same' – in just the same spirit as e.g. different Klein four-groups are 'essentially the same' in group theory. We will want to return to this point.

4.6 More categories

Let's continue our list of sorts of categories, first generalizing from our basic algebraic and order-theoretic examples in the last section, and then adding some geometric and other categories. And for brevity's sake, in most cases we now will jump straight to the maximal version living in our default universe of sets (i.e. the version which stands to other instances of the same general sort as e.g. Grp does to other categories of groups).

Categories of monoids and categories of groups are just the first of a family of cases, where the object-data are algebraic structures themselves comprising objects equipped with some functions and/or with certain distinguished objects picked out – and the arrows are the homomorphisms respecting the relevant amount of structure. Adding more structure to our object-data, then, we can get:

(C7) Ab is the category whose objects are (set-implementations of) abelian groups, and whose objects are (set-implementations of) group homomorphisms again.

(C8) Rng is, the category of rings, whose objects are predictably enough all rings and whose objects are ring homomorphisms (it would be boring to keep on repeating 'implementations'!)

(C9) And Bool is the category of Boolean algebras and structure-respecting maps between them.

We similarly have further categories of ordered objects. Enrich the notion of a pre-order, take as structures objects-equipped-with-the-richer-order, take enough order-respecting functions as arrows, and we get another kind of category. For example (taking maximal cases as before),

(C10) Pos is the category whose objects are all posets, sets-equipped-with-a-partial-order (where that's a pre-order which is anti-symmetric), and the arrows are all order-respecting maps again.

(C11) Tot is the category whose objects are all sets-equipped-with-a-total-order (where that's a partial order where any two objects stand in the order relation, one way round or the other). The arrows are as you would now expect.

And so on it goes!

Now for another paradigm type of case, namely geometric categories (even more central to the original development of category theory than the cases of algebraic categories or order categories).

(C12) Top is the category with

- (i) objects: all the topological spaces (in our favoured universe of course);
- (ii) arrows: the continuous maps between spaces.

(C13) Met is also a category: this has

(i) objects: metric spaces, which we can take to be a set of points S equipped with a real metric d;

(ii) arrows: the non-expansive maps, where – in an obvious notation – $f\colon (S,d) \to (T,e)$ is non-expansive iff $d(x,y) \geqslant e(f(x), f(y))$.

(C14) $\mathsf{Vect_k}$ is a category with

(i) objects: vector spaces over the field k (each such space comprising vectors equipped with vector addition and multiplication by scalars in the field k);

(ii) arrows: linear maps between the spaces.

And finally in this section, let's have a logical example.

(C15) Suppose $\mathcal{L}$ is a first-order formal language (the details don't matter). Then there is a category of propositions $\mathsf{Prop}_{\mathcal{L}}$ with

(i) objects: propositions, closed sentences $X, Y, \ldots$ of the formal language;

(ii) arrows: there is a (unique) arrow from X to Y iff $X \vDash Y$, i.e. X semantically entails Y.

The reflexivity and transitivity of semantic entailment means we get the identity and composition laws which ensure that this is a category.

4.7 The category of sets

(a) Categories like Mon and Ord whose objects are sets-equipped-with-some-structure and whose arrows are structure-respecting-functions are conventionally called *concrete* categories. As we have seen, lots of categories are not concrete in this sense – for example, neither a monoid-as-category nor a pre-ordering-as-category will count. We'll revisit the distinction between 'concrete' and 'abstract' categories in due course, and give a sharper technical account once we have the idea of a functor in play. But I thought I should mention the standard distinction straight away.

(b) Now, the monoids in Mon are sets equipped with not-very-much structure. Likewise for the pre-ordered sets in Preord. Going in one direction, we get concrete categories whose objects are sets equipped with a richer structure and whose arrows are functions constrained to respect this richer structure. Going in the other direction, we get categories of sets – i.e. categories whose objects are simply sets (equipped with no additional structure at all) and whose arrows are functions between these sets (any old functions so long as they are closed under composition, and we include the relevant identity functions: there is no requirement that functions respect structure because there is no structure to respect).

Here's the maximal case:

(C16) Set is the category with

(i) objects: all sets.
(ii) arrows: for any sets X, Y, every (total) set-function $f\colon X \to Y$ is an arrow.

There's an identity function on any set. And set-functions $f\colon A \to B$, $g\colon B \to C$ (where the source of g is the target of f) always compose. And so the axioms for being a category are evidently satisfied.

Some initial remarks:

(i) Note that the arrows in Set, like any arrows, must come with determinate targets/codomains. But we have already reminded ourselves that the standard way of treating functions set-theoretically is simply to define a function f as its *graph* $\hat{f}$, i.e. the set of pairs $\langle x, y \rangle$ such that $f(x) = y$. This definition is lop-sided in that it fixes the function's source/domain, the set of first elements in the pairs, but it doesn't determine the function's target. (For a quite trivial example, consider the Set-arrows $z\colon \mathbb{N} \to \mathbb{N}$ and $z'\colon \mathbb{N} \to \{0\}$ where both functions send every number to zero. These have the same graphs, but the functions have different targets and correspondingly different properties – e.g. the first isn't surjective, the second is.)

Perhaps set theorists themselves ought really to define a set-function $f\colon A \to B$ as a triple $\langle A, \hat{f}, B \rangle$. But anyway, that's how category theorists ought officially to regard arrows $f\colon A \to B$ in Set, and in other concrete categories too.

(ii) We should perhaps remind ourselves why there *is* an identity arrow for $\varnothing$ in Set. Vacuously, for *any* target set Y, there is exactly one set-function $f\colon \varnothing \to Y$, i.e. the one whose graph is the empty set. Hence in particular there is a function $1_\varnothing\colon \varnothing \to \varnothing$.

(iii) Note that in Set, the empty set is in fact the one and only set such that there is exactly one arrow *from* it to any other set. This gives us a simple example of how we can characterize a significant object in a category not by its internal constitution, so to speak, but by what arrows it has to and from other objects.

For another example, note that we can define singletons in Set by relying on the observation that there is exactly one arrow from any set *to* a singleton (why?).

(iv) So now choose a singleton $\{\bullet\}$, it won't matter which one (treat the bullet symbol here as a wildcard). Call your chosen singleton '1'. And consider the possible arrows (i.e. set-functions) from 1 to A.[3]

We can represent the arrow from 1 to A which sends the element of the singleton 1 to $x \in A$ as $\vec{x}\colon 1 \to A$ (the over-arrow here is simply a helpful reminder that we are notating an arrow). Then there is evidently a one-one

[3]We are overloading notation – here '1' is a special object, while in other contexts '1' is a special arrow, an identity arrow. You'll need to get used to this sort of thing, where we rely on context to disambiguate shared notations for objects and arrows.

correspondence between these arrows $\vec{x}$ and the elements $x \in A$. So talk of such arrows $\vec{x}$ is available as a category-speak surrogate for talking about elements x of A.

More on this sort of thing in due course: but it gives us another glimpse ahead of how we might trade in talk of sets-and-their-elements for categorial talk of sets-and-arrows-between-them.

(c) So far, so straightforward. But let's just pause to note again that the make-up of the category Set of course is relative to our background universe. We haven't determinately fixed that. But we'll continue to live with that for the moment. You can just interpret our talk of sets and the category Set in your preferred way assuming that this isn't too idiosyncratic!

And note that familiar size considerations that we touched on in §3.1(b) and §3.3(a) kick in again. The category of sets has all sets (in your favoured universe) as its objects. Unless you are an NF-iste,[4] however, there is no set of all sets – such a collection is, in a familiar way, 'too big' to be a set. Hence on the standard view, the category of sets is itself too big to be a set or to be modelled as a set. Not wanting to rule out the standard view of sets, that's another reason why our initial definition of a category did not say e.g. that a category always comprises a *set* of objects, but used a plural characterization.

4.8 Yet more examples

(a) Let's finish our initial list of examples of categories. And now we can go more briskly:

(C19) There is a category FinSet whose objects are the finite sets (i.e. sets with at most finitely many members), and whose arrows are the set-functions between such objects.

(C20) Pfn is the category of sets and *partial* functions. Here, the objects are all the sets again, but an arrow $f \colon A \to B$ is a function which is not necessarily everywhere defined on A (one way to think of such an arrow is as a total function $f \colon A' \to B$ where $A' \subseteq A$). Given arrows-qua-partial-functions $f \colon A \to B$, $g \colon B \to C$, their composition $g \circ f \colon A \to C$ is defined in the obvious way, though you need to check that this succeeds in making composition associative.

(C21) $\mathsf{Set}_\star$ is the category (of 'pointed sets') with

- (i) objects: all the non-empty sets, with each set A having a distinguished member $\star_A$.
- (ii) arrows: all the total functions $f \colon A \to B$ which map $\star_A$ to $\star_B$, for any non-empty sets A, B.

[4] That is to say, a devotee of Quine's deviant set theory NF which does have a universal set, and avoids paradox by constraining our comprehension principle.

As we'll show later, Pfn and $\mathsf{Set}_\star$ are in a good sense equivalent categories (challenge: pause to think why we should expect that).

(C22) The category Rel again has naked sets as objects, but this time an arrow $A \to B$ in Rel is (not a function but) any relation R between A and B. We can take this officially to be a triple $(A, \hat{R}, B)$, where $\hat{R} \subseteq A \times B$ is R's extension, the set of pairs $\langle a, b\rangle$ such that aRb.

The identity arrow on A is then the diagonal relation whose extension is $\{\langle a, a\rangle \mid a \in A\}$. And $S \circ R\colon A \to C$, the composition of arrows $R\colon A \to B$ and $S\colon B \to C$, is defined by requiring $a\, S{\circ}R\, c$ if and only if $\exists b(aRb \wedge bSc)$. It is easily checked that composition is associative.

So here we have yet another example where the arrows in a category are *not* functions.

(b) And that will surely do for the moment as an introductory list. There certainly is no shortage of categories of various kinds, then!

In fact, by this stage, you might very reasonably be wondering whether it isn't just *too* easy to be a category. If such very different sorts of structures as e.g. a particular small monoid on the one hand and the whole universe of topological spaces and their continuous maps on the other hand equally count as categories, how much mileage can there possibly be theorizing in general about categories and their interrelations?

Well, that's exactly what we hope to find out over the coming chapters.

Bibliography

Adámek, J., Herrlich, H., and Strecker, G., 2009. *Abstract and Concrete Categories: The Joy of Cats*. Mineola, New York: Dover Publications. URL http://www.tac.mta.ca/tac/reprints/articles/17/tr17.pdf. Originally published 1990.

Aluffi, P., 2009. *Algebra: Chapter 0*. Providence, Rhode Island: American Mathematical Society.

Awodey, S., 2010. *Category theory*, vol. 49 of *Oxford Logic Guides*. Oxford: Oxford University Press, 2nd edn.

Benacerraf, P., 1965. What numbers could not be. *Philosophical Review*, 74: 47–73.

Booth, D. and Ziegler, R. (eds.), 1996. *Finsler Set Theory: Platonism and Circularity*. Basel: Birkhäuser Verlag.

Church, A., 1956. *Introduction to Mathematical Logic*. Princeton, NJ: Princeton University Press.

Dummit, D. S. and Foote, R. M., 2004. *Abstract Algebra*. Hoboken, NJ: John Wiley, 3rd edn.

Finsler, P., 1926. Über die Grundlagen der Mengenlehre, I. *Mathematische Zeitschrift*, 25: 683–713. Reprinted and translated in Booth and Ziegler 1996: 103–132.

Incurvati, L., 2020. *Conceptions of Set and the Foundations of Mathematics*. Cambridge: Cambridge University Press.

Kunen, K., 1980. *Set Theory: An Introduction to Independence Proofs*. Amsterdam: Elsevier.

Leinster, T., 2014. *Basic Category Theory*. Cambridge: Cambridge University Press.

Mac Lane, S., 1997. *Categories for the Working Mathematician*. New York: Springer, 2nd edn.

Mac Lane, S. and Moerdijk, I., 1992. *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*. New York, Heidelberg, Berlin: Springer.

Maddy, P., 2017. Set-theoretic foundations. *Contemporary Mathematics*, 690: 289–322.

Oliver, A. and Smiley, T., 2016. *Plural Logic*. Oxford: Oxford University Press, 2nd edn.

Quine, W. V. O., 1963. *Set Theory and Its Logic*. Cambridge, MA: The Belknap Press of Harvard University Press.

Scott, D. S., 1980. Relating theories of the λ-calculus. In J. Hindley and J. Seldin (eds.), *To H. B. Curry, Essays on Combinatory Logic, Lambda Calculus and Formalism*, pp. 403–450. London: Academic Press.

Sellars, W., 1963. Philosophy and the scientific image of man. In *Science, Perception and Reality*. Routledge & Kegan Paul.

Tao, T., 2016. *Analysis*. Springer, 3rd edn.