Beginning Category Theory

Peter Smith

LOGIC MATTERS

This PDF is making (as yet) a very small start on revising the much-downloaded *Category Theory: A Gentle Introduction*. For the latest and most complete version of *Beginning Category Theory* and for related materials see the Category Theory page at the Logic Matters website.

Corrections, please, to ps218 at cam dot ac dot uk.

Contents

Preface

The project A few years ago, I put together a lengthy set of notes, *Category Theory: A Gentle Introduction*. I didn't really intend to write that near-book. But background notes for a different project grew and grew, and they started taking on a life of their own as I tried to organize them more logically. The emerging result was an elementary introduction to some entry-level category theory, a beginner's guide of the kind that I myself would have rather welcomed when starting out, and which I hoped that others might find both helpful and intriguing. At least in a rough and ready way, I tried to cover most of the really basic notions of category theory – explaining the very idea of a category, then treating e.g. limits, functors, natural transformations, representables, and adjunctions, with some pointers forward to further ideas. The aim was to get far enough to give a reader a reasonable grounding from which to tackle other texts of various kinds with some confidence.

Those notes were indeed rather long given their limited coverage, because they went at a pretty gentle pace. I don't apologize at all for this: there are plenty of fast-track alternative introductions available. But experience strongly suggests that getting a secure understanding of categorial[1] ways of thinking by initially taking things slowly does make later adventures exploring beyond the basics *very* much more manageable.

The last version of the notes as published online was still very much a work in progress; so there were chapters at different levels of development and with different degrees of integration with what's around them (and no doubt different levels of reliability). Despite their half-baked character, the notes have been downloaded well over a thousand times a month. Which is pleasing but also embarrassing, as I know how flawed they are.

I needed to set the *Gentle Introduction* aside for a few years, longer than I planned, to get on with some other projects. But those are at last finished: I now have time to return to thinking a bit about categories, and can start working on a new and (let's hope!) improved version of the notes. To avoid confusion, I'll for the moment give this a different title, hence *Beginning Category Theory*.

Here then are some initial chapters. All comments and corrections will be very gratefully received!

[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.

Who are these notes for? What do you need to bring to the party? I imagine one reader to be a mathematics student who wants a first introduction to some categorial 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 a relatively accessible introduction to give them an initial sense of what the categorial fuss is about, so that they can tell if they want to find out more.

Now, you obviously can't be well placed to appreciate how category theory gives us a story about the ways in which different parts of modern abstract mathematics hang together if you really know *nothing* beforehand about modern mathematics! But I try to presuppose relatively little. Suppose you know a few basic facts about groups, know something 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. 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.

Theorems as exercises There are currently no exercises in what follows – or at least, there are none explicitly labeled as such. However, almost all the proofs of the theorems you meet as you begin category theory are *very* straightforward. Surprisingly often, you just have to 'do the obvious thing': there's little ingenious trickery needed at the outset. So you can usually 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 at least, *an* answer) to the exercise. For a few tougher theorems, I give preliminary hints about how the argument ought to go.

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 that needs no proof), I also use '△' as an end-of-definition marker.

Thanks! Andrew Bacon, Malcolm F. Lowe and Mariusz Stopa very kindly sent embarrassingly long lists of corrections to the ancestor of these notes. A lot of the mistakes were obvious typos, but there were also enough mislabelled arrows or fumbling of notation mid-proof and the like that I should certainly apologize to readers who found themselves scratching their heads in puzzlement. I had further corrections from David Ozonoff, Zoltán Tóth, and Adrian Yee. 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 the structure-preserving maps between them. Then we can think of this family as forming a further structure – a structure-of-structures, if you like – something else to explore mathematically.

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 axioms. But we can also think of a whole family of groups, together with appropriate maps between them – i.e. homomorphisms which preserve group structure – as forming a further structure-of-structures.

Another example: any particular *topological space* is a structure, this time 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 preserve topological structure – forms another structure-of-structures.

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-preserving maps between them. We are interested in turn 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 objects), 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-preserving maps, the so-called *functors*, this time linking structures-of-structures (as when we map 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 then is *one* central mathematical imperative: 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.

And this is where category theory comes into play for us: it provides exactly what we need, at least 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, together with some of its connections with two other disciplines with different kinds of generality, logic and set theory.

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 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 note, category theory does much more than give us a helpful way of relating 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)

From its highly abstract vantage point, category theory crucially reveals *new* connections we hadn't made before. What are called 'adjunctions' are a prime example, as we will eventually see.

Making new connections 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 this book. I will stick to very much more elementary concerns, with an emphasis on unification and conceptual clarification. This will still give us more than enough to explore. And this way, I 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 intricate interconnections mean, however, 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 then 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 we said, many paradigm cases of categories are indeed structured-families-of-structures. And we go on to develop ways of describing what happens inside a category. In this new setting, we revisit e.g. many familiar ideas about maps between structures, and about ways of forming new structures by e.g. taking products or taking quotients.

Only after some extended exploration of categories taken singly do we then move up another level to consider *functors*, maps between categories (typically, maps between families of structures).

And then, 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 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*.

In short, then, my chosen route here into the basics of category theory steadily ascends through the increasing levels of abstraction in a particularly natural way (which has some logical appeal).

True, this does mean that it takes us quite a while to reach some of the *really* novel and exciting high-level categorial ideas.

However, this disadvantage is (I hope) considerably outweighed by the gain in secure understanding which comes from taking our gently sloping path. I will just have to do my best to make the views we glimpse along the way still seem interesting enough.

2 One structured family of structures

Category theory gives us a framework in which we can think systematically about structured families of mathematical structures: or at least, it is this aspect of the theory which is going to be our focus.

I said that one paradigm case of such a structured family comprises some groups organized by homomorphisms between them. So let's begin by reviewing some *very* elementary facts about groups. These facts will probably all be very familiar; but it will be useful to pick out some themes which are already there in pre-categorial mathematics, themes which we will later encounter again in a categorial guise.

You'll spot straight away that, in one respect, the definitions I give in this chapter are not quite the usual ones. But I'll explain the reason for their (only mildly) deviant character in Chapter 3, so do indulge me!

2.1 Groups revisited

(a) Here then is my preferred way of characterizing groups for now:

Definition 1. The objects G (including the distinguished object e), equipped with a binary operation $*$ (where for any x, y among G, $x * y$ is also among G), form a *group* iff

(i) $*$ is associative, i.e. for any x, y, z among G, $(x * y) * z = x * (y * z)$;
(ii) e acts as a group identity, i.e. for any x among G, $x * e = x = e * x$;
(iii) every object has a group inverse, i.e. for any x among G, there is at least one object y also among 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 them.

If e and e' are both identities for the group of objects G equipped with $*$, then $e = e * e' = e'$; so group identities are unique. Likewise, inverses are unique.

(b) I should emphasize the variety of objects and operations that can form a group. In fact, *any* object e, whatever you like, equipped with the only possible binary operation $*$ such that $e*e = e$, forms a trivial one-object group. Similarly, any two objects e, j, whatever you like, form a group when 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. non-zero rationals 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 indeed very many and various! But you knew that.

(c) We need to agree some notation. Let's use '$(G, *, e)$' simply to abbreviate 'the objects G equipped with the operation $*$ and with distinguished object e'. 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 $(G, *, e)$'.

As we have seen, 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 in general to using multiplication-like notation and to talk generically of group 'products'; we will correspondingly default to denoting the inverse of a group object x by x^{-1}. Additive notation, however, is commonly used when dealing with abelian groups in particular.

2.2 A quick word about 'objects'

Let me pause to make a point which we'll encounter again when categories are introduced.

Many philosophers and logicians argue – rightly, in my view – for an *absolute* type-theoretic distinction between objects-as-individuals and first-level functions sending objects to objects, and then second-level functions sending first-level functions to first-level functions, and so on up a type hierarchy (which also allows more complex types, with functions whose inputs and outputs are at mixed levels).

So 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 operation 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 indeed be objects-as-individuals (like numbers); but as already noted, the items can equally well be first-level functions (like permutations, i.e. bijections between numbers); or they can even be functions-of-functions, or of a still higher type.

2.3 New groups from old

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

Definition 2. $(G', *, e)$ is a *subgroup* of $(G, *, e)$ iff (i) G' are some of the objects G, and (ii) these objects 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, under addition, form a subgroup of the additive group of integers. For another example, the complex numbers on the unit circle, under multiplication, form a subgroup of the multiplicative group of non-zero complex numbers.

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

Definition 3. A scheme for pairing objects G with objects G' provides

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

(ii) a binary pairing function which we can notate '$\langle\ ,\ \rangle$' which sends x from among G and x' from among G' to a pair-object $\langle x, x'\rangle$ among O (with every pair-object being some such $\langle x, x'\rangle$);

(iii) two (unary) 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 the 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 group objects G and G' are in both cases natural numbers, then we could perfectly well take the pair-object $\langle m, n\rangle$ to be the number 2^m3^n, with the obvious pairing and unpairing functions.

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 an object x from G with an object x' from G' to a pair-object $\langle x, x'\rangle$. Let H be all the relevant pair-coding objects. Define $d = \langle e, e'\rangle$, and put $\langle x, x'\rangle \star \langle y, y'\rangle = \langle x * y, x' *' y'\rangle$. Then $(H, \star, d)$ is a *product* of the groups $\mathcal{G}$ and $\mathcal{G}'$. △

It is routine to check that $(H, \star, d)$ is indeed 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 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 (the entry at row r, column s, gives $r \star s$):

$\star$	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, then, 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 the 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 in an obvious sense: each way of forming a product group from a two-object group and itself always give us a group describable by reinterpreting the same table. And the point generalizes. Products of $\mathcal{G}$ with $\mathcal{G}'$ 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. If G are some objects, and $\sim$ is an equivalence relation defined over G, then a corresponding quotient scheme provides

(i) some quotient-objects O (which can be any suitable objects, which may or may not be already among G),

(ii) a unary function which we can notate '[ ]' which sends x from among G to a quotient-object $[x]$ among O (with every quotient-object being some such $[x]$), where

(iii) for all x, y among G, $[x] = [y]$ iff $x \sim y$. △

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

But note, but not that just as pair-objects in pairing schemes do not have to be sets, we similarly do *not* require $[x]$ to *be* a 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 $x \equiv_8 y$.

With our new definition to hand, we can now define the notion of a quotient group in two steps:

Definition 6. (i) If $\mathcal{G}$, i.e. $(G, *, e)$, is a group, and $\sim$ is an equivalence relation defined over the objects of $\mathcal{G}$, then $\sim$ *respects the group structure* of $\mathcal{G}$ iff, for any objects x, y, z from G, given $x \sim y$, then $x * z \sim y * z$ and $z * x \sim z * x$ (in other words, 'multiplying' equivalent objects by the same object yields equivalent results).

(ii) Suppose $\mathcal{G}$, i.e. $(G, *, e)$, is a group, and $\sim$ is an equivalence relation which respects its group structure, and suppose we also have a quotient scheme for $\sim$, which sends x among G to $[x]$. Let $[G]$ be all the objects $[x]$ for x among G, and define a binary operation $\star$ on the objects $[G]$ by putting $[x] \star [y] = [x * y]$. Then $([G], \star, [e])$ is a *quotient* of the original group $\mathcal{G}$ with respect to $\sim$, which we symbolize $\mathcal{G}/{\sim}$. △

Again, it is relatively routine to check that, as defined, this quotient $\mathcal{G}/{\sim}$ *is* indeed a group.

We just need to show that $\star$ is well-defined, i.e. the result of $\star$-multiplication does not depend on how we pick out the multiplicands. In other words, we need to show that if $[x] = [x']$ then (i) $[x] \star [y] = [x'] \star [y]$, and (ii) $[y] \star [x] = [y] \star [x']$. For (i), note that if $[x] = [x']$, then by definition $x \sim x'$, hence (since $\sim$ respects group structure) $x * y \sim x' * y$, hence by definition again $[x] \star [y] = [x'] \star [y]$. We derive (ii) similarly. It remains to check that $\mathcal{G}/{\sim}$ is then a group with $\star$ the group operation. But that's straightforward.

Note, 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 objects $[G]$ 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').

Let's take a quick example, to reinforce the point that the objects forming a quotient group need not be equivalence classes. Suppose $(Z, +, 0)$ is the additive group of the integers, $\mathcal{Z}$ for short, and consider again the equivalence relation of congruence mod 8 defined over Z. This equivalence relation evidently 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 together $\overline{8}$. Then $(\overline{8}, +_8, 0)$ is evidently a group (where $+_8$ is addition mod 8), and it is a quotient $\mathcal{Z}/{\equiv_8}$.

(d) Given groups $\mathcal{G}$ and $\mathcal{G}'$, do they always have a product? Given a group $\mathcal{G}$ and an equivalence relation $\sim$ which respects its group structure, is there always a quotient group $\mathcal{G}/{\sim}$?

On the one hand, nothing we have said so far assumes positive answers. On the other hand, it is very natural to work on the very modest assumption that such constructions are freely available, i.e. product schemes and quotient schemes are always available when we want them. Elementary discussions of group theory typically proceed as if that modest assumption is permitted.

2.4 Group homomorphisms

(a) Next, let's equally briskly recall some basic facts about structure-preserving maps between the groups. For now, my preferred – though again slightly deviant – style of definition is:

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 the objects G with values among H such that:

(i) for every x, y among G, $f(x * y) = fx \star fy$,

(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. It is immediate that a homomorphism also sends inverses to inverses: i.e. $f(x^{-1}) = (fx)^{-1}$.

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-preserving homomorphism we can use the notation $f\colon (G, *, e) \to (H, \star, d)$, or $f : \mathcal{G} \to \mathcal{H}$.

(b) An initial pair of trivial examples:

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

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

These cases remind us that, although homomorphisms are often described as 'preserving' group structure, this has to be understood with a large pinch of salt. Homomorphisms can in fact suppress many or most aspects of structure simply by mapping distinct objects to one and same value.

Three more elementary examples:

(3) There is a homomorphism from $\mathcal{Z}$, the additive group of integers $(Z, +, 0)$, 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.

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

(5) Let R be the real numbers, and 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 $j\colon (R, +, 0) \to (C^*, \times, 1)$ 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 an 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. For (1) we, of course, simply take $g \circ f$ ('g following f') applied to an object x among the objects of $\mathcal{G}$ to be $g(f(x))$ and then check that $g \circ f$ so defined satisfies 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) A quick but important remark. Note that this, our very first theorem, is *not* a mere logical consequence of our definitions of groups and group homomorphisms. Our proof depends on invoking background assumptions about functions, such as 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:

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 objects of $\mathcal{G}$ and the objects of $\mathcal{H}$.

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 the group comprising e, j, equipped with the only possible group operation $*$, and the group comprising e', j', equipped with $*'$. Then the map which sends the group identity e to the group identity e' and j to j' is obviously a group isomorphism.

(2) There are two automorphisms from the additive group $\mathcal{Z}$ 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 $\mathcal{Q}$ 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 the numbers $1, 3, 5, 7$ equipped with multiplication mod 8. 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, when suitably reinterpreted.

(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 (we 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 of $\mathcal{G}$ with $\mathcal{G}'$; 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 from $\mathcal{G}$ and x' from $\mathcal{G}'$ 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 suitable equivalence 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 by definition the underlying function $f\colon G \to H$ is a bijection and so 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, and g is its two-sided inverse, 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 also an isomorphism. And given that isomorphisms are homomorphisms with two-sided inverses which are homomorphisms, it is immediate that the inverse of an isomorphism is also an isomorphism. Therefore, just 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 just two automorphisms from $\mathcal{Z}$ 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 by a non-zero rational' is an automorphism for the additive group $\mathcal{Q}$, and stretchings 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 looked at structure-preserving maps between groups. We now bring the two themes together.

(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 $\mathcal{H}$, there is a homomorphism $f\colon \mathcal{G} \to \mathcal{H}$ such that that 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 all the objects which are f-images of objects from among G, so they include d, i.e. $f(e)$. Define $f(\mathcal{G})$, the f-image of the group $\mathcal{G}$, in the obvious way as $(f[G], \star, d)$. We now need to check this too is a group, and hence a subgroup of $\mathcal{H}$.

(i) Suppose y_1 and y_2 are among $f[G]$. By assumption, they are f-images of some objects x_1, x_2 among G. So we have $y_1 \star y_2 = fx_1 \star fx_2 = f(x_1 * x_2)$, and hence $y_1 \star y_2$ will also be among $f[G]$ as required.

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

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 from $\mathcal{G}$ 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 skip past product groups for now, and next consider quotient groups arising from suitable equivalence relations. We then have the following result:

Theorem 6. *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 $f(\mathcal{G})$, 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 an equivalence 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 we can indeed define $\mathcal{G}/{\sim}$. 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 $\sim$, and it is then 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 certain claims about the structure of certain groups, this time about their quotient structure, for corresponding claims about homomorphisms between groups.

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

(c) Similarly, it turns out that claims about the structure of product groups can also be traded in for claims about corresponding homomorphisms between groups (we use the fact that pairing schemes essentially involve pairing and unpairing functions that behave in the right way). But I'll leave the proof of this for later. For now, I'll just flag up the 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 of central importance for category theory.

2.8 'Identical up to isomorphism'

(a) 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*. In other words, despite the differences between their objects, the groups are the same at least as far as their structural properties are concerned – 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 very happy way to putting things. We have just reminded ourselves that $\mathcal{K}_2$ and $\mathcal{K}_3$ are isomorphic groups. But $\mathcal{K}_2$ 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', in the sense of being summed up by the same table differently interpreted.

A rather happier, and widely used, way of putting things is this: $\mathcal{K}_2$ and $\mathcal{K}_3$ are identical *up to isomorphism*. And (now reading that quotation more charitably) for many purposes, group theory can indeed ignore the differences between groups which are identical up to isomorphism.

(b) It is common, then, to talk of *the* Klein group $\mathcal{K}$, *the* permutation group of three elements $\mathcal{S}_3$, *the* free group over the generators G, and so on. And in most contexts, we can let this pass quite happily. Though if we are being pernickety, such talk can be cashed out in one of two or perhaps three ways:

1. Most simply, talk of *the* Klein group can typically be treated as just generalizing talk about Klein groups. So 'the Klein group is abelian' is to be understood as simply saying that any Klein group is abelian. Similarly, 'There is a unique homomorphism from the Klein group to the one object group' says that for any Klein group and any one object group, there is a unique homomorphism from the first to the second. And so on.

2. Sometimes, though, a specific paradigm case, a canonical exemplar, is introduced. For example, *the* free group with certain generators is often defined by a very particular construction using 'words' formed from the generators.

3. Occasionally, and more mysteriously, some flirt with the idea that, as well as concrete Klein groups (to return to that example), i.e. 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. It is this purely abstract group comprising de-natured elements which is then said to be, properly speaking, *the* Klein group.

For now, we'll hang fire on the question whether the third option makes much sense. We will in due course need to revisit this sort of question.

2.9 Categories of groups

(a) We said at the beginning of the first chapter that 'suitably structured families of structures' are prime examples of categories. Then at the start of this chapter, we picked out a paradigm case, namely a family of groups organized by homomorphisms between them.

We just impose two very natural conditions on such a *category of groups*. First, its homomorphisms are closed under composition. And second, the identity homomorphisms for each relevant group are included. It's that simple!

(b) But now let's say exactly the same thing again, but this time in laborious detail, for clarity's sake. So:

Definition 9. A *category of groups* comprises

1. some groups *Grp*, and
2. some group homomorphisms *Hom*,

where these groups and homomorphisms are governed by the following conditions:

Sources and targets For each homomorphism $f\colon \mathcal{G} \to \mathcal{H}$ among *Hom*, both its source group $\mathcal{G}$ and its target group $\mathcal{H}$ are among *Grp*.

Composition For any two homomorphisms $f\colon \mathcal{G} \to \mathcal{H}$, $g\colon \mathcal{H} \to \mathcal{I}$ among *Hom*, where the target of f is the source of g, the homomorphism $g \circ f\colon \mathcal{G} \to \mathcal{I}$ is also among *Hom*.

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

The homomorphisms also satisfy the following conditions:

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

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

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

(c) 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 boring cases! Things can get more interesting when the groups in a category start to communicate (so to speak).

Consider next, then, the category which comprises all the finite groups whose objects are natural numbers together with all the isomorphisms between them. 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; and we have enough homomorphisms in play to enable us to begin to use them 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.

But first, we do really need to pause to say something about the universe where at least initially we will take categories of groups to live (along with all the other categories we will meet).

3 Groups and sets

The last chapter gave some reminders about groups and their homomorphisms, enough for our present purposes. And everything we said, at least before we mentioned categories, will probably be very familiar. Except in one respect. I almost entirely avoided talking about *sets*.

This is the mildly deviant feature of the presentation. There was a reason – a good reason, as I hope to make clear – for introducing groups in my way; but there is a good reason for now bringing sets back to centre stage. This chapter explains.

3.1 Sets, virtual classes, plurals

(a) We need to begin by getting an important distinction into focus.

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

But if this is the guiding idea, then the first point to note is that *a great deal of elementary informal set talk is really no more than a façon de parler*. Yes, it is a useful and familiar idiom for talking about many things at once; but in many elementary contexts informal talk of a set doesn't really carry any serious commitment to there being any *additional* object over and above those many things. On the contrary, apparent singular talk about *the set of Xs* can often be paraphrased away into talk directly about those Xs, without loss of content. Talk about the set of prime numbers, for example, can typically just be taken as a way of talking about the prime numbers themselves.

When it can be paraphrased away, talk of sets is said to be talk of *virtual classes*.[1] And the idea that there is a distinction to be made between sets and virtual classes is an old one. Here is Paul Finsler, writing a century ago:

> It would surely 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

[1] See W.V.O. Quine's famous discussion in the opening chapter of his *Set Theory and its Logic*: "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)

> 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.[2]

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 choice of particular word here, 'class' vs 'set'. What matters is the distinction between non-committal, eliminable, talk – talk of merely virtual sets/classes/pluralities (whatever we call them) – and uneliminable talk of sets as entities in their own right.

(b) I'm not so sure that Finsler is right, though, about the inconvenience of plural talk. Indeed, for clarity's sake, I do think the best policy is simply to eschew class talk when we can, and to stick to plural locutions when there's no particular need to invoke sets in the Cantorian sense. So that will be the general policy in this book.

Do note that there is nothing suspect or unnatural about the use of plural terms. Consider, for example, terms such as 'the real numbers between 0 and 1', 'the points where line L intersects curve C', 'the finite groups of order 8', 'Hilbert's axioms for geometry', 'the symmetries of a rectangle', 'the ordinals', etc. Mathematicians of course are always using 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 Xs as disguised singular talk referring to *the set of* X's (where the set is a single item and something distinct, over and above its members). For you already know that *that* can't always be done. We can't, for example, trade in universally generalizing plural talk about the ordinals for singular talk about the set of ordinals because, on standard assumptions, 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).

The same goes, as we will see, when it comes to defining categories (and this matters). We can't in general treat a category as comprising a *set* of items together with the maps between them. For there may be too many relevant items to form a set. We will meet an example at the end of the chapter.

(c) In sum, plural talk – for example, of the kind I used in talking about groups in the previous chapter – is in perfectly good logical order as it is, taken at face value, without needing to be re-interpreted as referring to sets.[3]

And, crucially, we will need plural talk – and preferably not disguised as talk of (virtual) classes – in framing our theory of categories.

[2] Finsler 1926, p. 106, quoted in Incurvati 2020, p. 3.

[3] That is still true, even if your measure of being in good logical order is formalizability: 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's *Plural Logic* (2016).

3.2 Group theory again

(a) Now let's return to thinking in general terms about what it takes to develop group theory, initially as informal mathematics in the standard sort of way.

As noted at the time, even as soon as we reach our elementary Theorem 1 we are going beyond the mere logical consequences of our definitions of groups and group homomorphisms. So what do we need to bring to the table to get group theory going? Roughly: the usual mathematical stock-in-trade of a body of facts about *functions* together with a repertoire of available *constructions*.

For example, we generously assume that any association of inputs to single outputs (nicely specifiable or not) constitutes a function; so functions always do compose when they can, and composition is associative. We assume that there can exist functions of any possible type we want (i.e. functions of objects, functions of functions, functions of functions-of-functions, and functions of mixed types too). We assume facts about injections, surjections, bijections. We assume that it makes sense, e.g. to talk about all the bijections between some objects and those same objects. And so on, and so forth.

Similarly, we assume that we can construct (what will serve as) pairs ad libitum; and we assume that when an equivalence relation partitions some objects we can somehow get representatives for the partitions. In other words, in our earlier terms, we assume pairing schemes and quotient schemes are available whenever we want them. And we not only can construct pairs and finite tuples but infinite sequences too. We also assume that we can freely construct 'copies' of whatever structures we already have. And so on, and so forth.

That's intentionally vague. But we are just reflecting how standard developments of group theory just help themselves at the outset to whatever unproblematic informal background assumptions are useful as they go along.

(b) The everyday informal mathematics of functions and constructions is not intrinsically set-theoretic (not even when it deals with pairs etc., a point we will return to). And we can informally develop group theory quite extensively without talking of sets or making essentially set-theoretic assumptions; part of the point of the previous chapter was to show how to make a start on this.

Suppose, however, that we do adopt a more conventional idiom from the outset; so instead of saying that a group $(G, *, e)$ comprises some objects G equipped with a suitable binary operation $*$ and a distinguished object e, we define a group as comprising a non-empty *set* of objects, etc. The set idiom need *still* be doing no heavy lifting at all (as we continue to work in informal mathematics). Paulo Aluffi, for example, frankly remarks on this at the beginning of his fine book *Algebra, Chapter 0*: the conventional informal set idiom which he adopts is, he says, 'little more than a system of notation and terminology" (Aluffi 2009, p. 1).

(c) So far, so unexciting. But there remains, of course, a decidedly interesting project in the offing, the project of regimenting the background stock-in-trade of assumptions about functions and constructions which we bring to bear in developing group theory (and other theories) by the ordinary standards of informal mathematics. And *here* by far the best-known way to go is set-theoretic: rather remarkably, a universe of sets, in no-doubt familiar ways, in fact provides proxies or surrogates for all the functions and constructions we can want (and our set theory also provides comforting confirmation that our usual informal stock-in-trade can be made to cohere consistently together). But still, it is important to highlight the distinction between the gadgetry we need in developing group theory, for example, and the set-theoretic underpinnings we might offer for that gadgetry. For that leaves room for the thought that other underpinnings might be available.

For one example, a suitable *topos* (that's a distinctively category-theoretic notion) can arguably provide an alternative, even improved, framework in which we can regiment our ordinary mathematical gadgetry. And that claim would be very puzzling if we have already jumped too quickly to assuming that ordinary mathematics is already fixedly set-theoretic through and through.

3.3 Set-theoretic surrogates

(a) I've just talked cautiously of a universe of sets as providing *surrogates* for the functions and constructions of informal mathematics. Why the caution?

Consider, for a first simple example, the case of one-place functions. Agree on some standard way of implementing ordered pairs as sets, e.g. as Kuratowski pairs $\langle x, y\rangle_K$ ($=_{\text{def}}$ $\{\{x\}, \{x, y\}\}$). Then here is a familiar (and uncontentious!) definition:

Definition 10. Given a function $f\colon X \to Y$, the *graph* of f is the set Γ_f of ordered pairs $\langle x, y\rangle_K$ where x is among the objects X and y is among Y, and $fx = y$. △

An orthodox policy, at least in its baldest form, *identifies* a function $f\colon X \to Y$ with its graph Γ_f.[4]

However, we really should resist this contentious 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 applies to too many things

[4]Some would prefer, for rather good reasons which chime with category theory, to identify a function with a set-theoretic *triple* whose members are the function's graph, its domain (treated as a set, rather than plurally, of course), and its co-domain. But our worries about the simpler version of the orthodoxy will carry over, mutatis mutandis, to the fancier version, so we need not delay just now over this complication.

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.

(b) Well, next note that the implementation of a function as a set of ordered pairs involves arbitrary choices of implementation scheme.

(i) For a start, it is arbitrary to fix on Kuratowski's particular implementation of pairs as sets.
(ii) And even relative to a choice of set-theoretic 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 $\langle x, y\rangle$ – indeed some textbooks do just this. And 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 conventionally defined – we have made the uniquely *right* choice, i.e. the choice that correctly identifies which set that function really is. But if there is no determinate fact of the matter about which sets functions are, then *functions aren't sets*.

(c) We can press the point further: *functions just aren't the right logical type of thing to be sets*. As Alonzo Church puts it:

> 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)

And just to show that it isn't just logicians who care about this, here is Terence Tao on the same theme:

> 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)

For example, a function such as the factorial defined over the natural numbers is, of its nature, the type of thing which yields a numerical value when given a 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 to e.g. sets of ordered pairs of numbers in particular.[5]

Which 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

[5] A well-known Fregean metaphor might help. Functions of their nature are 'unsaturated' and have an *arity*, 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', it is self-standing, with no empty slots waiting to be filled.

regress, we had better not seek to again trade in this evaluation function for another set.

(d) We can make parallel remarks about another orthodoxy here. Recall another (uncontentious) definition:

Definition 11. Given a binary relation R which holds between objects X and Y, the *extension* of R is the set E_R of ordered pairs $\langle x, y\rangle_K$ where x is among X, y is among Y, and xRy. △

Then the familiar (but contentious) orthodoxy *identifies* the relation R with its extension E_R.

But again we should definitely resist. For a start, some relations are 'too big' to have extensions according to standard set theories (consider e.g. the relation that holds between a singleton and its sole member). And in any case the arbitrariness built into the conventional rendition of ordered pairs prevents us from justifiably saying that a relation really *is* its extension as defined. And if there is no determinate fact of the matter about which sets relations are, then *relations aren't sets.*

And again, we can press the point further: *relations just aren't the right logical type of thing to be sets.* Start this time with the following observation:

> it lies in the nature of a relation that it holds or does not hold of things. (Oliver and Smiley 2016, p. 156)

But a set doesn't, of its nature, hold true of anything; and that applies to any set, including sets of ordered pairs. Hence an extension isn't the type of thing that a binary relation is.[6]

(e) Since functions aren't sets, the best we can do in a set-theoretic environment is to provide graphs as *surrogates* (or *proxies* or *implementations*) for functions. Then we can render a claim of the form $f(x) = y$ by a corresponding set-theoretic claim of the form $\langle x, y\rangle_K \in \Gamma_f$.

Similarly, since relations aren't sets, the best we can do in a set-theoretic environment is to provide extensions as *surrogates* for relations (all bar one), and so render a claim e.g. of the form xRy by a corresponding set-theoretic claim of the form $\langle x, y\rangle_K \in E_R$. In set theory, the relation $\in$ is left as the sole relation recognized as such.

Back then to structures like groups or well-ordered objects. These comprise some objects and functions, or some objects and relations on those objects, and so on. But there are strictly speaking no functions or relations (other than membership) in a universe of sets – we have to make do with proxies for them. So we get only surrogates for groups or surrogates for well-orderings a universe of sets. Similarly for other structures.

[6] The same Fregean metaphor might help. Relations of their nature are 'unsaturated' and have an arity, a certain number of empty slots waiting to be filled appropriately when the relation is applied to the right number of things. By contrast, as noted before an object like a set is already 'saturated', it is self-standing, with no empty slots waiting to be filled.

Which is all just fine, of course. Our beef isn't with the metaphysics of functions etc., and proxies for functions and the rest will serve various mathematical purposes perfectly well. So three cheers for set theory, in its place! But still, it *is* important for our future purposes to be clear about the role that the universe of sets can play here. It provides *one* generous arena where we can find proxies for all the structures we want.[7] However – and here is the important point again – this does leave the door open to the possibility that other kinds of universe might do the same job. They might even do the job better in some respects, e.g. by more faithfully respecting the type-differences between objects-as-individuals, functions, and relations.

3.4 'The' category of groups?

(a) Let's return to categories of groups.

The finite groups whose objects are natural numbers are countable, and so are the homomorphisms between these groups. Hence the category we defined as comprising them is equally a tamely countable structure-of-structures. But there are much larger, more inclusive, categories of groups. Indeed, we might wonder: is there perhaps an all-inclusive category of *all* groups and *all* the homomorphisms between them?

"But can this really make sense? For a start, can we stably pin down *all* the groups? To take a silly example, if I cut out a new non-equilateral rectangle, then – lo and behold! – won't there spring into being a new Klein group, the group of its own rotation/reflection symmetries?" Fair questions, given that we were previously entirely permissive about where we can find groups: on our definition, we just need some new objects (in the broad sense) and a suitable operation on them and we get another group. But on the other hand, a new physically realized Klein group is surely neither here nor there as far as group theory is concerned. As we said before, group theory will for most purposes ignore the differences between groups which are identical up to isomorphism; it can concentrate on more abstract exemplars.

Well, suppose we can assume that we are working in a capacious enough mathematical universe which has copies of all the groups we will ever want (so we won't care about any additional isomorphic copies of these groups which are roaming outside in the wild). Then *that* universe can be the arena in which we can hope to locate a determinate category of 'all' groups and their homomorphisms And where can we find a capacious enough mathematical universe? We have already trailed the now entirely predictable answer: Take a large enough universe of sets. Then we will find surrogates for all the groups we want there; and – now happily working with surrogates – there will consequently be a category Grp living there which comprises all these groups and the homomorphisms between them.

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

(b) But we now have something of a presentational quandary.

In developing category theory – our abstract theory for handling structures of structures – there is (as we have been stressing in previous sections) a good reason for *not* baldly assuming that mathematical structures must all have their home in some universe of sets. So, on the one hand, we'd perhaps rather like to avoid taking on any specifically set-theoretic commitments at the outset.

On the other hand, we'll soon want to talk about large categories like Grp which supposedly comprise (at least proxies for) 'all' groups and the homomorphisms between. But as we've just said, if that's to make sense, it seems that we will need to think of the relevant groups as living in some definite but sufficiently capacious universe. And a universe of sets is the only familiar candidate we are likely to have available. So the obvious default is indeed to understand categories like Grp as living in a universe of sets.

Now, we could lean in the first direction and indeed try to proceed at some remove from any direct engagement with sets. But this would in fact involve too big a departure from the standard mode of presentation of elementary category theory, which would be quite inappropriate for this kind of introduction. There's nothing for it, then, but to start off by going in the second direction. So we now *will* be assuming (and for quite a long time) that we are working in a universe of sets which is capacious enough to contain at least proxies for all the structures that we want. And then we can indeed sensibly talk about categories like Grp.

(c) But do note that what Grp comprises will be relative to our favoured universe of sets, and that's not a unique choice.

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 a model to work in? Even if we go second-order, that still doesn't determine isn't a unique universe – there are different models of second-order ZFC with different heights. But then, on second thoughts, do we need such rich universes as models of full ZFC for ordinary mathematical purposes? It has been argued e.g. that the much weaker Mac Lane set theory is strong enough to model standard mathematics which is not directly connected with set theory or logic. And arguably more radically deviant set theories like NFU provide equally competent generous arenas for modelling the gadgetry of ordinary mathematics.

Now, we plainly don't want to get bogged down into further investigations of such contentious issues here at the very beginning. What to do? In fact, it is quite common for introductions to category theory to talk loosely about large inclusive categories like Grp as if they are determinate mathematical structures, without fussing too much about the relativity to our choice of a background universe of sets. We will do the same. Because, as we will see, this relativity will turn out to be harmless for most purposes. It will be a while, however, before we can show exactly why. And it will be a while longer before we can perhaps begin to untie our developing category theory from its initial anchorage in a world of sets. There is a lot of ground to cover first.

Bibliography

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

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.

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

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.

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.