Beginning Mathematical Logic

A Study Guide: Part I

Peter Smith

LOGIC MATTERS

First published in this format, Logic Matters 2021

The latest version of this Guide is always available at logicmatters.net/tyl

Contents

Preface

Before I retired from the University of Cambridge, it was my greatest good fortune to have secure, decently paid, university posts for forty years in leisurely times, with almost total freedom to follow my interests wherever they meandered. Like most of my contemporaries, for much of that time I didn't really appreciate how *very* lucky I was. This Study Guide to mathematical logic textbooks is intended to give a little back by way of heartfelt thanks.

The Guide is aimed at two main groups of readers – philosophers who want to go on beyond their first introductory course to learn some more serious logic, and mathematicians wanting to get to grips with an under-taught but exciting area of mathematics. Why not two separate Guides? Because it would be difficult to decide what should go where. After all, a number of philosophers develop interests in more purely mathematical corners of the broad field of logic. And a number of mathematicians find themselves becoming interested in more foundational/conceptual issues. Rather than impose artificial divisions, I provide here a single but wide-ranging menu for everyone to select from as their interests dictate. So . . .

Don't be scared off by the Guide's length! This is due both to its breadth of coverage and also to its starting just half a step beyond what is sometimes rudely called 'baby logic' and then going quite a long way towards upper graduate-level treatments of a variety of topics. Simply choose the parts which are most relevant to your background and your interests, and you will be able to cut the Guide down to more manageable proportions. There is a lot of signposting and there are also explanatory overviews to enable you to pick your way through.

However, if you are hoping for help with *very* elementary logic (perhaps as encountered by philosophers in their first formal logic courses), then let me say straight away that this guide is not designed for you. The only chapter that even briefly mentions logic at this level is the initial introduction for philosophers. All the rest is about rather more advanced – and eventually *very* much more advanced – material.

Most of the recommendations in this Guide point to published *books*. True, there are quite a lot of on-line lecture-notes that university teachers have made available. Some of these are excellent. But they do tend to be terse, and often *very* terse (as entirely befits material originally intended to support a lecture course).

They are usually not as helpful as fully-worked-out book-length treatments for students needing to teach themselves.

So where can you find the books mentioned here? They should in fact be held by any large-enough university library which has been trying over the years to maintain core collections in mathematics and philosophy (and if the local library is too small, books should be borrowable through some inter-library loans system). Since it is assumed that you will by default be using library copies of books, I have *not* made cost or being currently in print a significant consideration. However:

> I have marked with one star* books that are available new or second-hand at a reasonable price (or at least are unusually good value for the length and/or importance of the book).
>
> I have marked with two stars** those books for which e-copies are freely and legally available, and links are provided.
>
> Where articles or encyclopaedia entries have been recommended, these can almost always be freely downloaded, again with links supplied.
>
> I give short tinyurl links in square brackets when convenient: so, for example, '[opensettheory]' is short for 'tinyurl.com/opensettheory'.

We must pass over in silence the question of downloading books from a certain well-known and extremely well-stocked copyright-infringing PDF repository. That is between you and your conscience. I could not possibly comment . . .

This re-titled version of the Guide expands and revises earlier editions of my much-downloaded *Teach Yourself Logic*. Many thanks, then, to all those who have commented on different versions of *TYL* over a decade: and as always, further comments and suggestions are always gratefully received.

Preliminaries

The page left intentionally blank

1 A brief introduction for philosophers

1.1 Why this Guide for philosophers?

It is an odd phenomenon, and a very depressing one too. Logic, beyond a very elementary introductory level, is taught less and less, at least in UK philosophy departments. Fewer and fewer philosophers with serious logical interests seem to get appointed to permanent posts.

Yet logic itself is, of course, no less exciting and rewarding a subject than it ever was, and the amount of significant formally-informed work in philosophy is ever greater as time goes on. Moreover, logic is far too important to be left entirely to the mercies of technicians from maths or computer science departments with different agendas (who often reveal an insouciant casualness about basic conceptual issues that will matter to the more philosophical reader).

So how is a real competence in logic to be passed on if there are not enough courses, or indeed if there are none at all? It seems that many beginning graduate students in philosophy will need to teach themselves logic from books, either solo or by organizing their own study groups (local or online).

In a way, this is perhaps no real hardship; there are some wonderful books written by great expositors out there. But *what* to read and work through? Logic books can have a *very* long shelf life, and you shouldn't at all dismiss older texts when starting out on some topic area: so there's more than a sixty year span of publications to select from. There are hundreds of good books that might feature somewhere in a Study Guide such as this.

Philosophy students evidently need a Guide if they are to find their way around the very large literature old and new, with the aim of teaching themselves enjoyably and effectively. This is my attempt to provide one.

1.2 But how far do you want to go?

This Guide begins by looking at the standard 'mathematical logic' curriculum – taking this to cover first-order logic, a little model theory, some formal arithmetic and computability theory, and some set theory.

Now, you will certainly want to know this amount of core logic and indeed rather more, if you aim to be a philosopher of mathematics for example, or want to work on various areas of 'philosophical logic'/philosophy of language. But for many other purposes, you can get by with less. Maybe it will in fact suffice to read the relevant parts of

> Eric Steinhart, *More Precisely: The Math You Need to Do Philosophy** (Broadview, 2nd edition 2017: there is a companion website with answers to exercises at [moreprecisely].)
>
> The author writes: "The topics presented ... include: basic set theory; relations and functions; machines; probability; formal semantics; utilitarianism; and infinity. The chapters on sets, relations, and functions provide you with all you need to know to apply set theory in any branch of philosophy. The chapter of machines includes finite state machines, networks of machines, the game of life, and Turing machines. The chapter on formal semantics includes both extensional semantics, Kripkean possible worlds semantics, and Lewisian counterpart theory. The chapter on probability covers basic probability, conditional probability, Bayes theorem, and various applications of Bayes theorem in philosophy. .... The chapters on infinity cover recursive definitions, limits, countable infinity, Cantor's diagonal and power set arguments, uncountable infinities, the aleph and beth numbers, and definitions by transfinite recursion. *More Precisely* is designed both as a text book and reference book to meet the needs of upper level undergraduates and graduate students. It is also useful as a reference book for any philosopher working today."

Steinhart's book is admirable, and will give many philosophers a handle on some technical ideas going well beyond first-year logic and which they really should know just a little about, without all the hard work of doing a full mathematical logic course. What's not to like? This could indeed be enough for you.

Still, there will always be discriminating philosophers who want to go further and study logic in much greater depth than a book like Steinhart's can provide. If you are one, read on!

1.3 Assumed background

So what do you need to bring to the party, if you are going to begin tackling some of the books recommended in later chapters of this Guide? You should be well-enough prepared if you have worked through a book like one of the following:

1. Peter Smith, *Introduction to Formal Logic*** (2nd edition, CUP, 2020; corrected version now freely downloadable from logicmatters.net/ifl). This was the first year text in Cambridge for a decade. Written as an accessible teach-yourself book, it covers basic propositional and predicate logic. The first edition did logic 'by trees'. This new second edition instead focuses on natural deduction; but material on trees is still available in the form of online supplements.

2. P. D. Magnus, Tim Button, and others, *forallx*** (now part of the Open Logic Project and freely available from [forallx].) Also focuses on natural deduction, and is considerably brisker and less expansive than my book.

3. Nicholas Smith, *Logic: The Laws of Truth* (Princeton UP 2012). This excellent introduction is very clearly written, and has many virtues (particularly if you like your texts to go rather slowly and discursively). The first two parts of the book focus on logic by trees. But the third part ranges wider, including a brisk foray into natural deduction. There are some extras too, going significantly beyond the basics. And there is a rich mine of end-notes. The book's website is at [lawsoftruth].

However, if your initial logic course was based instead on some more elementary text book like Sam Guttenplan's *The Languages of Logic*, Howard Kahane's *Logic and Philosophy*, or Patrick Hurley's *Concise Introduction to Logic*, then you might struggle with the initial suggestions in the main part of the present Guide. But this will inevitably vary a lot from person to person. So just make a start on the suggested mathematical logic texts and see how things work out; if the going gets too tough, backtrack to read one of the introductory books like my own, skipping quickly over what you already know.

While if you have only done an 'informal logic' or 'critical reasoning course', and haven't any mathematical background, then you very probably *will* need to read a good introductory formal logic text before tackling the more advanced topics covered in the body of this Guide.

1.4 How to prove it

Experience shows that being able to handle (say) natural deduction proofs in a formal system doesn't always go with being able to construct good informal proofs. For example, one of the few meta-theoretic results that might be met in a first logic course is the so-called expressive completeness of the set of formal connectives $\{\wedge, \vee, \neg\}$. And the proof of this result is based on a very simple idea. But some philosophy students who ace the part of the end-of-course exam asking for *formal* proofs inside a deductive system find themselves all at sea when asked to outline an acceptable *informal* proof elaborating that proof idea.

Another example: it is only too familiar to find philosophers introduced to set notation not being able even to make a start on giving a correct informal proof that $\{\{a\}, \{a, b\}\} = \{\{a'\}, \{a', b'\}\}$ if and only if $a = a'$ and $b = b'$.

Well, if you *are* one of those who jumped through the formal hoops but were unclear about how to set out elementary mathematical proofs (e.g. proofs from the 'metatheory' of elementary logic or from very introductory set theory), then you will certainly profit by reading an introductory book on proof-writing. Here are two suggestions:

4. Daniel J. Velleman, *How to Prove It: A Structured Approach** (CUP, 3rd edition, 2019). From the Preface: "Students ... often have trouble the first time that they're asked to work seriously with mathematical proofs, because they don't know 'the rules of the game'. What is expected of you if you are asked to prove something? What distinguishes a correct proof from an incorrect one? This book is intended to help students learn the answers to these questions by spelling out the underlying principles involved in the construction of proofs." There are chapters on the propositional connectives and quantifiers, and on key informal proof-strategies for using them; there are chapters on relations and functions, a chapter on mathematical induction, and a final chapter on infinite sets (countable vs. uncountable sets).

   This is a truly excellent student text, and working through it from the beginning could be exactly what you need to get you prepared for the serious study of logic. Even if you were one of those comfortable with the informal proofs, you will probably still profit from skipping and skimming through the book (perhaps paying especial attention to the chapter on mathematical induction).

5. Joel David Hamkins, *Proof and the Art of Mathematics** (MIT Press, 2020) From the blurb: "This book offers an introduction to the art and craft of proof-writing. The author ... presents a series of engaging and compelling mathematical statements with interesting elementary proofs. These proofs capture a wide range of topics ... The goal is to show students and aspiring mathematicians how to write [informal!] proofs with elegance and precision." A less conventional text than Velleman's. Attractively written (though it has to be said rather uneven in level and tone). Readers without much of a mathematical background at all could still well enjoy this, and will learn a good deal, e.g. about proofs by induction. Lots of striking and memorable examples.

2 An even briefer introduction for mathematicians

2.1 Why this Guide for mathematicians?

Mathematics students soon pick up a passing acquaintance with some *very* basic notions about sets and some logical symbolism. But usually they get no systematic introduction to formal logic. And while they may be told early on that set theory provides a foundation for mathematics (in some sense), most students won't encounter even the outlines of a detailed set theory. Indeed, there are full university maths courses in good UK universities with precisely *zero* courses offered on the core mathematical logic curriculum – first-order logic and basic model theory, the theory of computability, set theory. And the situation can be equally patchy elsewhere.

So if you want to teach yourself some serious logic, where should you start? What are the topics you might want to cover? What textbooks are likely to prove accessible, engaging, and rewarding to work through? As with other areas of mathematics, the textbook options are many and various, and logic books can have a very long shelf life. So there's more than a sixty year span of publications to select from, and that's hundreds of books. This Guide will give you pointers to the geography of the subject-area, and offer recommendations of some of the best options for self-study.

True, this is put together by someone who, apart from a few guest mini-courses, has taught in philosophy departments and who has never been a research mathematician. Which no doubt gives a distinctive tone to the Guide. Still, mathematics remains my first love, and most of the books I recommend on core topics are very definitely paradigm *mathematics* texts: I won't be leading you astray.

2.2 Assumed background

There is no specific mathematical knowledge you need before tackling the entry-level mathematical logic books surveyed in the first main part of this Guide. They don't presuppose very much 'mathematical maturity', so you really can just dive in.

I did however recommend to philosophers a quite excellent book (written by a mathematician for mathematicians) which talks in an introductory way about logical notation and its use in proofs – namely Daniel J. Velleman, *How to Prove It: A Structured Approach** (CUP, 3rd edition, 2019). This very well-regarded book could serve as a very useful way of organizing and reinforcing your informal logical ideas if your grasp really is *very* shaky and fragmentary.

On the other hand, suppose you are already know how to use the so-called truth-functional connectives symbolized '$\wedge$' or '&', '$\vee$', '$\neg$' or '$\sim$', and '$\rightarrow$'. Suppose too that you are quite comfortable with the use of formal quantifiers, e.g. as in $\forall\epsilon\exists\delta$ used in stating the ϵ/δ definition of continuity. And suppose you are already entirely happy with various proofs by induction – as surely you are! Then you won't strictly speaking *need* Velleman's introduction (though I guess it could do no harm to speed-read the book, slowing down if anything looks a bit unfamiliar).

3 The Guide, and how to use it

This chapter describes the overall structure and coverage of this Guide. I also offer a little advice about how to make the best use of it.

3.1 "Est omnis divisa in partes tres"

There is a fourth preliminary chapter, on *'naive' set theory* – which reviews the concepts and constructions typically taken for granted in quite elementary mathematical writing (not just in texts about logic). After that, the main Guide, like Gaul, divides into three parts.

In headline terms,

Part I covers the now standard core mathematical logic curriculum at an introductory level.

Part II expands the range of logical topics, while still remaining at roughly the same level of mathematical difficulty.

Part III then gives pointers forward to more advanced treatments of some of the areas covered in Parts I and II.

The chapters in Parts I and II carve up the broad field of logic in a pretty conventional way: but of course, even these 'horizontal' divisions into different subfields can in places be a little arbitrary. And the 'vertical' divisions between the entry-level coverage in Parts I and II and the further explorations of the same areas in Part III are necessarily going to be a *lot* more arbitrary. At least in retrospect, everyone will agree that e.g. the elementary theory of ordinals and cardinals belongs to the basics of set theory, while explorations of 'large cardinals' or independence proofs via 'forcing' are decidedly more advanced. But in most areas, there are fewer natural demarcation lines between the entry-level basics and more advanced work. Still, it is surely very much better to impose *some* such structuring than to heap everything together.

Let me now say a *little* more about the contents of the first two Parts, and about the role of Part III. Don't worry, of course, if these headline gestures are not immediately very clear – all will be explained in due course!

3.2 On Part I: Core topics

The traditional menu of core topics has indeed remained fairly fixed ever since e.g. Elliott Mendelson's classic textbook *Introduction to Mathematical Logic* first published in 1964. And this menu is reflected in the following chapters. First, therefore,

Chapter 5 discusses *classical first-order logic* (FOL), which is at the fixed centre of any mathematical logic course.

As we will see in Part II, there is a variety of ways of extending and/or deviating from FOL. But there is one extension it is worth knowing about straight away (in order to understand some themes you'll meet in later chapters in Part I). So:

Chapter 6 goes beyond first-order logic by briefly looking at *second-order logic*. (Second-order languages have more ways of forming general propositions than first-order ones.)

You can then take the topics of the following three chapters in whatever order you choose:

Chapter 7 introduces a modest amount of *model theory* which, roughly speaking, explores how formal theories relate to the structures they are about.

Chapter 8 looks at one particular kind of formal theory, i.e. *formal arithmetics*, and relatedly at the theory of computable functions: we arrive at proofs of epochal results such as Gödel's incompleteness theorems.

Chapter 9 is on *set theory* proper – beginning again fairly informally, examining basic notions of cardinals and ordinals, constructions of number systems in set theory, the role of the axiom of choice, etc. We then look at the standard formal axiomatization of ZFC, and nod towards alternatives.

3.3 On Part II: More logic

What else should you encounter when beginning serious logic? There is a whole area of logic which tends to be passed by in introductory texts, so

Chapter **??** says something about *proof theory*. OK, that label is pretty unhelpful given that most areas of logic deal with proofs, but it conventionally points to investigations of a cluster of issues about the structure of proofs and the consistency of theories, etc. Some would say that at least the more elementary parts of proof theory really belong earlier in this Guide, as part of the initial study of FOL.

Then, as we said, there is variety of ways of deviating from the classical first-order paradigm, some of which we already touch on in discussing proof theory. In particular, *Chapter* **??** introduces one important variety of 'non-classical' logic, namely *intuitionist logic*, which drops the classical principle that, for any proposition, either it or its negation is true.

We need to explain why we might be interested in dropping the Law of Excluded Middle like this: and indeed we will see that intuitionist logic rather naturally appears in a number of different contexts. And in developing intuitionist logic, we encounter a new way of thinking about the meanings of the logical operators, using so-called 'possible world semantics'. This takes us into the territory of

Chapter **??** on *modal logic*. Modal logic deals, in the first instance, with logical notions of necessity and possibility – of what is true 'in all possible worlds' (as they say), as opposed to what happens to be true 'in the actual world' but could have been otherwise (i.e. can be false in other possible worlds). Modal logics are of particular interest to philosophers; but they also turn out to have important applications e.g. of interest to computer scientists.

There is indeed something to be said for first tackling modal logics and their possible-world semantics in general, before looking at the particular application to intuitionist logic.

This still leaves us with further varieties of logic to discuss. These are touched on more briskly in

Chapter **??** on *other logics*. We look at a selection of topics, probably of most interest to philosophically-minded logicians. These include so-called relevant logics (where we impose stronger requirements on the relevance of premisses to conclusions in genuinely valid arguments), free logics (where we no longer presuppose that e.g. names in a formal language actually name something), and plural logics (where we can e.g. cope with terms denoting more than one thing, like 'Russell and Whitehead' and 'the Brontës').

3.4 How chapters in Parts I and II are structured

From Chapter 5 to Chapter **??**, each chapter includes at least one *overview* of its topic(s). These are intended to give helpful pointers to the coverage of the chapter; but if these necessarily brisk headlines sometimes mystify, feel free to just skim through. The overviews have occasional footnotes which hint at more technical points; these can certainly be skipped.

Overviews are followed by a list of main recommended texts for the chapter's topic(s), put into what strikes me as a sensible reading order. I then offer some suggestions for alternative/additional reading at about the same level. Since it is always quite illuminating to know just a little of the background history of a topic, I also give some brisk references for those who might be interested.

The earliest versions of this Guide kept largely to positive recommendations: I didn't pause to explain the reasons for the then absence of some well-known books. This was partly due to considerations of length which have now gone by the wayside; but also I wanted to keep the tone enthusiastic, rather than to start criticizing or carping. However, enough people kept asking what I think about some alternative X, or asking why the old warhorse Y wasn't mentioned, for me to change my mind. So I have added episodes to some chapters where I give reasons why I *don't* particularly recommended certain books.

3.5 On Part III: Going further

The chapters in Parts I and II are intended to provide a coherent and reasonably systematic survey of basic areas in and around the core of mathematical logic, at an advanced undergraduate/beginning graduate student level. And once you have studied enough of the suggested readings, you should in fact be in a very good position to continue into more advanced work on particular topics under your own steam. So you probably won't really *need* Part III of this Guide which gives my own suggestions for more advanced reading. But it is here anyway, for what it is worth. The coverage of various topics is now much more varied and rather patchy: the recommendations can be many or few (or non-existent!) depending on my own personal interests and knowledge. Make what use of it you will.

3.6 Strategies for self-teaching from logic books

We cover a good deal of ground in this Guide which is one reason for its initially daunting length. But there is another reason, connected to a point which I now really want to highlight:

> I very strongly recommend tackling an area of logic by reading a series of books which *overlap* in level (with the next one covering some of the same ground and then pushing on from the previous one), rather than trying to proceed by big leaps.

In fact, I probably can't stress this bit of advice too much (which, in my experience, applies equally to getting to grips with any new area of mathematics). This approach will really help to reinforce and deepen understanding as you encounter and re-encounter the same material, coming at it from somewhat different angles, with different emphases.

Exaggerating only a little, there are many instructors who say 'This is the textbook we are using/here is my set of notes: take it or leave it'. But you will *always* gain from looking at a variety of treatments, perhaps at slightly different levels. The multiple overlaps in coverage in the reading lists in later chapters, which help make the Guide as long as it is, are therefore fully intended. They also mean that you should always be able to find the options that best suit your degree of mathematical competence and your preferences for textbook style.

To repeat: you will certainly miss a lot if you concentrate on just one text in a given area, especially at the outset. Yes, do very carefully read one or two central texts, chosing books that work for you. But do also cultivate the crucial further habit of judiciously skipping and skimming through a number of other works so that you can build up a good overall picture of an area seen from various angles of approach, and with different degrees of sophistication.

While we are talking about strategies for self-teaching, I suppose I should add a quick remark on the question of doing exercises.

> Mathematics is, as they say, not merely a spectator sport: so you should try some of the exercises in the books as you read along, in order to check and reinforce comprehension. On the other hand, don't obsess about this, and do concentrate on the exercises that look interesting and/or might deepen understanding.

Note that some authors have the irritating(?) habit of burying quite important results among the exercises, mixed in with routine homework. It is therefore always a good policy to skim through the exercises in a book even if you don't plan to work on answers to very many of them.

Oddly – even in these days where books can have websites to support them – it isn't that common for logic texts to have very detailed worked solutions to exercises available. I will try to highlight those authors who are more helpful than usual in this respect.

3.7 Choices, choices

So what has guided my choices of texts to recommend?

Different people find different expository styles congenial. What is agreeably discursive for one reader might be irritatingly verbose and slow-moving for another. For myself, I do particularly like books that are good on conceptual details and good at explaining the motivation for the technicalities while avoiding needless complications, excessive hacking through routine detail, or misplaced 'rigour'. Given the choice, I prefer a treatment that highlights intuitive motivations and doesn't rush too fast to become too abstract: this is especially what we want in books to be used for self-study. (There's a certain tradition of masochism in older maths writing, of going for brusque formal abstraction from the outset with little by way of explanatory chat: this is quite unnecessary in other areas, and just because logic is all about formal theories, that doesn't make it any more necessary here.)

The selection of readings in the following chapters no doubt reflects these tastes. But overall, while I have no doubt been opinionated, I don't think that I have been very idiosyncratic: indeed, in many respects I have probably been really rather conservative in my choices. So nearly all the readings I recommend will very widely be agreed to have significant virtues (even if other logicians would have different preference-orderings).

4 A little informal set theory

Notation, concepts and constructions from basic set theory are often presupposed in quite elementary mathematical texts – including some of the introductory logic texts mentioned in the following chapters, before we get round to officially studying set theory itself. So in §4.1 I note what you ideally should know about sets at the outset. It isn't a lot! – and to be honest, you could probably manage pretty well for a while without even this much. For now, we proceed 'naively' – i.e. we proceed quite informally, and will just *assume* that the various constructions we talk about are permitted, etc. §4.2 gives recommended readings on naive basic set theory for those who need them. In §4.3 I point out that in many elementary contexts, the conventional use of set-talk can in fact be eliminated without serious loss.

4.1 Sets: a checklist of some basics

(a) So what elementary ideas *should* you be familiar with, given our limited current purposes? Let's have a quick checklist:

(i) A set is a single thing, which exists over and above its members but whose identity is fully determined by its members. So A and B are one and the same set if and only if whatever is a member of A is a member of B and vice versa (that's the *extensionality* principle).

(ii) We use the likes of '$\{a, b, c, d\}$' to denote the set whose members are a, b, c, d. And we use the likes of '$\{x \mid x \text{ is } F\}$', which can be read as 'the set of all x such that x is F', to denote the set of things (in some domain) which are F.

Membership is symbolized by '$\in$' – hence, for example, $b \in \{a, b, c, d\}$ and $4 \in \{x \mid x \text{ is even}\}$.

The subset relation is symbolized by '$\subseteq$' so $A \subseteq B$ is true if and only if every member of A is a member of B. A is a proper subset of B, $A \subset B$, if $A \subseteq B$ but $A \neq B$.

The membership and subset relations need to be sharply distinguished from each other. And note in particular that the singleton set $\{a\}$ is to be distinguished from its sole member a: thus $a \in \{a\}$ and $\{a\} \subseteq \{a\}$, but not $a = \{a\}$ and not $a \subseteq \{a\}$.

(iii) If A, B are sets, so are their union, intersection and their powersets.

The union of A and B, $A \cup B$, is the set which contains just those things which are members of at least one of A and B.

The intersection of A and B, $A \cap B$, is the set which contains all and only those things which are members of both A and B. If intersections are always to exist, then we have to allow sets which contain no members (since A and B might not overlap). By extensionality, the empty set must be unique; it is standardly symbolized $\emptyset$.

The powerset of A, $\mathcal{P}(A)$, is the set whose members are all and only the subsets of A: note this assumes that sets are themselves things which can be members of other sets.

(iv) Sets are in themselves unordered. Because they have the same members, $\{1, 2, 3, 4\}$ counts as the same set as $\{4, 1, 3, 2\}$ (and the same set as $\{1, 2, 3, 4, 1, 2, 3, 4\}$). But we often need to work with ordered pairs, ordered triples, ordered quadruples, ..., *tuples* more generally.

We use '$\langle a, b\rangle$' – or often simply '(a, b)' – for the ordered pair, first a, then b. So, while $\{a, b\} = \{b, a\}$, by contrast $\langle a, b\rangle \neq \langle b, a\rangle$.

Once we have ordered pairs available, we can use them to define ordered triples: $\langle a, b, c\rangle$ can be defined as first the pair $\langle a, b\rangle$, then c, i.e. as $\langle\langle a, b\rangle, c\rangle$. Then the quadruple $\langle a, b, c, d\rangle$ can be defined as $\langle\langle a, b, c\rangle, d\rangle$. And so it goes.

To deal with tuples, then, it is enough just to have ordered pairs available. And we can define ordered pairs using unordered sets in various ways: all we need is some definition for ordered pairs which ensures that $\langle a, b\rangle = \langle a', b'\rangle$ if and only if $a = a'$ and $b = b'$. The following works and is standard: $\langle a, b\rangle =_{\text{def}} \{\{a\}, \{a, b\}\}$.

(v) The Cartesian product $A \times B$ of the sets A and B is the set whose members are all the ordered pairs whose first member is in A and whose second member is in B (cf. Cartesian coordinates). In an obvious notation, $A \times B$ is $\{\langle x, y\rangle \mid x \in A \;\&\; y \in B\}$. Cartesian products of n sets are defined as sets of n-tuples, again in the obvious way.

(vi) If R is a binary relation between members of the set A and members of the set B, then its extension is the set of ordered pairs $\langle x, y\rangle$ (with $x \in A$ and $y \in B$) such that x is R to y. So the extension of R is a subset of $A \times B$.

The extension of an n-place relation is the set of n-tuples of things which stand in that relation. In the unary case, where P is a property defined over some set A, then we can simply say that the extension of P is the set of members of A which are P.

For certain purposes, we might want to individuate properties (relations) in a fine-grained way, so that distinct properties (relations) can have the same extension. For example, you might say that the property of being a terrestrial featherless biped and the property of being human are different properties; but they do have the same extension (or so the conventional story goes). For many mathematical purposes, however, we can treat properties and relations extensionally; i.e. we regard properties with the same extension as being the same property, and regard relations with the same extension as being the same relation. Indeed, we can often simply treat a property (relation) as if it *is* its extension.

(vii) Likewise, the extension (or graph) of a unary function f which sends members of A to members of B is the set of ordered pairs $\langle x, y\rangle$ (with $x \in A$ and $x \in B$) such that $f(x) = y$. So the extension of f is the same set as the extension of the corresponding relation R such that Rxy if and only if $f(x) = y$. Similarly for many-place functions. For many purposes, we treat functions extensionally, regarding functions with the same extension as the same. Again we often treat a function as if it *is* its extension, i.e. we identify a function with its graph.

(viii) Relations can, for example, be reflexive, symmetric, transitive; equivalence relations are all three. Functions can be injective, surjective, bijective, etc. Those are familiar pre-set-theoretic ideas. When we officially identify relations and functions with their extensions, these various features become corresponding features of the sets which are extensions.

One point to highlight. If $\equiv$ is an equivalence relation defined over some set, it partitions that set into equivalence classes (we never say 'equivalence sets'!) of objects standing in that relation. If $[x]$ is the equivalence class (with respect to $\equiv$) containing x, then $[x] = [y]$ if and only if $x \equiv y$.

(ix) Two sets are equinumerous just if we can match up their members one-to-one, i.e. when there is a one-to-one function, a bijection, between the sets. A set is countably infinite if and only if it is equinumerous with the natural numbers.

And here we get to the first exciting claim – *there are infinite sets which are not countably infinite*: a standard example is the set of infinite binary strings.[1] And this is just the beginning of a story about how sets can have different infinite 'sizes' or cardinalities. Indeed, there is an unending sequence of ever-larger infinite cardinalities. But at this stage you need to know little more than this bald fact: further elaboration can wait.

---

[1] Here's the argument – fun, if you've not seen it before! Suppose we try to match up the strings one-to-one with the natural numbers, perhaps like this:

0: 01011100101 ...
1: 10011000100 ...
2: 10100111111 ...
3: 10100111110 ...
4: 10100111011 ...
5: 00111101111 ...
6: ......

Now 'go down the diagonal [from northwest to southeast] and flip digits' to get the binary string which starts 110110 .... By construction, this new binary string differs from the first string on our list in the first place, differs from the second in the second place, the third in the third place, etc., so cannot appear anywhere in our given list. Hence, that particular list cannot contain all the infinite binary strings after all. And the argument evidently generalizes: no countably infinite list of binary strings can contain all possible binary strings.

(x) There's one further, rather less elementary, idea that you should perhaps also meet sooner rather than later, so that you recognize any passing references to it. This is the Axiom of Choice. In one version, this says that, given an infinite collection of sets, there is a choice function – i.e. a function which, given as input any one of those sets in the family, 'chooses' a single member from that set as output. (Bertrand Russell's toy example: given an infinite collection of pairs of socks, there is a function which chooses one sock from each pair.)

Note that while other principles for forming new sets (e.g. unions, powersets) determine what the members of the new set are, Choice just tells us that there *is* a set (the extension of the choice function) which plays a certain role, without pinning down its members. At this stage you basically just need to know that Choice is a principle which is implicitly or explicitly invoked in many mathematical proofs. But you should also know that it is independent of other basic set-theoretic principles (and there are set theories in which it doesn't hold) – which is why we often explicitly note when, in more advanced logical theory, a result does indeed depend on Choice.

(b) An important observation before proceeding.

You've very probably met Russell's Paradox. Say a set is *normal* if it isn't a member of itself. The set of musketeers $\{\mathrm{Athos}, \mathrm{Porthos}, \mathrm{Aramis}\}$ is not another musketeer and so is not a member of itself; again, the set of prime numbers isn't itself a prime number, so also isn't a member of itself; so these are indeed normal sets.

Now we ask: is there a Russell set R whose members are all and only the normal sets?

No. For if there were, it would be a member of itself if and only if wasn't – think about it! – which is impossible. The putative set R is, in some sense, 'too big'. Hence, if we overshoot and suppose that for *any* property (including the property of being a normal set) there is a set which is its extension, we get into deep trouble.

Here we need to avoid tripping over one of those rather annoying terminological divergences. Some people use 'naive set theory' to mean, specifically, a theory which makes that simple but hopeless assumption that any property has a set as an extension. As we've just seen, naive set theory in *this* sense is inconsistent. Many others, however, use 'naive set theory' just to mean set theory developed informally, without rigorous axiomatization, but guided by unambitious low-level principles. In this second sense, we have been proceeding naively in this chapter, but let's hope we are still on track for a consistent story! Thus, we were careful in (vi) to assign extensions just to those properties and relations that are defined over domains we are already given as sets. True, our story so far is silent about exactly which putative sets *are* the kosher ones – i.e. are not 'too big' to be to be problematic. However, important though this is, we can leave this topic until Chapter 9 when we turn to set theory proper. Low-level practical uses of sets in 'ordinary' mathematics seem remote from such problems, so hopefully we can proceed safely for now.

4.2 Recommendations on informal basic set theory

You only need a very modest mathematical background for the topics on our checklist to be already entirely familiar; and if they are, you can now skip over these first reading suggestions. But non-mathematicians should find one of the following to be exactly what they need:

1. Tim Button, *Set Theory: An Open Introduction*** (Open Logic Project), Chapters 1–5. Available at [opensettheory]. Read Chapter 1 for some interesting background. Chapter 2 introduces basic notions like subsets, powersets, unions, intersections, pairs, tuples, Cartesian products. Chapter 3 is on relations (treated as sets). Chapter 4 is on functions. Chapter 5 is on the size of sets, countable vs uncountable sets, Cantor's Theorem. At this stage in his book, Button is proceeding naively in our second sense, with the promise that everything he does can be replicated in the rigorously axiomatized theory he introduces later.

   Button writes, here as elsewhere, with very admirable clarity. So this is warmly recommended.

2. David Makinson, *Sets, Logic and Maths for Computing* (Springer, 3rd edn 2020), Chapters 1 to 3. This is exceptionally clear and very carefully written for students without much mathematical background. Chapter 1 reviews very basic facts about sets. Chapter 2 is on relations. Chapter 3 is on sets. Again this can be warmly recommended (though you might want to supplement it by following up his reference to Cantor's Theorem).

Now, Makinson doesn't mention the Axiom of Choice at all. While Button does eventually get round to Choice in his Chapter 16; but the treatment there depends on the set theory developed in the intervening chapters, so isn't appropriate for us just now. Instead, the following two pages should be enough for the present:

3. Tim Gowers et al. eds, *The Princeton Companion to Mathematics* (Princeton UP, 2008), §III.1: The Axiom of Choice.

We return to set theory in Chapter 9. But let me mention two other things that – whether you are a philosopher or mathematician – you could usefully read in advance. First, there is an old but much used short text:

4. Paul Halmos, *Naive Set Theory** (1960: republished by Martino Fine Books, 2011).

The purpose of this famous book, Halmos says in his Preface, is "to tell the beginning student ... the basic set-theoretic facts of life, and to do so with the minimum of philosophical discourse and logical formalism". Again he is proceeding naively in our second sense. True he tells us about some official axioms as he goes along, but he doesn't explore the development of set theory inside a resulting formal theory: this is informally written in an unusually conversational style for a maths book, concentrating on the motivation for various concepts and constructions. This is all done very approachably. You could read the first fifteen – very short – chapters now, leaving the rest for later.

And it is well worth seeking out this famous discussion where we meet Russell's infinite collection of socks:

5. Bertrand Russell, *An Introduction to Mathematical Philosophy*** (1919), Chapter XII, 'Selections and the Multiplicative Axiom'. Available at [russellimp].

The 'Multiplicative Axiom' is Russell's name for a version of the Axiom of Choice. (In fact, the whole of Russell's lovely book is *still* a wonderful read if you have any interest in the foundations of mathematics!)

4.3 Virtual classes, real sets

A set, we said following Cantor, is a unity, a single thing in itself over and above its members. But if that is the guiding idea, then it is important to note that *a great deal of elementary informal set talk in mathematics 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 apparent talk of a set doesn't really carry any serious commitment to there being any *additional* object, a set, over and above those many things. On the contrary, in such contexts, apparent talk about a *set of* F*s* can very often be paraphrased away into direct talk about those Fs, without any loss of content.

Here is just one example, relevant for us. It is usual to say something like this: (1) "A set of formulas Γ logically entails the formula φ if and only if any structure which makes every member of Γ true makes φ true too". Don't worry for now about the talk of structures: just note that the reference to a *set* of formulas and it *members* is doing no work here. It would do just as well to say (2) "Some formulas G (whether they are zero, one or many) logically entail φ if and only if every structure which makes all of those formulas G true makes φ true too". The set version (1) adds nothing important to the plural version (2).

When it can be paraphrased away like this, talk of sets is just talk of virtual sets – or rather, in the more conventional idiom, it's talk of of *virtual classes.*

One source for this conventional usage is W.V.O. Quine's famous discussion in the opening chapter of his *Set Theory and its Logic* (1963):

> Much ... of what is commonly said of classes with the help of '$\in$' can be accounted for as a mere manner of speaking, involving no real reference to classes nor any irreducible use of '$\in$'. ... [T]his part of class theory ... I call the virtual theory of classes.

You will find that this same usage plays an important role in set theory in some treatments of so-called 'proper classes' as distinguished from sets. For example, in his standard book *Set Theory* (1980), Kenneth Kunen writes

> Formally, proper classes do not exist, and expressions involving them must be thought of as abbreviations for expressions not involving them.

And here is Paul Finsler, writing much earlier in 1926 (as quoted by Luca Incurvati, in his *Conceptions of Set*):

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

Nothing really hangs, of course, on the particular terminology here, 'classes' vs 'sets'. What matters is the distinction between non-committal, eliminable, talk – talk of merely virtual sets/classes/pluralities (whichever idiom we use) – and uneliminable talk of sets as entities in their own right.

Now this certainly isn't the place to start discussing exactly where and why, as we get into the serious study of logic, we do eventually get entangled with sets as genuine entities. I am just flagging up that, at least when starting out on logic, you *can* take a lot of set talk just as a non-committal light-weight idiom, simply there to conveniently enable us to talk in the singular about many things at once.

Part I

Core Topics in Mathematical Logic

The page left intentionally blank

5 First-order logic

Let's get down to business! This chapter starts with an overview of the basics of classical first-order logic (or predicate logic, quantificational logic, call it what you will: I'll use 'FOL' for short). At this level, the most obvious difference between various treatments of FOL is in the choice of proof-system: so I will comment on some options in a little more detail. Then I highlight the main self-study recommendations. These are followed by some suggestions for parallel and further reading. After a short historical section, the chapter ends with some additional comments, mostly responding to frequently asked questions.[1]

5.1 FOL: a general overview

FOL deals with deductive reasoning that turns on the use of 'propositional connectives' like *and*, *or*, *if*, *not*, and on the use of 'quantifiers' like *every*, *some*, *no*. But in ordinary language (and even in informal mathematics) these logical operators work in quite complex ways, introducing the kind of obscurities and possible ambiguities we want to avoid in logically transparent arguments. What to do?

From the time of Aristotle, logicians have used a 'divide and conquer' strategy that involves introducing restricted 'formalized' languages. We tackle a stretch of reasoning by first reformulating it in a suitable regimented language with tidier logical operators, and then we can evaluate the reasoning once recast into this more well-behaved form. This way, we have a division of labour. There's first the task of working out the intended structure of the original argument as we render it into an ambiguous formal language. Then there's the separate business of assessing the validity of the resulting regimented argument.

In FOL, therefore, we use appropriate formal languages which contain, in particular, tidily-disciplined surrogates for the propositional connectives *and*,

[1] A note to philosophers. If you *have* carefully read a substantial introductory logic text for philosophers such as Nick Smith's, or even my own, you will already be familiar with (versions of) an amount of the material covered in this chapter. However, you will now begin to see topics being re-presented in the sort of mathematical style and with the sort of rigorous detail that you will necessarily encounter more and more as you progress in logic. You do need to start feeling entirely comfortable with this mode of presentation at an early stage. So it is well worth working through even rather familiar topics again, this time with more mathematical precision.

or, *if*, *not* (standardly symbolized $\wedge$, $\vee$, $\rightarrow$ and $\neg$), plus replacements for the ordinary language quantifiers (roughly, using $\forall \mathsf{x}$ for *every x is such that* ..., and $\exists \mathsf{y}$ for *some y is such that*...).

Although the fun really starts once we have the quantifiers in play, it is very useful to develop FOL in two main stages. So:

(a) Typically, we start by introducing propositional languages whose built-in logical apparatus comprises just the propositional connectives, and discuss the propositional logic of arguments framed in these languages. This gives us a very manageable setting in which to first encounter a whole range of logical concepts and strategies.

(b) We then move on to develop the syntax and semantics of richer formal languages which add the apparatus of so-called first-order quantification, and explore the logic of arguments rendered into such languages.

Let's have some more detail.

(a.i) We first look at the *syntax* of propositional languages, defining what count as the well-formed formulas of such languages. If you have already encountered such languages, you will now get to know how to prove various things about them that might seem obvious and that you perhaps previously took for granted – for example, that 'bracketing works' to avoid ambiguities, so every well-formed formula has a unique parsing.

(a.ii) On the *semantic* side, you need to understand the idea of a *valuation* for a propositional language. We start with an assignment of truth-values, *true* vs *false*, to the atomic components of our languages; and then we explain how to evaluate complex sentences involving the connectives by using the 'truth-functional' interpretation of the connectives.

We can then define the key semantic relation of *(tautological) entailment*, where a set of sentences Γ (tautologically) entails the sentence φ when any valuation which makes all the sentences in Γ true makes φ true too. We will explore some of the key properties of this semantic entailment relation, and learn how to calculate whether the relation holds.

(a.iii) Different textbook presentations of stages (a.i) and (a.ii) tend to be very similar; but now the path forks. For the next topic will be a deductive *proof-system* for propositional logic, and there is a variety of such systems to choose from. For a start, we have

1. Old-school 'axiomatic' systems.
2. Natural deduction done Gentzen-style.
3. Natural deduction done Fitch-style.
4. 'Truth trees' or 'semantic tableaux'.
5. Sequent calculi.
6. Resolution calculi

Different proof systems for classical propositional logic will (as you'd expect) be equivalent – meaning that, given some premisses, we can warrant the same conclusions in each system. However, they do differ considerably in their intuitive appeal and user-friendliness, as well as in some of their more technical features.[2] Note, though: apart from some looking at a few illustrative examples, we won't be much interested in producing lots of derivations *inside* a chosen proof system; the focus will be on establishing results *about* the systems.

The easiest for beginners to work with are (3) Fitch-style deductions and (4) truth trees – which is why the majority of elementary logic books for philosophers introduce one or other (or both) systems. By contrast, (5) the sequent calculus (in its most interesting form) really comes into its own in more advanced work (in proof theory), while (6) resolution calculi are perhaps of particular concern to computer scientists interested in automating theorem proving. Introductory mathematical logic text books, however, usually focus on either (1) axiomatic or (2) Gentzen-style proof systems. True, axiomatic systems in their raw state can initially be pretty horrible to use – but they can be made a bit less painful once you learn some basic dodges (like the use of the so-called 'Deduction Theorem'). By comparison, Gentzen-style systems are initially much more attractive – there is a reason they are called *natural* deduction systems!

In due course, the educated logician will want to learn at least a little about *all* these proof styles – at the minimum, you should eventually get a sense of how they respectively work, and come to appreciate the interrelations between them. But we'll start – as is usual at this level – by looking at (1) and (2).

(a.iv) At this stage, then, we have two quite different ways of defining what makes for a deductively good argument in propositional logic:

> We said that a set of premisses Γ *semantically entails the conclusion* φ if every possible valuation which makes Γ all true makes φ true.
>
> We can now also say that Γ *yields the conclusion* φ *in the proof-system* S if there is an S-type derivation of the conclusion φ from premisses in Γ.

Of course, we want these two approaches to fit together. We want our favoured proof-system S to be *sound* – it shouldn't give false positives. In other words, if there is an S-derivation of φ from Γ, then φ really *is* semantically entailed by Γ. We also want our favoured proof-system S to be *complete* – we want it to capture all the correct semantic entailment claims. In other words, if φ is semantically entailed by the set of premisses Γ, then there is indeed some S-derivation of φ from premisses in Γ.

[2] In fact, even once you've picked your favoured general *type* of proof-system to work with, there are many more choices to be made before landing on a particular system. Francis Jeffry Pelletier and Allen Hazen published an interesting survey of logic texts aimed at philosophers using natural deduction, available at [pellhazen]. They note, for example, that thirty texts use a variety of Fitch-style system; and rather remarkably no two of these have *exactly* the same system of rules for FOL! And even at the level of propositional logic there are notable divergences.

So we will want to establish both the soundness and the completeness of our favoured proof-system S for propositional logic (axiomatic, natural deduction, whatever). Now, these two results will hold no terrors! However, in establishing soundness and completeness for propositional logics we will encounter some useful strategies which can later be beefed-up to give us soundness and completeness results for stronger logics.

(b.i) Having warmed up with propositional logic, we can now turn to full FOL (predicate logic, quantificational logic). Again, *syntax* first. And while the syntax of propositional logic is quite straightforward, a story needs to be told about why FOL expressions of generality are structured as they are.

Compare the ordinary-language sentences (i) 'Socrates is wise' and (ii) 'Everyone is wise'. Here, the quantifier expression 'everyone' can occupy the same place as the name in (i), giving us another grammatical sentence (ii). Similarly, we can replace the name 'Juliet' in (iii) 'Romeo loves Juliet' with another quantifier expression to get the equally grammatical (iv) 'Romeo loves someone'. In FOL, however, while we might render (i) as simply Ws,[3] (ii) will get rendered by something like $\forall\mathsf{xWx}$ (roughly, everyone x is such that x is wise). Similarly if (iii) is rendered Lrj, then (iv) gets rendered by something like $\exists\mathsf{xLrx}$ (roughly, someone x is such that Romeo loves x). You need to understand the basic rationale for this apparently more complex 'quantifier/variable' mechanism for expressing generalizations.

(b.ii) Turning to *semantics*: the first key idea we need is that of a *structure*, a (non-empty!) domain of objects equipped with some properties, relations and/or functions. We here treat properties etc. extensionally – see §4.1(a.vi). In other words, we can think of a property as a set of objects from the domain, a binary relation as a set of pairs from the domain, and so on – though, heeding the point of §4.3, we can arguably take the talk of sets here in a non-committal way.

Then, crucially, you need to grasp the idea of an *interpretation* of a language in such a structure; names get assigned objects in the domain, a one-place predicate gets assigned a property, i.e. a set of objects from the domain (its extension – intuitively, the objects it is true of), and so on. Such an interpretation of the elements of a first-order language then generates a unique valuation (a unique assignment of truth-values) for every sentence of the interpreted language. How does it do that? We need a proper formal semantic story with the bells and whistles to explain how the interpretation of a language fixes the valuations of quantified sentences. There are in fact three slightly different though ultimately equivalent ways of spinning the story.

We can now introduce the idea of a *model* for a set of sentences, i.e. an interpretation in a structure which makes all the sentences true together. And we can again define a relation of semantic entailment, this time for FOL expressions. The set of FOL sentences Γ semantically entails φ when any interpretation in any structure which makes all the sentences Γ true also makes φ true too. Or

[3] The inversion of the subject-predicate order is conventional but not deeply significant.

for short, any model for Γ is a model for φ. You'll again need to know some of the basic properties of this entailment relation.

(b.iii) We next need to explore a proof system for FOL. Corresponding to the six types of system we mentioned for propositional logic, you can again encounter six different types of proof system, with their varying attractions. To repeat, you'll want at some future point to find out about all these styles of proof: but we will be looking at axiomatic systems and at one kind of natural deduction.

(b.iv) As with propositional logic, we will want *soundness* and *completeness* results which show that our chosen proof system for FOL doesn't overshoot (giving us false positives) or undershoot (leaving us unable to derive some semantically valid entailments). In other words, if S is our proof system, Γ a set of FOL sentences, and φ a particular sentence, we need to show

1. If there is an S-proof of φ from premisses in Γ, then Γ semantically entails φ.
2. If Γ semantically entails φ, then there is an S-proof of φ from premisses in Γ.

The completeness theorem (2), by the way, comes in two versions, a weaker version where Γ has have only finitely many members, and a stronger version which allows Γ to be infinite.

And it is at *this* point, it might well be said, that the study of FOL becomes really interesting: in particular, establishing completeness involves rather more sophisticated ideas than anything we have met before. So we will take this as our main target destination in an initial treatment of FOL.

(b.v) Proofs in formal systems are finite objects, so can only call on a finite number of premisses. But the strong completeness theorem for FOL, as we said, allows Γ to have an infinite number of members. And this combination of facts has an easy but important corollary, the so-called *finiteness* or *compactness* theorem for sentences of FOL languages:

3. If every finite subset of Γ has a model, so does Γ.

In other words, if each finite selection of sentences from Γ has an interpretation on which it is true, then there is an interpretation which makes all of Γ true together.[4]

We will find that this compactness theorem has numerous applications in model theory.

[4]Why so? Compactness is equivalent to the claim that if Γ has no model, then there is some finite $\Delta \subseteq \Gamma$ which has no model. And *that* is equivalent to the claim that if Γ semantically entails a contradiction, there is some finite $\Delta \subseteq \Gamma$ which entails a contradiction. OK: suppose that Γ, possibly infinite, *does* semantically entail some contradiction $\perp$. Then by strong completeness there is will be a proof of $\perp$ from premisses in Γ in your favourite proof system. But proofs are finite objects and hence this proof can only use some finite subset Δ of Γ as premisses. So we can derive $\perp$ from premisses in Δ in your favourite proof system. But that system is sound, which means that Δ must semantically entail $\perp$.

5.2 A little more about proof styles

The most striking difference between various treatments of FOL is in the choice of proof system. For orientation, especially if this is new to you, it might be useful if I say something brief and introductory about axiomatic vs natural deduction systems. (You can find out about other proof systems – particularly truth trees and sequent calculi – in e.g. Bostock's book mentioned in §5.4.)

(a) You will be familiar with the informal idea of an axiomatized theory. We are given some axioms and some deductive apparatus is presupposed or explicitly supplied. Then the theorems of the theory are whatever can be derived from the axioms by using the deductive apparatus. Similarly:

> In an *axiomatic* logical system, we adopt a set of basic logical truths as axioms, and explicitly specify a (usually very restricted) set of allowed rules of inference. A proof from some given premisses to a conclusion then has the simplest possible structure. It is just a linear sequence of sentences, each one of which is either (i) one of the premisses, or (ii) one of the logical axioms, or (iii) follows from earlier sentences in the proof by one of the rules of inference, with the whole sequence ending with the target conclusion.

Informal deductive reasoning, however, is not relentlessly linear like this. We don't require that each proposition in a proof (other than a given premiss or a logical axiom) has to follow from what's gone before. We often step sideways (so to speak) to make some new temporary assumption 'for the sake of argument'. For example, we may say 'Now suppose that P is true'; we go on to show that, given what we've already established, our supposition leads to a contradiction; we then drop or 'discharge' the temporary supposition and conclude that *not*-P. That's how a reductio ad absurdum argument works. For another example, we may again say 'Suppose that P is true'; we go on to show that we can now derive Q; we then discharge the temporary supposition and conclude that *if* P*, then* Q. That's how we often argue for a conditional proposition. Noting this,

> A *natural deduction* system of logic aims to formalize patterns of reasoning now including proofs where we argue by making and then later discharging temporary assumptions. We will need some way of keeping track of which temporary assumptions are in play and for how long. Two non-linear styles of layout are popular. Gerhard Gentzen in his doctoral thesis of 1933 introduced *trees*. And a *multi-column layout* was later popularized by Frederick Fitch in his classic 1952 logic text, *Symbolic Logic: an Introduction*.

(b) Let's have some *very* quick illustrations, starting with an example of an axiomatic system for propositional logic. In this system M, to be found e.g. in Mendelson's classic *Introduction to Mathematical Logic*, the only propositional connectives built into the basic language of the theory are '$\to$' and '$\neg$' (for 'if

... then ...' and 'not'). The axioms are then all sentences of the language which are instances of the following schemas (so you get an axiom by systematically replacing the schematic letters A, B, C with sentences – possibly quite complex ones – from our formal propositional language):

Ax1. $(A \to (B \to A))$
Ax2. $((A \to (B \to C)) \to ((A \to B) \to (A \to C)))$
Ax3. $((\neg B \to \neg A) \to ((\neg B \to A) \to B))$

M's one and only rule of inference is *modus ponens* (MP), which is the obvious rule: from A and $(A \to C)$ you can infer C.

Now, consider the inference 'if Jack missed his train, he'll be late; if he's late, we'll need to reschedule; so if Jack missed his train, we'll need to reschedule'. Evidently valid. So let's now show that we can correspondingly argue from the premisses $(\mathsf{P} \to \mathsf{Q})$ and $(\mathsf{Q} \to \mathsf{R})$ to the conclusion $(\mathsf{P} \to \mathsf{R})$ in M:

1.	$(\mathsf{P} \to \mathsf{Q})$	premiss
2.	$(\mathsf{Q} \to \mathsf{R})$	premiss
3.	$((\mathsf{Q} \to \mathsf{R}) \to (\mathsf{P} \to (\mathsf{Q} \to \mathsf{R})))$	instance of Ax1
4.	$(\mathsf{P} \to (\mathsf{Q} \to \mathsf{R}))$	from 2, 3 by MP
5.	$((\mathsf{P} \to (\mathsf{Q} \to \mathsf{R}) \to ((\mathsf{P} \to \mathsf{Q}) \to (\mathsf{P} \to \mathsf{R})))$	instance of Ax2
6.	$((\mathsf{P} \to \mathsf{Q}) \to (\mathsf{P} \to \mathsf{R}))$	from 4, 5 by MP
7.	$(\mathsf{P} \to \mathsf{R})$	from 1, 6 by MP

Which wasn't too difficult!

OK, let's try another example. The inference from the premiss $(\mathsf{P} \to (\mathsf{Q} \to \mathsf{R}))$ to the conclusion $(\mathsf{Q} \to (\mathsf{P} \to \mathsf{R}))$ is also intuitively valid – think about it! And, as we'd therefore hope, we can again derive the conclusion from the premiss in Mendelson's M. But this time the shortest M-proof I know has nineteen lines, starting of course

1. $(\mathsf{P} \to (\mathsf{Q} \to \mathsf{R}))$ premiss

and then going (I kid you not!) via

12. $(((\mathsf{Q} \to ((\mathsf{P} \to \mathsf{Q}) \to (\mathsf{P} \to \mathsf{R}))) \to ((\mathsf{Q} \to (\mathsf{P} \to \mathsf{Q})) \to (\mathsf{Q} \to (\mathsf{P} \to \mathsf{R})))) \to$
$(((\mathsf{Q} \to ((\mathsf{P} \to \mathsf{Q}) \to (\mathsf{P} \to \mathsf{R}))) \to$
$(\mathsf{Q} \to (\mathsf{P} \to \mathsf{Q}))) \to ((\mathsf{Q} \to ((\mathsf{P} \to \mathsf{Q}) \to (\mathsf{P} \to \mathsf{R}))) \to (\mathsf{Q} \to (\mathsf{P} \to \mathsf{R})))))$

which, if you look at it *very* hard and count brackets, can be seen to be an instance of the axiom schema Ax2. Even with that hint, however, it isn't exactly obvious how actually to complete the proof and arrive at

19. $(\mathsf{Q} \to (\mathsf{P} \to \mathsf{R}))$

This is rather typical of axiomatic systems (at least in their purest form): finding proofs even of simple results can be decidably unpleasant – though there are dodges for mitigating the pain somewhat in practice.

And in the case of Mendelson's system, things aren't helped either by having just the two connectives $\to$ and $\neg$ built in. For this means that M can only handle expressions involving (say) $\wedge$ – i.e. *and* – by treating these as shorthand for expressions using the two original connectives. So Mendelson treats an expression of the form $(A \wedge B)$ as short for the corresponding $\neg(A \to \neg B)$ (and that supposed equivalence is not exactly obvious). For Mendelson, therefore, the trivial inference from $(\mathsf{P} \wedge \mathsf{Q})$ to P requires a quite contorted derivation using negation and the conditional. All *very* unnatural!

(c) Let's look at our examples again, but now done in a Gentzen-style natural deduction system.

First then, arguing informally for a moment, what's a natural way of showing that, given the two premisses $(\mathsf{P} \to \mathsf{Q})$ and $(\mathsf{Q} \to \mathsf{R})$, we can infer $(\mathsf{P} \to \mathsf{R})$? We can reason in two phases.

1. We start by making an temporary additional assumption for the sake of argument; we suppose for the moment that, alongside the given premisses, P is in fact also true. Then modus ponens using the first premiss gives us Q. From that and the second premiss, another modus ponens inference give us R.

2. In short, supposing P, we can derive R. Hence, if P is true then R is, i.e. we have $(\mathsf{P} \to \mathsf{R})$.

The first phase of the argument, where the temporary assumption is in play, just uses modus ponens again. The novelty comes in the second phase were we 'discharge' that supposition. We apply a new rule of *conditional proof* (CP); this tells us that if we have a proof from the temporary supposition A (plus background assumptions) to the interim conclusion C, then we can go on to drop that supposition and infer $(A \to C)$ (from the same background assumptions).

Now let's formalize this mode of arguing. As we said, the first phase of the argument just uses modus ponens again. But rather than setting this out in a *vertical column*, Mendelson-style, let's now set things out in the form of a *branching tree*, like this:

$$\frac{\dfrac{\mathsf{P} \qquad (\mathsf{P} \to \mathsf{Q})}{\mathsf{Q}} \qquad (\mathsf{Q} \to \mathsf{R})}{\mathsf{R}}$$

A horizontal line (at least in these cases) indicates that the sentence below the line is inferred from those immediately above it – and a sentence without a horizontal line above it is an assumption (either a fixed premiss or a temporary assumption waiting to be discharged). Note that this time we don't need any commentary to tell us which earlier lines an inference step is derived from – we can read that off simply from the geometry of the proof.

We've shown then that, supposing that P, we can derive R (using some other assumptions). Hence, moving to the second phase of the argument, we can now discharge the assumption P which we made for the sake of argument (while keeping the other assumptions in play), and apply conditional proof (CP) to infer that *if* P, *then* R. We'll signal that we've discharged the assumption by enclosing it in square brackets; so applying (CP) turns the previous proof into this:

$$\dfrac{\dfrac{\dfrac{[\mathsf{P}] \qquad (\mathsf{P} \to \mathsf{Q})}{\mathsf{Q}} \qquad (\mathsf{Q} \to \mathsf{R})}{\mathsf{R}}}{(\mathsf{P} \to \mathsf{R})}$$

And now just the unbracketed sentences at the tips of branches are left as 'live' assumptions. So this is our Gentzen-style proof from the remaining premisses (P → Q) and (Q → R) to the conclusion (P → R). Already neater and more natural than the Mendelson proof.

(d) Let's show the same idea in play again. This time we will tackle that inference from (P → (Q → R)) to the conclusion (Q → (P → R)) which was horrible to derive in Mendelson's M. How can we warrant this inference, Gentzen-style?

Well, evidently the premiss (P → (Q → R)) plus the two additional suppositions P and Q entails R. We just have to use modus ponens twice in an argument we can set out like this:

$$\dfrac{\mathsf{Q} \qquad \dfrac{\mathsf{P} \qquad (\mathsf{P} \to (\mathsf{Q} \to \mathsf{R}))}{(\mathsf{Q} \to \mathsf{R})}}{\mathsf{R}}$$

Now, since P plus some other assumptions entail R, we can discharge the assumption P and derive *if* P *then* R, while keeping the other assumptions in play:

$$\dfrac{\dfrac{\mathsf{Q} \qquad \dfrac{[\mathsf{P}]^{(1)} \qquad (\mathsf{P} \to (\mathsf{Q} \to \mathsf{R}))}{(\mathsf{Q} \to \mathsf{R})}}{\mathsf{R}}}{(\mathsf{P} \to \mathsf{R})}\,(1)$$

And this time, for extra clarity, we tag both the assumption which is discharged and the corresponding inference line where the discharging takes place with matching labels '(1)'.

OK, at this point, we have used the assumption Q (with another premiss) to derive the conclusion (P → R). But we can now use CP again to discharge that assumption, this time using (2) to tag both the assumption which is discharged and the corresponding inference line where the discharging takes place.

$$\dfrac{\dfrac{\dfrac{[\mathsf{Q}]^{(2)} \qquad \dfrac{[\mathsf{P}]^{(1)} \qquad (\mathsf{P} \to (\mathsf{Q} \to \mathsf{R}))}{(\mathsf{Q} \to \mathsf{R})}}{\mathsf{R}}}{(\mathsf{P} \to \mathsf{R})}\,(1)}{(\mathsf{Q} \to (\mathsf{P} \to \mathsf{R}))}\,(2)$$

So, as we wanted, we end up with just the one live, undischarged, premiss remaining at the top of the proof-tree, and the desired conclusion at the bottom.

This is really rather elegant, with the layout clearly revealing what depends on what, and reflecting natural modes of reasoning. Or at least, that's the story! But I have perhaps already said more than enough by way of a first gesture towards axiomatic vs Gentzen-style natural deduction systems. It's time for some detailed reading suggestions . . .

5.3 Main recommendations on FOL

A preliminary reference. In my elementary logic book I carefully explain the 'design brief' for the languages of FOL, spelling out the rationale for the quantifier-variable notation. This might be helpful parallel reading when working through your chosen main text(s), at the point when that notation is introduced:

1. Peter Smith, *Introduction to Formal Logic*** (2nd edn), Chapters 26–28. Downloadable from logicmatters.net/ifl.

Unsurprisingly, there is a *very* long list of texts which cover FOL. But the point of this Guide is to choose. So here are my top recommendations, starting with one-and-a-third stand-out books which, taken together, make an excellent introduction:

2. Ian Chiswell and Wilfrid Hodges, *Mathematical Logic* (OUP 2007). This nicely written text is very approachable. It is written by mathematicians primarily for mathematicians. However, it is only one notch up in actual difficulty from some introductory texts for philosophers like mine or Nick Smith's, though – as its title might suggest – it does have a notably more mathematical 'look and feel'. It should in fact be entirely manageable for self study by philosophers and mathematicians alike (philosophers can skip over a few of the more mathematical illustrations).

   The briefest headline news is that authors explore a Gentzen-style natural deduction system. But by building things up in three stages – so after propositional logic, they consider an important fragment of first-order logic before turning to the full-strength version – they make e.g. proofs of the completeness theorem for first-order logic unusually comprehensible. For a more detailed description see my book note on C&H, [CHbooknote].

   Very warmly recommended, then. For the moment, you only *need* read up to and including §7.6 (under two hundred pages). But having got that far, you might as well read the final few sections and the Postlude too! The book has brisk solutions to some of the exercises.

Next, you should complement C&H by reading the first third of the following excellent book:

3. Christopher Leary and Lars Kristiansen's *A Friendly Introduction to Mathematical Logic*** (1st edn by Leary alone, Prentice Hall 2000; 2nd edn Milne Library 2015; and now downloadable at [friendlylogic]).

   There is a great deal to like about this book. Chs. 1–3, in either edition, do indeed make a friendly and helpful introduction to FOL. The authors use an axiomatic system, though this is done in a particularly smooth way. At this stage you could stop reading after the definition of compactness in §3.3, which means you will be reading just 87 pages.

   Unusually, L&K dive straight into a treatment of first-order logic without spending an introductory chapter or two on propositional logic: in a sense, as you will see, they let propositional logic look after itself. But this happily means (in the present context) that you won't feel that you are labouring through the very beginnings of logic one more time than is really necessary – this book therefore dovetails very nicely with C&H.

   Again written by mathematicians, some illustrations of ideas can presuppose a smattering of background mathematical knowledge; but philosophers will miss very little if they occasionally have to skip an example (and the curious can always resort to Wikipedia, which is quite reliable in this area, for explanations of some mathematical terms). The book ends with extensive answers to exercises.

   I like the overall tone of L&K very much indeed, and say more about this admirable book in another book note, [LKbooknote].

As an alternative to the C&H/L&K pairing, the following slightly more conventional book is also exceptionally approachable:

4. Derek Goldrei's *Propositional and Predicate Calculus: A Model of Argument* (Springer, 2005) is explicitly designed for self-study. Read up to the end of §6.1 (though you could skip §§4.4 and 4.5 for now, leaving them until you turn to elementary model theory).

   While C&H and the first third of L&K together cover overlapping material twice, Goldrei – in a comparable number of pages – covers very similar ground once, concentrating on a standard axiomatic proof system. So this is a relatively gently-paced book, allowing Goldrei to be more expansive about fundamentals, and to give a lot of examples and exercises with worked answers to test comprehension along the way. A great amount of thought has gone into making this text as clear and helpful as possible. Some may find it occasionally goes a bit too slowly, though I'd say that this is erring on the right side in an introductory book: if you want a comfortingly manageable text, you should find this particularly accessible. As with C&H and L&K, I like Goldrei's tone and approach a great deal.

   But since Goldrei uses an axiomatic system throughout, do eventually supplement with at least a brief glance at a Gentzen-style natural deduction proof system.

These three main recommended books, by the way, have all had very positive reports over the years from student users.

5.4 Some parallel and slightly more advanced reading

The material covered in the last section is so very fundamental, and the alternative options so very many, that I really do need to say at least something about a few other books. So in this section I list – in rough order of difficulty/sophistication – a small handful of further texts which could well make for useful parallel or additional reading. In the final section of the chapter, I will mention some other books I've been asked about.

I'll begin with a book written by a philosopher for philosophers:

5. David Bostock, *Intermediate Logic* (OUP 1997). From the preface: "The book is confined to . . . what is called first-order predicate logic, but it aims to treat this subject in very much more detail than a standard introductory text. In particular, whereas an introductory text will pursue just one style of semantics, just one method of proof, and so on, this book aims to create a wider and a deeper understanding by showing how several alternative approaches are possible, and by introducing comparisons between them." So Bostock ranges more widely than the books I've so far mentioned; he does indeed usefully introduce you to tableaux ('truth trees') *and* an Hilbert-style axiomatic proof system *and* natural deduction *and* even a sequent calculus as well. Anyone could profit from at least a quick browse of his Part II to pick up the headline news about the various approaches.

   Bostock eventually touches on issues of philosophical interest such as free logic which are not often dealt with in other books at this level. Still, the discussions mostly remain at much the same level of conceptual/mathematical difficulty as e.g. my own introductory book. He proves completeness for tableaux in particular, which I always think makes the needed construction seem particularly natural. *Intermediate Logic* should therefore be, as intended, particularly accessible to philosophers who haven't done much formal logic before and should, if read in parallel, help ease the transition to coping with the more mathematical style of the books recommended in the last section.

Note, unlike our main recommendations, Bostock does discusses tableaux ('truth trees'). If you are a philosopher, you may well have already encountered them in your introductory logic course. If not, as an alternative to Bostock,

6. My elementary introduction to truth trees for propositional logic available at [proptruthtrees] will give you the basic idea in an accessible way.

Next, even though it is giving a second bite to an author we've already met, I must mention a rather different discussion of FOL:

7. Wilfrid Hodges, 'Elementary Predicate Logic', in the *Handbook of Philosophical Logic*, Vol. 1, ed. by D. Gabbay and F. Guenthner, (Kluwer 2nd edition 2001). This is a slightly expanded version of the essay in the first edition of the *Handbook* (read that earlier version if this one isn't available), and is written with Hodges's usual enviable clarity and verve. As befits an essay aimed at philosophically minded logicians, it is full of conceptual insights, historical asides, comparisons of different ways of doing things, etc., so it very nicely complements the textbook presentations of C&H, L&K and/or Goldrei.

   Read at this stage the very illuminating first twenty short sections.

Now, as a follow up to C&H, I recommended L&K's *A Friendly Introduction* which uses an axiomatic system. As an alternative, here is an older (and, in its day, much-used) text which should certainly be very widely available:

8. Herbert Enderton, *A Mathematical Introduction to Logic* (Academic Press 1972, 2002). This also focuses on a Hilbert-style axiomatic system, and is often regarded as a classic of exposition. However, it does strike me as somewhat more difficult than L&K, so I'm not surprised that students often report finding it a bit challenging *if used by itself as a first text*. Still, it is an admirable and very reliable piece of work which you should be able to cope with well if used as a supplementary second text, e.g. after you have tackled C&H.

   Read up to and including §2.5 or §2.6 at this stage. Later, you can finish the rest of that chapter to take you a bit further into model theory. For more about this classic, see [enderlogicnote].

I can also note another much-used text which has gone through multiple editions and should again be in any library; it is a very useful natural-deduction based alternative to C&H. Later chapters of this book are also mentioned later in this Guide as possible reading for more advanced work, so it could be worth making early acquaintance with ...

9. Dirk van Dalen, *Logic and Structure* (Springer, 1980; 5th edition 2012). The early chapters up to and including §3.2 provide an introduction to FOL via Gentzen-style natural deduction. The treatment *is* often approachable and written with a relatively light touch. However, it has to be said that the book isn't without its quirks and flaws and inconsistencies of presentation (though perhaps you have to be an alert and rather pernickety reader to notice and be bothered by them). Still, having said that, the coverage and general approach is good.

   Mathematicians should be able to cope readily. I suspect, however, that the book would occasionally be tougher going for philosophers if taken from a standing start – which is another reason why I have recommended beginning with C&H instead. For more on this book, see [dalenlogic].

Next, I should certainly mention the outputs from the Open Logic Project. This is an entirely admirable, collaborative, open-source, enterprise inaugurated by Richard Zach, and very much work in progress. You can freely download the latest full version and various sampled 'remixes' from [openlogic]. In an earlier version of this Guide, I said that "although this is referred to as a textbook, it is perhaps better regarded as a set of souped-up lecture notes, written at various degrees of sophistication and with various degrees of more book-like elaboration." But things have moved on: in particular, the mix of chapters on propositional and quantificational logic in the following have been expanded and developed considerably, and are now much more book-like:

10. Richard Zach and others, *Sets, Logic, Computation*** (Open Logic). There's a lot to like here. In particular, Chapter 9 could make for very useful supplementary reading on natural deduction. Chapter 8 tells you about a sequent calculus (a slightly odd ordering!). And Chapter 10 on the completeness theorem for FOL should also prove a very useful revision guide. My sense is that overall these discussions probably will still go somewhat too briskly for some readers to work as a stand-alone introduction for initial self-study without the benefit of lecture support, which is why this doesn't feature as one of my principal recommendations in the previous section: however, your mileage may vary. And certainly, chapters from this project could/should be *very* useful for reinforcing/revision. You can download *SLC* from [slcopen].

So much, then, for reading on FOL running on more or less parallel tracks to the main recommendations in the preceding section. We continue exploring FOL semantically – the relation between theories couched in first-order languages and the structures they describe – in Chapter 7 on model theory. We later return to consider more syntactic aspects of FOL proofs in various styles in Chapter **??** on proof theory. We also consider various logics deviating from FOL in Chapter 6 and in Part II. For now, let me mention just two more books, on core FOL:

First, we go up a step in mathematical sophistication, to an absolute classic, short but packed with good things:

11. Raymond Smullyan, *First-Order Logic** (Springer 1968, Dover Publications 1995). This is terse, but those with a taste for mathematical elegance can certainly try its Parts I and II, just a hundred pages, after the initial recommended reading in the previous section. This beautiful little book is the source and inspiration of many modern treatments of logic based on tree/tableau systems. Not always easy, especially as the book progresses, but a delight for the mathematically minded.

Taking things in a new direction, don't be put off by the title of

12. Melvin Fitting, *First-Order Logic and Automated Theorem Proving* (Springer, 1990, 2nd end. 1996). This is a wonderfully lucid book by a renowned expositor. Yes, at a number of places in the book there are illustrations of how to implement various algorithms in Prolog. But either you can easily pick up the very small amount of background knowledge about Prolog that's needed to follow everything that is going on (and that's quite fun) or you can in fact just skip those implementation episodes while still getting the principal logical content of the book.

    As anyone who has tried to work inside an axiomatic system knows, proof-discovery for such systems is often hard. Which axiom schema should we instantiate with which wffs at any given stage of a proof? Natural deduction systems are nicer. But since we can, in effect, make any new temporary assumption we like at any stage in a proof, again we still need to keep our wits about us if we are to avoid going off on useless diversions. By contrast, tableau proofs (a.k.a. tree proofs) can pretty much write themselves even for quite complex FOL arguments, which is why I used to introduce formal proofs to students that way (in teaching tableaux, we can largely separate the business of getting across the idea of formality from the task of teaching heuristics of proof-discovery). And because tableau proofs very often write themselves, they are also good for automated theorem proving. Fitting explores both the tableau method and the related so-called resolution method which we mentioned as, yes, a sixth style of proof!

    This book's approach is, then, rather different from most of the other recommended books (except perhaps for Smullyan's book). However, I do think that the fresh light thrown on first-order logic makes the slight detour through this extremely clearly written book *vaut le voyage*, as the Michelin guides say. (If you don't want to take the full tour, however, there's a nice introduction to proofs by resolution in Shawn Hedman, *A First Course in Logic* (OUP 2004): §1.8, §§3.4–3.5.)

5.5 A little history (and some philosophy too)?

(a) Classical FOL is a powerful and beautiful theory. Its treatment, in one version or another, is always the first and most basic component of modern textbooks or lecture courses in mathematical logic. But how did it get this status?

The first system of formalized logic in anything like the contemporary sense – Frege's system in his *Begriffsschrift* of 1879 – allows higher-order quantification (and Frege doesn't identity FOL as a subsystem of distinctive interest). The same is true of Russell and Whitehead's logic in their *Principia Mathematica* of 1910–1913. It is not until Hilbert and Ackermann in their rather stunning short book *Mathematical Logic* (original German edition 1928) that FOL is highlighted, in a neat axiomatic system, under the label 'the restricted predicate calculus'.

(b) Those early formalized logics were all given an axiomatic presentation (though notationally very different from each other). And, as an aside, it is worth noting that the axiomatic approach reflects a broadly shared philosophical stance on the very nature of logic. Thus Frege thinks of logic as a science, in the sense of a body of truths governing a special subject matter (they are fundamental truths governing logical operations such as negation, conditionalization, quantification, identity). And in *Begriffsschrift* §13, he extols the general procedure of axiomatizing a science to reveal how a bunch of laws hang together: 'we obtain a small number of laws [the axioms] in which . . . is included, though in embryonic form, the content of all of them'. So it is not surprising that Frege takes it as appropriate to present logic axiomatically too.

In a rather different way, Russell also thought of logic as a science; he thought of it as in the business of systematizing the most general truths about the world. A special science like chemistry tells us truths about certain kinds of constituents of the world and certain of their properties; for Russell, logic tells us absolutely general truths about *everything*. If you think like *that*, treating logic as (so to speak) the most general science, then of course you'll again be inclined to regiment logic as you do other scientific theories, ideally by laying down a few 'basic laws' and then showing that other general truths follow.

Famously, Wittgenstein in the *Tractatus* reacted radically against Russell's conception of logic. For him, logical truths are *tautologies*. They are not deep ultimate truths about the most general, logical, structure of the universe; rather they are *empty* claims in the sense that they tell us nothing informative about how the world is: they merely fall out as byproducts of the meanings of the basic logical particles.

That last idea can be developed in more than one way. But one approach is Gentzen's in the 1930s. He thought of the logical connectives as getting their meanings from how they are used in inference (so grasping their meaning involves grasping the inference rules governing their use). For example, grasping 'and' involves grasping, inter alia, that from A *and* B you can (of course!) derive A. Similarly, grasping the conditional involves grasping, inter alia, that a derivation of the conclusion C from the temporary supposition A warrants an assertion of *if* A *then* C. Suppose then that A *and* B; then we can derive A (by one of the rules for 'and'). And reflecting on that little suppositional inference, we see that the rule of inference which partly gives the meaning of 'if' entitles us to assert *if* A *and* B, *then* A. And hence those logical inference rules enable us to derive that logical truth 'for free' (from no remaining assumptions).

If that is right, and if the point generalizes, then we don't have to see such logical truths as reflecting deep facts about the logical structure of the world (whatever that could mean): logical truth falls out just as a byproduct of the inference rules whose applicability is, in some sense, built into the very meaning of e.g. the connectives and the quantifiers.

It is a nice question how far we should buy that sort of story about the nature of logical truth. But whatever your judgement on that, there surely *is* something odd about thinking with Frege and Russell that a systematized logic as primarily aiming to regiment a special class of ultra-general *truths*. Isn't logic at bottom about good and bad reasoning practices, about what makes for a good proof? Shouldn't its prime concern be the correct styles of valid inference? So shouldn't a formalized logic highlight *rules of proof-building* (perhaps in natural deduction style) rather than stressing *axiomatic truths*?

(c) Back to the history of the technical development of logic. Here are three suggestions for reading:

13. An obvious starting place is the following clear and judicious *Stanford Encyclopaedia* article: William Ewald, 'The Emergence of First-Order Logic'** (*SEP*, 2018). Available at [emergenceFOL]
14. If you want rather more, José Ferreiros's 'The Road to Modern Logic – an Interpretation'**, *Bulletin of Symbolic Logic* 7 (2001): 441–484, is also readable and very helpful. Available at [roadtologic].
15. For a longer, and rather bumpier, read – you'll probably need to skim and skip! – you could also try dipping into the more wide-ranging piece by Paolo Mancosu, Richard Zach and Calixto Badesa, 'The Development of Mathematical Logic from Russell to Tarski: 1900–1935'** in Leila Haaparanta, ed., *The History of Modern Logic* (OUP, 2009, pp. 318–471). Available at [developlogic].

5.6 Postscript: Other treatments?

I will end this chapter by responding to a variety of Frequently Asked Questions, mostly questions raised in response to earlier versions of the Guide. Occasionally, I have to be pretty negative.

A blast from the past: What about Mendelson? Somewhat to my surprise, perhaps the most frequent question I used to get asked in response to early versions of the Guide is 'But what about Mendelson, Chs. 1 and 2'? Well, Elliott Mendelson's *Introduction to Mathematical Logic* (Chapman and Hall/CRC, 6th edn. 2015) was first published when I was a student and the world was a great deal younger. The book was I think the first modern textbook of its type (so immense credit to Mendelson for that), and I no doubt owe my whole career to it – it got me through tripos!

It seems that some others who learnt using the book are in their turn still using it to teach from. But let's not get too sentimental! It has to be said that the book in its first incarnation was often brisk to the point of unfriendliness, and the basic look-and-feel of the book hasn't changed a great deal as it has run through successive editions. Mendelson's presentation of axiomatic systems of logic are quite tough going, and as the book progresses in later chapters through formal number theory and set theory, things if anything get somewhat less reader-friendly. Which certainly doesn't mean the book won't repay battling with. But unsurprisingly, fifty years on, there are many rather more accessible and more amiable alternatives for beginning serious logic. Mendelson's book is a landmark worth visiting one day, but I can't recommend starting there. For a little more about it, [mendelsonlogic].

(If you *do* want an old-school introduction from the same era, you might more enjoy Geoffrey Hunter, *Metalogic** (Macmillan 1971, University of California Press 1992). This is not groundbreaking like Smullyan's *First-Order Logic* is, nor is it as comprehensive as Mendelson: but it was an exceptionally good textbook from a time when there were few to choose from. Read Parts One to Three at this stage. And if you are finding it rewarding reading, then do eventually finish the book: it goes on to consider formal arithmetic and proves the undecidability of first-order logic, topics we consider in Chapter 8. Unfortunately, the typography – from pre-LaTeX days – isn't at all pretty to look at: this can make the book's pages initially appear rather unappealing. But in fact the treatment of an axiomatic system of logic is extremely clear and accessible. It might be worth blowing the dust off your library's copy!)

Five more recent mathematical logic texts: What about Ebbinghaus, Flum and Thomas? Hedman? Hinman? Rautenberg? Kaye? We start with H.-D. Ebbinghaus, J. Flum and W. Thomas, *Mathematical Logic* (Springer, 2nd edn. 1994). This is the English translation of a book first published in German in 1978, and appears in a series 'Undergraduate Texts in Mathematics', which indicates the intended level. The book is often warmly praised and is (I believe) quite widely used – and there is a lot of material here, often covered well. But revisiting the book, I can't find myself wanting to recommend it as a good place to *start*, either for philosophers or for mathematicians. The core material on the syntax and semantics of first-order logic in Chs 2 and 3 is presented more accessibly and more elegantly elsewhere. And the treatment of a sequent calculus Ch. 4 strikes me as poor, with the authors (by my lights) mangling some issues of principle and certainly failing to capture the elegance that a sequent calculus can have. For more on this book, see my [EFTbooknote].

Shawn Hedman's *A First Course in Logic* (OUP, 2004) is subtitled 'An Introduction to Model Theory, Proof Theory, Computability and Complexity'. So there is no lack of ambition in the coverage! The treatment of basic FOL is patchy. It is pretty clear on semantics, and the book can be recommended to more mathematical readers for its treatment of more advanced model-theoretic topics (see §7.3 in this Guide). But Hedman offers a peculiarly ugly not-so-natural deductive system. As already noted, however, he is good on so-called resolution proofs. For more about what does and what doesn't work in Hedman's book, see [hedmanbook].

Peter Hinman's *Fundamentals of Mathematical Logic* (A. K. Peters, 2005) is a massive 878 pages, and as you'd expect covers a great deal. Hinman is, however, not really focused on deductive systems for logic, which don't make an appearance until over two hundred pages into the book. This is not, then, the place to start with FOL, though the book contains some useful supplementary material once you have got hold of the basics from elsewhere. I do think, however, that most readers will find Hinman pretty tough going. For more about what does and what doesn't work in his book, see [hinmanbook].

Wolfgang Rautenberg's *A Concise Introduction to Mathematical Logic* (Springer, 2nd edn. 2006) has some nice touches. But its first hundred pages on FOL are rather *too* concise to serve most readers as a first introduction; and its preferred formal system is not a 'best buy' either. Good revision material, though.

Finally, Richard Kaye wrote an attractively written 1991 classic on models of Peano Arithmetic (we will meet this in Chapter 8). So I had high hopes for his later *The Mathematics of Logic* (CUP 2007). "This book", he writes, "presents the material usually treated in a first course in logic, but in a way that should appeal to a suspicious mathematician wanting to see some genuine mathematical applications. ... The main goal is an understanding of the mathematical content of the Completeness Theorem for first-order logic, including some of its mathematically more interesting applications. ... This book is unusual, however, since I do not present the main concepts and goals of first-order logic straight away. Instead, I start by showing what the main mathematical idea of 'a completeness theorem' is, with some illustrations that have real mathematical content." So the reader is taken on a mathematical journey starting with König's Lemma (I'm not going to explain that here!), and progressing via order relations, Zorn's Lemma (an equivalent to the Axiom of Choice), Boolean algebras, and propositional logic, to completeness and compactness of first-order logic. Does this unusual route work? I am not at all convinced. It seems to me that the journey is made too bumpy and the road taken is too uneven in level for this to be appealing as an early trip through first-order logic. But if you already know a fair amount of this material from more conventional presentations, the different angle of approach in this book could be interesting and illuminating.

What about a reasonably introductory text with a more proof-theoretic slant? Well, it's true that there are some questions about systems of FOL which can be tackled at a quite introductory level, yet which aren't addressed by any of the readings so far mentioned. For a simple example, suppose we are working from given premisses in a formal proof system and (i) have so far derived A and also derived B; then (ii) we can (rather boringly!) infer the conjunction $A \wedge B$ (remember '$\wedge$' means *and*). Now, suppose later in the same proof (iii) we appeal to that conjunction $A \wedge B$ to derive A. We wouldn't have gone *wrong*; but obviously we have gone on a pointless detour, given that at stage (i) we have already derived A. There is evident interest in the question of how to eliminate such detours and other pointless digressions from proofs. Gentzen famously started the ball rolling in his discussions of how to 'normalize' proofs in his natural deduction systems, and he showed how normalization results can be used to derive other important properties of the proof systems.

For a first encounter with this sort of topic, you could look at Jan von Plato's *Elements of Logical Reasoning** (CUP, 2014). This is based on the author's lectures for a general introductory FOL course. But a lot of material is touched on in a relatively short compass as von Plato talks about a range of different natural deduction and sequent calculi; I suspect that, without prior knowledge of ND and without classroom work to round things out, this book might not be as accessible as the author intends. Still, if you want to know about variant ways of setting up ND systems, about proof-search, about the relation with so-called sequent calculi, etc., then the first half of this book makes a reasonably clear start. However, my own recommendation would be to concentrate for the moment on the core topics in FOL covered by the books we have mentioned previously, and then later dive into proof theory proper, covered in Chapter **??**.

Designed for philosophers: Why not The Logic Book? What about Sider? What about Bell, DeVidi and Solomon? Many US philosophers have had to take courses based on *The Logic Book* by Merrie Bergmann, James Moor and Jack Nelson (first published by McGraw Hill in 1980; a sixth edition was published – at a quite ludicrous price – in 2013). I doubt that those students much enjoyed the experience! This is a large book, over 550 pages, starting at about the level of my introductory book, and going as far as metalogical results like a full completeness proof for FOL, so its coverage overlaps with the main recommendations of §5.3. But while reliable enough, it all strikes me, like some other readers who have commented, as *very* dull and laboured, and often rather unnecessarily hard going. You can certainly do better.

Theodore Sider – a very well-known philosopher – has written a text called *Logic for Philosophy** (OUP, 2010) which I've repeatedly been asked to comment on. The book in fact falls into two halves. The second half (about 130 pages) is on modal logic, and I will return to that in Chapter **??**. The first half of the book (almost exactly the same length) is on propositional and first-order logic, together with some variant logics, so is very much on the topic of this chapter. But while the coverage of modal logic is quite good, I can't at all recommend the first half of this book: I explain why in a book note, [siderbook].

A potential alternative to Bostock at about the same level, and which can initially look promising, is John L. Bell, David DeVidi and Graham Solomon's *Logical Options: An Introduction to Classical and Alternative Logics* (Broadview Press 2001). This book covers a lot pretty snappily – for the moment, just Chapters 1 and 2 are relevant – and some years ago I used it as a text for second-year seminar for undergraduates who had used my own tree-based book for their first year course. But many students found the exposition too terse, and I found myself having to write very extensive seminar notes. If you want some breadth, you'd do better sticking with the more expansive Bostock.

But I don't want to finish on a negative note: so finally ...

Puzzles galore: What about some of Smullyan's other books? I have already warmly recommended Smullyan's 1968 classic *First-Order Logic*. He went on to write some classic texts on Gödel's theorem and on recursive functions, which we'll be mentioning later. But as well as these, Smullyan wrote many 'puzzle' based-books aimed at a wider audience, including the justly famous *What is the Name of This Book?** (Dover Publications reprint of 1981 original, 2011).

More recently, he wrote *Logical Labyrinths* (A. K. Peters, 2009). From the blurb: “This book features a unique approach to the teaching of mathematical logic by putting it in the context of the puzzles and paradoxes of common language and rational thought. It serves as a bridge from the author’s puzzle books to his technical writing in the fascinating field of mathematical logic. Using the logic of lying and truth-telling, the author introduces the readers to informal reasoning preparing them for the formal study of symbolic logic, from propositional logic to first-order logic, … The book includes a journey through the amazing labyrinths of infinity, which have stirred the imagination of mankind as much, if not more, than any other subject.” Smullyan starts, then, with puzzles, e.g. of this kind: you are visiting an island where there are Knights (truth-tellers) and Knaves (persistent liars) and then in various scenarios you have to work out what’s true from what the inhabitants say about each other and the world. And, without too many big leaps, he ends with first-order logic (using tableaux), completeness, compactness and more. This is no substitute for standard texts, but – for those with a taste for being led up to the serious stuff via sequences of puzzles – an entertaining and illuminating supplement.

Smullyan’s later *A Beginner’s Guide to Mathematical Logic** (Dover Publications, 2014) is more conventional. The first 170 pages are relevant to FOL. A rather uneven read, it seems to me; but again an engaging supplement to the main texts recommended above.

6 Second-order logic, briefly

Classical first-order logic contrasts along one dimension with various non-classical logics, and along another dimension with second-order and higher-order logics. We'll leave the exploration of non-classical logics to Part II of the Guide. But we will say something about second-order logics in this chapter.

Theories expressed in first-order languages with a first-order logic have their limitations. That's a theme that will recur when we look at model theory (Chapter 7), at theories of arithmetic (Chapter 8), and at set theory (Chapter 9); and you will occasionally find contrasts being drawn with theories expressed in second-order languages with a second-order logic. So, although it is a judgement call, I think it is worth knowing something early on about second-order logic – if only the very little which is sketched in the overview here. Then you can either wait to get a proper grip on second-order logic later, as and when needed, or you can read up somewhat more in advance now.

6.1 A preliminary note on many-sorted logic

(a) As you will have seen, FOL is standardly presented as having a single 'sort' of quantifier, in the sense that all the quantifiers in a given language run over one and the same domain of objects. But this is artificial, and certainly doesn't conform to everyday mathematical practice.

To take an example which will be very familiar to mathematicians, consider the usual practice of using one style of variable for scalars and another for vectors, as in the rule for scalar multiplication:

(1) $$a(\mathbf{v}_1 + \mathbf{v}_2) = a\mathbf{v}_1 + a\mathbf{v}_2.$$

If we want to make the implied generality here explicit, we could very naturally write

(2) $$\forall a \forall \mathbf{v}_1 \forall \mathbf{v}_2 (\mathbf{v}_1 + \mathbf{v}_2) = a\mathbf{v}_1 + a\mathbf{v}_2,$$

with the first quantifier running just over scalars, and the other quantifiers running just over vectors. It seems odd, then, to insist that, if we want to formalize our theory of vector spaces, we should follow FOL practice and use only *one* sort of variable and so render the rule for scalar multiplication along the lines of

(3) $$\forall \mathsf{x} \forall \mathsf{y} \forall \mathsf{z} ((\mathsf{S}\mathsf{x} \land \mathsf{V}\mathsf{y} \land \mathsf{V}\mathsf{z}) \to \mathsf{x}(\mathsf{y} + \mathsf{z}) = \mathsf{x}\mathsf{y} + \mathsf{x}\mathsf{z}),$$

i.e. 'Take any three things in our domain, if the first is a scalar, the second is a vector, and the third is a vector, then ...'.

In sum, the theory of vector spaces is naturally regimented using a *two-sorted* logic, with two sorts of variables running over two different domains.

(b) Generalizing, why not allow a *many-sorted* logic – allowing multiple independent domains of objects, with different sorts of variables restricted to running over the different domains? And it isn't hard to set up such a revised version of FOL (it *is* still first-order, as the quantifiers are still of the familiar general type, running over objects in the relevant domains). The syntax and semantics of a many-sorted language can be defined quite easily. Syntactically, we will just need to keep a tally of the sorts assigned to the various names and variables, and we will also need rules about which sorts of terms can go into which slots in predicates and in function-expressions. Semantically, we assign a domain for each sort of variable, and then proceed pretty much as in the one-sorted case. Assuming that each domain is non-empty (as in standard FOL) the inference rules for a deductive system will then look entirely familiar. And the resulting logic has the same nice technical properties as standard FOL; crucially, you can prove soundness, completeness and compactness theorems in just the same ways.

(c) As so often in the formalization game, we are now faced with a cost/benefit trade-off. We can get the benefit of somewhat more natural regimentations of mathematical practice, at the cost of having to use a slightly more complex many-sorted logic. Or we can pay the price of having to use less natural regimentations – we systematically translate propositions like (2) into restricted quantifications like (3) – and get the benefit of a simpler-in-practice logic.[1]

So you pays your money and you takes your choice. For most purposes, logicians take the second option and stick to standard FOL. That's because at the end of the day they care less about elegance when regimenting this or that theory and more about having a simple-but-powerful logical system to work with.

6.2 Second-order logic: an overview

(a) Now we turn from 'sorts' to 'orders'. It will help to fix ideas if we begin with an easy arithmetical example; so consider the informal principle of induction:

(1) Take *any* numerical property X; if (i) zero has X and (ii) any number which has X passes it on to its successor, then (iii) *all* numbers must share property X.

This holds, of course, because every natural number is either zero or is an eventual successor of zero (i.e. is either 0 or $S0$ or $SS0$ or $SSS0$ or ...): there are

[1] We also get some added flexibility on the second option. The use of a sorted quantifier $\forall\alpha\mathsf{F}\alpha$ with the usual logic presupposes that there is at least one thing in the relevant domain, while a corresponding restricted quantification $\forall\mathsf{x}(\mathsf{Ax} \to \mathsf{Fx})$, where A picks out the relevant sort which α is supposed to run over, leaves open the possibility that there is nothing of that sort.

no stray numbers outside that sequence, so a property that percolates down the sequence eventually applies to each and any number.

There is no problem about expressing some particular *instances* of the induction principle in a first-order language. For example, suppose P is a formal one-place predicate expressing an arithmetical property: then we can write

(2) $$(\mathsf{P0} \land \forall \mathsf{x}(\mathsf{Px} \to \mathsf{Psx})) \to \forall \mathsf{x}\, \mathsf{Px}$$

where the small-'x' quantifier runs over the natural numbers and 's' expresses the successor function. But how can we state the *general* principle of induction in a formal language? The natural candidate is something like this:

(3) $$\forall \mathsf{X}((\mathsf{X0} \land \forall \mathsf{x}(\mathsf{Xx} \to \mathsf{Xsx})) \to \forall \mathsf{x}\, \mathsf{Xx}).$$

Here the big-'X' quantifier is a new type of quantifier, which unlike the small-'x' quantifier, quantifies 'into predicate position'. In other words, it quantifies into the position occupied in (2) by the predicate 'P', and is intended to run over all *properties* of numbers, so that (3) indeed formally renders (1). But this kind of quantification – second-order quantification – is not available in standard first-order languages of the kind that you now know and love.

If we want to stick with an arithmetic theory framed in a first-order language L which just quantifies over numbers, the best we can do is to use a template or schema and say something like

(4) For any arithmetical L-predicate $\varphi(\;)$, simple or complex, the corresponding sentence $(\varphi(\mathsf{0}) \land \forall \mathsf{x}(\varphi(\mathsf{x}) \to \varphi(\mathsf{sx})) \to \forall \mathsf{x}\, \varphi(\mathsf{x})$ is an axiom.

However (4) is much weaker than the informal (1) or the equivalent formal version (3) on its intended interpretation. For (1/3) tells us that induction holds for *any property at all*; while, in effect, (4) only tells us that induction holds for *those properties that can be expressed by some L-predicate $\varphi(\;)$*.

Something like (3), then, is the natural way of formalizing the full principle of induction. So why not allow second-order quantification, quantification into predicate position?

(b) Again, it isn't difficult to extend the syntax and semantics of first-order languages to allow for second-order quantification.

The required added *syntax* is unproblematic. Recall how we can take a formula $\varphi(\mathsf{n})$ containing some occurrence(s) of the name 'n', swap out the name on each occurrence for a particular (small) variable, and then form a first-order quantified wff like $\forall \mathsf{x}\varphi(\mathsf{x})$. We just need now to add the rule that we can take a formula $\varphi(\mathsf{P})$ containing some occurrence(s) of the unary predicate 'P', swap out the predicate for some (big) variable and then form a second-order quantified wff like $\forall \mathsf{X}\varphi(\mathsf{X})$. Fine print apart, that's straightforward.

The standard *semantics* is equally straightforward. Again we model the story about the second-order quantifiers on the account of first-order quantifiers. So fix a domain of quantification. Then recall that $\forall \mathsf{x}\varphi(\mathsf{x})$ is true on a given interpretation of its language just when $\varphi(\mathsf{n})$ remains true, however we vary the object in the domain which is assigned to the name 'n' as its interpretation. Similarly $\forall \mathsf{X}\varphi(\mathsf{X})$ is true on an interpretation just when $\varphi(\mathsf{P})$ remains true, however we vary the subset of the domain which is assigned to the predicate 'P' as *its* interpretation (i.e. its extension). Again, there's fine print; but you get the general idea.

We'll want to expand the syntactic and semantic stories further to allow second-order quantification over binary and other relations and over functions too; but these expansions raise no new issues.

We can then define the relation of semantic consequence for formulas in our extended languages including second-order quantifiers in the now familiar way: some formulas Γ semantically entail φ just in case every interpretation that makes all of Γ true makes φ true.

(c) So the situation is this. There are quite a few familiar mathematical claims which, like the arithmetical induction principle, are naturally regimented using quantifications over properties (and/or relations and/or functions). And there is no problem about augmenting the syntax and semantics of our formal languages to allow such second-order quantifications, and we can carry over the definition of semantic entailment to cover sentences in the resulting second-order languages.

Moreover, theories framed in second-order languages turn out to have nice properties which are lacked by their first-order counterparts. For example, a theory of arithmetic with the full second-order induction principle (3) will be 'categorical', in the sense of having just one kind of structure as a model (a model built from a zero, its eventual successors, and nothing else). On the other hand, a first-order theory of arithmetic which has to rely on a limited induction principle like (4) will have models of quite different kinds (as well as the intended model with just a zero and its eventual successors, there will be an infinite number of different 'non-standard' models which have unwanted junk in their domains).

The obvious question which arises is *why have we followed the standard modern practice of privileging FOL*? Why not adopt a second-order logic from the outset as our preferred framework for regimenting mathematical arguments? – after all, as noted in §5.5, early formal logics like Frege's allowed more than first-order quantifiers.

(d) The short answer is: *because there can be no sound and complete formal deductive system for second-order logic.*

There can be be sound but partial deductive systems S for a language including second-order quantifiers. So we have the one-way conditional that, whenever there is an S-proof from premisses in Γ to the conclusion φ, then Γ indeed semantically entails φ. But the converse fails. We can't have a respectable formal system S (where it is decidable what's a proof, etc.) such that, whenever Γ semantically entails φ, there is an S-proof from premisses in Γ to the conclusion φ. In other words, once second-order sentences are in play, we can't fully capture the relation of semantic entailment in a formal deductive system.

(e) It is helpful to contrast the case of a two-sorted first-order language of the kind we met in the previous section. In that case, the two sorts of quantifier get interpreted independently – fixing the domain of one doesn't fix the domain of the other. And it is because each sort of quantifier, as it were, stands alone, a familiar kind of first-order logic continues to each seperately.

But in second-order logic it is quite different. For note that on the standard semantic story, it is now the *same* domain which fixes the intepretation of both kinds of quantifier – i.e. one and the same domain both provides the objects for the first-order quantifiers to range over, and also provides the sets of objects (i.e. *all* the subsets of the original domain) for the second-order quantifiers to range over. The interpretations of the two kinds of quantifier are tightly connected, and this makes all the difference, and blocks the possibility of a complete deductive system for second-order logic.

(Technical note: If we drop the requirement of 'standard' or 'full' semantics that the second-order big-'X' quantifiers run over exactly *all* the subsets of the domain of the corresponding first-order small-'x' quantifiers, we will arrive what's called 'Henkin semantics' or 'general semantics'. And on this semantics we can regain a completeness theorem, but we lose the nice other features that second-order theories have on their natural 'standard' semantics.)

(f) It's of course not supposed to be *obvious* that we can't have a complete deductive system for second-order logic on the standard semantics, any more than it is obvious that we *can* have a complete deductive system for first-order logic![2] And it isn't obvious either what the significance this technical result might be. In fact, the whole question of the status of second-order logic leads to a tangled debate.

Let's briefly touch on one thread of the debate. On the usual story, when we give the semantics of FOL, we interpret one-place predicates by assigning them *sets* as extensions. And when we now add second-order quantifiers, we are adding quantifiers which are correspondingly interpreted as ranging over all these possible extensions. So, you might well ask, why not frankly rewrite (3), for example, in the form

(5) $$\forall \mathsf{X}((0 \in \mathsf{X} \land \forall \mathsf{x}(\mathsf{x} \in \mathsf{X} \to \mathsf{sx} \in \mathsf{X}) \to \forall \mathsf{x}\, \mathsf{x} \in \mathsf{X}),$$

[2] Let's outline the easy proof that strong completeness fails in the second-order case (the proof that weak completeness fails is harder). It is enough to show that compactness fails (compare §5.1, fn.4).

First recall that, for any finite n, we can form a (first-order) sentence we'll abbreviate $\exists n$ which is true if and only if the domain contains at least n objects.

Now, we can also form a second-order sentence $\exists\infty$ which is true in all and only those structures which have an infinite domain. Roughly, we write down a sentence saying that there is a binary relation (second-order quantification!) which relates each object to a distinct one, in a way that gives us an unending chain of objects which never repeats.

Then we consider the infinite set of sentences

$$\Gamma =_{\mathrm{def}} \{\exists 1, \exists 2, \exists 3, \exists 4, \ldots, \neg\exists\infty\}$$

Any *finite* subset $\Delta \subset \Gamma$ has a model (because there will be a maximum number n such that $\exists n$ is in Δ – and then all the sentences in Δ, which might include $\neg\exists\infty$, will be true in a structure whose domain contains exactly n objects). Compactness would then imply that Γ has a model. But that's impossible. No structure can have a domain which both *does* have at least n objects for every n and also *doesn't* have infinitely many objects.

making it explicit that the big-'X' variable is running over sets? Well, we *can* do that. Though if (5) is to replicate the content of (3) on its standard semantics, it is crucial that the big-'X' variable has to run over *all* the subsets of the domain of the small-'x' variable.

And now some would say that, because (3) can be rewritten as (5), this just goes to show that in using second-order quantifiers we are straying into the realm of set theory. But others would push the connection in the other direction. They would start by arguing that the invocation of sets in the explanation of second-order semantics, while conventional, is actually dispensable (in the spirit of §4.3; see the papers by Boolos mentioned in the next section). So this means that (5) in fact dresses up an induction principle (3) which is not in essence set-theoretic in misleadingly fancy clothing.

So we are left with a troublesome question: is second-order logic really just some "set theory in sheep's clothing" (as the philosopher W.V.O. Quine famously quipped)? We can't pursue this further here (though I give some pointers in the next section for philosophers who want to tackle the issue). Fortunately, for the purposes of getting to grips with the logical material of the next few chapters, you just need to grasp a few basic technical facts about second-order logic, and in particular note that – as announced – while second-order theories can have nice properties, their logic escapes being captured in a formal deductive system.

6.3 Recommendations on many-sorted and second-order logic

If you want to know more about the formal details of many-sorted first-order languages and their logic, what little you need is covered in four clear pages by

1. Herbert Enderton, *A Mathematical Introduction to Logic* (Academic Press 1972, 2002), §4.3.

So let's turn straight to second-order logic.

For a brief review, saying only a little more than the overview in the last section, see

2. Richard Zach and others, *Sets, Logic, Computation*** (Open Logic) §11.3, excerpted at [openlogicSOL].

You could then look e.g. at the rest of Chapter 4 of the Enderton book we just mentioned. Or better, read

3. Stewart Shapiro, 'Higher-order Logic', in S. Shapiro, ed., *The Oxford Handbook of the Philosophy of Mathematics and Logic* (OUP, 2005).

You can skip §3.3; but §3.4 touches on Boolos's ideas and is relevant to the question of how far second-order logic presupposes set theory. Shapiro's §5, 'Logical choice', is an interesting discussion of what's at stake in adopting a second-order logic.

And that's probably all you need at this stage (and don't worry if some points will only become clearer when you've done some model theory and some formal arithmetic). However, to nail down some of the technical basics you can very usefully supplement the explanations in Shapiro with the admirably clear

4. Tim Button and Sean Walsh, *Philosophy and Model Theory** (OUP, 2018), Chapter 1.

This chapter reviews, in a particularly helpful way, first-order syntax and various ways of developing its semantics; and then it compares the first-order case with second-order semantics, both 'full' semantics and 'Henkin' semantics.

If these readings leave you still wanting to fill out the technical story about second-order logic a little further, you will then want to dive into the self-recommending

5. Stewart Shapiro, *Foundations without Foundationalism: A Case for Second-Order Logic*, Oxford Logic Guides 17 (Clarendon Press, 1991), Chs. 3–5 (with Ch. 6 for enthusiasts).

And philosophers who have Shapiro's wonderfully illuminating book in their hands, will also be intrigued by the initial philosophical/methodological discussion in his first two chapters here. This whole book is a modern classic, remarkably accessible, and important too for the contrasting side-light it throws on FOL.

Shapiro – in both his *Handbook* essay and in his earlier book – mentions Boolos's arguments against treating second-order logic as essentially set-theoretical. Just because he is so readable, let me mention the thought-provoking

6. George Boolos, 'On Second Order Logic' and 'To Be is to Be a Value of a Variable (or to Be Some Values of Some Variables)', both reprinted in his wonderful collection of essays *Logic, Logic, and Logic* (Harvard UP, 1998).

You can then follow up some of the critical discussions of Boolos mentioned by Shapiro.

Finally, if you are left wanting still more, there is an up-to-the-minute sophisticated review of second order logic with many further pointers to an extensive literature (but going way beyond the basics) here:

7. Jouko Väänänen, 'Second-order and Higher-order Logic', *Stanford Encyclopedia of Philosophy*, available at [sep-vaan].

7 Beginning model theory

The high point of a first serious encounter with FOL is the proof of the completeness theorem. Introductory mathematical logic texts then usually discuss at least a couple of quick corollaries of the proof – the *compactness theorem* and the *downward Löwenheim-Skolem theorem*. And so we take initial steps into what we can call Level 1 model theory. Further along the track we will encounter Level 3 model theory (I am thinking of the sort of topics covered in e.g. the later chapters of the now classic texts by Wilfrid Hodges and David Marker). In between, there is a stretch of what we can think of as Level 2 theory – still relatively elementary, relatively accessible without too many hard scrambles, but going somewhat beyond the very basics.

Putting it like this in terms of 'levels' is of course only for the purposes of rough-and-ready organization: there are no sharp boundaries to be drawn. In a first mathematical logic course, though, you should certainly get your head around Level 1 model theory. Then tackle as much Level 2 theory as grabs your interest. Level 3 topics are definitely more advanced: so we leave them to Part III of the Guide. But what topics can we assign to the first two levels?

7.1 Elementary model theory: an overview

(a) Model theory is about mathematical structures and about how to characterize and classify them using formal languages. Put it another way, it concerns the relationship between a mathematical theory (regimented as a collection of formal sentences) and the structures which 'realize' that theory (i.e. the structures which we can interpret the theory as being true of, i.e. the structures which provide a model for the theory).

It will help to be have in mind a sample range of theories and corresponding structures. For example, it is good to know just a little about theories of arithmetic, algebraic theories (like the theory of groups or of Boolean algebra), theories of various kinds of order, etc., and to know just a little about some of the structures which provide models for these theories. Mathematicians will already be familiar with informally presented examples: philosophers will probably need to do a small amount of preparatory homework here (but the first reading recommendation in the next section should provide enough to start you off).

Here are some initial themes we'll need to explore:

(1) We'll be interested in relations between structures. Most simply, one structure can be a substructure of another, or can extend another. Or we can map one structure to another in a way that preserves structural information – so, for example, a structure-preserving map can send one structure to a copy embedded inside another structure. In particular, we will be interested in the case where there's an isomorphism between structures, so that each is a replica of the other (as far as their structural features are concerned).

We will similarly be interested in relations between languages for describing structures – we can expand or reduce the non-logical resources of a language, potentially giving it greater or lesser expressive power. So we will want to know something about the interplay between these expansions and reductions of structures and corresponding languages.

(2) How much can a language tell us about a structure? For a toy example, take the structure $(\mathbb{N}, <)$, i.e. the natural numbers equipped with their standard order relation. And consider the first-order formal language whose sole bit of non-logical vocabulary is a symbol for the order relation (let's re-use $<$ for this, with context making it clear that this now *is* an expression belonging to a formal language!). Then, note that we can e.g. define the successor relation over $\mathbb{N}$ in this language, using the formula

$$\mathsf{x} < \mathsf{y} \land \forall \mathsf{z}(\mathsf{x} < \mathsf{z} \to (\mathsf{z} = \mathsf{y} \lor \mathsf{y} < \mathsf{z}))$$

For evidently a pair of numbers x, y satisfies this formula if y comes immediately after x in the ordering. And given we can define the successor relation, we can now e.g. define 0 as the number in the structure $(\mathbb{N}, <)$ which isn't a successor of anything.

Now contrast the structure $(\mathbb{Z}, <)$, i.e. the integers equipped with *their* standard order relation, with the corresponding formal language where $<$ gets re-interpreted accordingly. The same formula as before still suffices to define the successor relation over $\mathbb{Z}$. But this time, we obviously can't define 0 as the integer which isn't a successor (all integers are successors!). And in fact no other expression from the formal language whose sole bit of non-logical vocabulary is the order-predicate $<$ will define the zero in $(\mathbb{Z}, <)$ – in other word, the ordering relation gives only the relative position of integers, but doesn't fix the zero.

OK, those were indeed toy examples! But they illustrate a very important class of questions of the following form: which objects and relations in a particular structure can be pinned down, which can be defined, using expressions from a first-order language for the structure?

(3) Moving from what can be defined by particular expressions to the question of what gets fixed by a whole theory, we can ask how varied the models of a given theory can be. At one extreme, we have a theory like group theory, whose models are wildly various. At the other extreme, a theory like second-order Peano Arithmetic is *categorical* – its models will all 'look

the same', i.e. are all isomorphic with each other. Categoricity is good when we can get it: but when is it available? We'll return to this in a moment.

(4) Instead of going from a theory to the structures which are its models, we can go from structures to theories. Given a class of structures, we can ask: is there a first-order theory for which just these structures are the models? Or given a particular structure, and a corresponding language with appropriate names, predicates and functional expressions), we look at the set of all the sentences in the language which are true of the structure. We can now ask, when can all those sentences be regimented into a nicely axiomatized theory? Perhaps we can find a finite collection of axioms which entails all those truths about the structure: or if a finite set of axioms is too much to hope for, perhaps we can at least get a set of axioms which are nicely disciplined in some other way. And when is the theory for a structure (the set of sentences it makes true) decidable, in the sense that a computer could work out what sentences are truths of the theory?

(b) Now, you have already met a pair of fundamental results linking semantic structures and sets of first-order sentences – the soundness and completeness theorems. And these lead to a pair of fundamental model-theoretic results. The first of these we've met before, at end of §5.1:

(5) *The finiteness or compactness theorem.* If every finite subset of Γ has a model, so does Γ.

For our second result, revisit a standard completeness proof for FOL, which shows that any syntactically consistent set of sentences from a first-order language (set of sentences from which you can't derive a contradiction) has a model. Looking at the details of the proof, and assuming that we are dealing with normal first-order languages (with a countable vocabulary), you'll find that it in fact establishes the stronger result that a syntactically consistent set of sentences will have a *countable* model – in effect, it has a model whose domain is just (some or all) the natural numbers. Which gives us

(6) *The downward Löwenheim-Skolem theorem.* Suppose a bunch of sentences Γ from a first-order language L has a model (however large); then Γ has a countable model.

Why so? Suppose Γ has a model. Then it is syntactically consistent in your favoured proof system (for if we could derive absurdity from Γ then, by the soundness theorem, Γ would semantically entail absurdity, i.e. would be semantically inconsistent after all and have no model). And since Γ is syntactically consistent then, by our proof of completeness, Γ has a countable model.

Note: compactness and the L-S theorem are both results about models, and don't themselves mention proof-systems. So you'd expect we ought to be able to prove them directly without appeal to the completeness theorem which mentions proof-systems. And we can!

(c) An easy argument shows that we can't consistently have (i) for each n a sentence $\exists n$ which is says that there are at least n things, (ii) a sentence $\exists\infty$ which is true in all and only infinite domains, and also (iii) compactness (see §6.2 fn.2). We noted that in the second-order case, we have (i) and (ii), so that rules out compactness. In the first-order case, we have (i) and (iii), so that implies

(7) There is no first-order sentence $\exists\infty$ which is true in all and only structures with infinite domains.

That's a nice mini-result about the limitations of first-order languages. But now let's note a second, much more dramatic, such result.

Suppose L_A is a formal first-order language for the arithmetic of the natural numbers. The precise details don't matter; but to fix ideas, suppose L_A's built-in non-logical vocabulary comprises the binary function expressions $+$ and $\times$ (with their obvious interpretations), the unary function expression S (expressing the successor, *next number*, function),[1] and the constant $\mathsf{0}$ (denoting zero). So note that L_A then has a sequence of expressions $\mathsf{0}, \mathsf{S0}, \mathsf{SS0}, \mathsf{SSS0}, \ldots$ which can serve as numerals, denoting 0, 1, 2, 3, ...

Now let the theory T_{true}, i.e. *true arithmetic*, be the set of *all* true L_A sentences. Then we can show the following:

(8) As well as being true of its 'intended model' – i.e. the natural numbers with their distinguished element zero and the successor, addition, and multiplication functions and ordering relation defined over them – T_{true} is *also* true of differently-structured, non-isomorphic, models.

This can be shown again by an easy compactness argument.[2]

---

[1] There's no significance in the choice of small-'s' earlier and big-'S' now – it's just a matter of local readability!

[2] Here's the proof idea, which is very neat. For brevity, write $\overline{\mathsf{n}}$ as short for n occurrences of S followed by $\mathsf{0}$: so $\overline{\mathsf{n}}$ is denotes n.

Let's add to the language L_A the single additional constant 'c'. And now consider the theory T^+_{true} formed in the expanded languages, which has as its axioms all of T_{true} plus the infinite supply of extra axioms $\mathsf{0} \neq \mathsf{c}$, $\overline{\mathsf{1}} \neq \mathsf{c}$, $\overline{\mathsf{2}} \neq \mathsf{c}$, $\overline{\mathsf{3}} \neq \mathsf{c}, \ldots$.

Now observe that any finite collection of sentences $\Delta \subset T^+_{true}$ has a model. Because Δ is finite, there will be a some *largest* number n such that the axiom $\overline{\mathsf{n}} \neq \mathsf{c}$ is in Δ; so just interpret c as denoting $n+1$ and give all the other vocabulary its intended interpretation, and every sentence in the finite set Δ will by hypothesis be true on this interpretation.

Since any finite collection of sentences $\Delta \subset T^+_{true}$ has a model, we can now appeal to compactness to conclude that T^+_{true} itself has a model. That model, as well as having a zero and its successors, must also have in its domain a distinct non-standard 'number' c to be the denotation of the new name c (where c is distinct from the denotations of $\mathsf{0}, \overline{\mathsf{1}}, \overline{\mathsf{2}}, \overline{\mathsf{3}}, \ldots$). And since the new model must still make true the old T_{true} sentence which says that everything in the domain has a successor, there will in addition be *more* non-standard numbers to be successor of c, the successor of *that*, etc.

Now take a structure which is a model for T^+_{true}, with its domain including non-standard numbers. Then in particular it makes true all the sentences of T^+_{true} which don't feature the constant c. But these are just the sentences of the original T_{true}. So this structure will still make all T_{true} true – even though its domain contains more than a zero and its successors, and so not 'look like' the original intended model.

And this is really rather remarkable! Formal first-order theories are our standard way of regimenting informal mathematical theories: but now we find that even T_{true} – the set of *all* first-order L_A truths taken together – still fails to pin down a unique structure for the natural numbers.

(d) And, turning now to the L-S theorem, we find that things only get worse. Again let's take a dramatic example.

Suppose we aim to capture the set-theoretic principles we use as mathematicians, arriving at the gold-standard Zermelo-Fraenkel set theory with the Axiom of Choice, which we regiment as the first-order theory ZFC. Then:

(9) ZFC, on its intended interpretation, makes lots of infinitary claims about the existence of sets much bigger than the set of natural numbers. But the downward Löwenheim-Skolem theorem tells us that, all the same, assuming ZFC is consistent and has a model at all, it has an unintended countable model (despite the fact that ZFC has a theorem which on the intended interpretation says that there are uncountable sets). In other words, ZFC has an interpretation in the natural numbers. Hence our standard first-order formalized set theory certainly fails to pin down the wildly infinitary universe of sets.

This result has even more of a wow factor! What is emerging then, in these first steps into model theory, are some unexpected(?) expressive limitations of first-order formalized theories.

(e) These limitations can be thought of as one of the main themes of Level 1 model theory (meaning the cluster of topics often touched on in introductory books on FOL).

At Level 2, we can pursue this theme further, starting with the upward Löwenheim-Skolem theorem which tells us that if a theory has an infinite model it will also have models of all larger infinite sizes (as you see, then, you'll need some basic grip on the idea of the hierarchy of different cardinal sizes to make full sense of this sort of result). Hence

(10) The upward and downward Löwenheim-Skolem theorems tell us that first-order theories which have infinite models won't be categorical – i.e. their models won't all look the same because they can have domains of different infinite sizes. For example, try as we might, a first-order theory of arithmetic will always have non-standard models which 'look too big' to be the natural numbers with their usual structure, and a first-order theory of sets will always have non-standard models which 'look too small' to be the universe of sets as we intuitively conceive it.

But if we can't achieve full categoricity (*all* models looking the same), perhaps we can get restricted categoricity results for some theories (telling us that all models *of a certain size* look the same) – when is this possible? An example you'll find discussed: the theory of dense linear orders is countably categorical (i.e. all its models of the size of the natural numbers are isomorphic – a lovely result due to Cantor); but it isn't categorical at the next infinite size up. On the other hand, theories of first-order arithmetic are not even countably categorical.

Still at Level 2, there are results about which theories are *complete* in the sense of entailing either φ or $\neg\varphi$ for each relevant sentence φ, and how this relates to being categorical at a particular size. And there is another related notion of so-called model-completeness: but let's not pause over that.

Instead, let's mention just one more fascinating topic that you will encounter early in your model theory explorations:

(11) We can take a standard first-order theory of the natural numbers and use a compactness argument to show that it has a non-standard model which has an element c in the domain distinct from (and indeed greater than) zero or any of its successors. Similarly, we can take a standard first-order of the real numbers and use another compactness argument to show that it has a non-standard model with an element r in the domain such that that $0 < |r| < 1/n$ for any natural number n. So in this model, the non-standard real r is non-zero but smaller than any rational number, so is *infinitesimally* small. And indeed our model will have non-standard reals infinitesimally close to any standard real.

In this way, we can build up a model of *non-standard analysis* with infinitesimals (where e.g. a differential really can be treated as a ratio of infinitesimally small numbers – in just the sort of way that we all supposed wasn't respectable!).

(f) An issue arising from all this. You may well have already encountered in a maths course a proof that a theory usually called 'Peano Arithmetic' is categorical – all its models do indeed structurally look the same. So how does that square with the result that first-order arithmetics are *not* categorical?

The answer is straightforward, and already flagged up before:

(12) As already indicated in (3), the version of Peano Arithmetic which is categorical is a *second-order* theory – i.e. a theory which quantifies not just over numbers but over numerical properties, and has a second-order induction principle. Going second-order makes all the difference in arithmetic, and in other theories too like the theory of the real numbers.[3]

But why? To understand what is going on here, you need to understand something about the contrast between first-order theories and second-order ones. See our previous chapter!

(g) Finally, for a little more by way of a general introductory overview of model theory, you could also usefully look at Wilfrid Hodges's piece in the *Stanford Encyclopaedia of Philosophy* 'Model Theory', at [sepmodel].

[3] With second-order set theory, however, the situation is rather more complicated.

7.2 Main recommendations for beginning first-order model theory

As we've already noted, when exploring model theory you will very quickly meet claims that involve the idea of there being different infinite cardinalities, and you will come across occasional passing references to the Axiom of Choice. Let's take it you are familiar enough with these basic set-theoretic ideas (perhaps from the readings suggested back in Chapter 4).

A number of the introductions to FOL that I noted in $5.4 have treatments of the Level 1 basics; I'll be recommending one in a moment, and will return to some of the others in the next section on parallel reading.

Going just a little beyond, the very first volume in the prestigious and immensely useful Oxford Logic Guides series is Jane Bridge's short *Beginning Model Theory: The Completeness Theorem and Some Consequences* (Clarendon Press, 1977). This neatly takes us through some Level 1 and a few Level 2 topics. But the writing, though very clear, is also rather terse in an old-school way; and the book – not unusually for that publication date – looks like photo-reproduced typescript, which is off-putting to read.

What, then, are the more recent options?

1. I have already sung the praises of Derek Goldrei's *Propositional and Predicate Calculus: A Model of Argument* (Springer, 2005) for the accessibility of its treatment of FOL in the first five chapters. You should now read Goldrei's §§4.4 and 4.5 (on which I previously said you could skip), and then Chapter 6 'On some uses of compactness'.

   In a little more detail, §4.4 introduces some axiom systems describing various mathematical structures (partial orderings, groups, rings, etc.): this section could be particularly useful to philosophers who haven't really met the notions before. Then §4.5 introduces the notions of substructures and structure-preserving isomorphisms. After proving the compactness theorem in §6.1 (as a corollary of his completeness proof), Goldrei proceeds to use it in §§6.2 and 6.3 to show various theories can't be finitely axiomatized, or can't be nicely axiomatized at all. §6.4 introduces the Löwenheim-Skolem theorems and some consequences, and the following section introduces the notion of 'diagrams' and puts it to work. The final section, §6.6 considers issues about categoricity, completeness and decidability. All this is done with the same admirable clarity as marked out Goldrei's earlier chapters.

Goldrei, though, goes quite slowly and doesn't get very far (it is Level 1 model theory). If you want to take a further step (Level 2), here are two suggestions. Neither is quite ideal, but each has virtues. The first is

2. María Manzano, *Model Theory*, Oxford Logic Guides 37 (OUP, 1999). This book aims to be an introduction at the kind of levels we are currently concerned with. And standing back from the details, I do like the way

that Manzano structures her book. The sequencing of chapters makes for a very natural path through her material, and the coverage seems very appropriate for a book at Levels 1 and 2. After chapters about structures (and mappings between them) and about first-order languages, she proves the completeness theorem again, and then has a sequence of chapters on various core model-theoretic notions and proofs. Overall, this should all be tolerably accessibly (especially if not your very first encounter with model theoretic ideas).

However, it seems to me that the discussions at some points would have benefitted from rather more informal commentary, motivating various choices. Sometimes the symbolism is unncessarily heavy-handed. And there are some other infelicities. But overall – though as I say not ideal – Manzano's text could work well enough as a follow-up to Goldrei. See my note [manzanobook] for more details.

Another option is to look at the first two-thirds of the following book, which is explicitly aimed at undergraduate mathematicians, and is at approximately the same level of difficulty as Manzano:

3. Jonathan Kirby, *An Invitation to Model Theory* (CUP, 2019). As the blurb says, "The highlights of basic model theory are illustrated through examples from specific structures familiar from undergraduate mathematics." Now, one thing that usually *isn't* already familiar to most undergraduate mathematicians is any serious logic: so Kirby's book is an introduction to model theory that doesn't presuppose a previous FOL course. So he has to start with some rather speedy explanations in Part I about first-order languages and interpretations in structures.

   The book *is* then nicely arranged. Part II of the book is on 'Theories and Compactness', Part III on 'Changing Models', and Part IV on 'Characterizing Definable Sets'. (I'd say that some of the further Parts of the book, though, go a bit beyond what you need at this stage.)

   Kirby writes admirably clearly; but his book goes pretty briskly and would have been improved – at least for self-study – if he had slowed down for some more classroom asides. So I can imagine that some readers would struggle with parts of this short book if were treated as a sole introduction to model theory. However, again if you have read Goldrei, it should be very helpful as an alternative or complement to Manzano's book (For a little more about it, see [kirbybooknote].

We noted that first-order theories behave differently from second-order theories where we have quantifiers running over the properties and functions defined over a domain, as well as over the objects in the domain. For more on this see the readings on second-order logic suggested in §6.3.

7.3 Some parallel and slightly more advanced reading

Let's begin with a recent book with an enticing title:

5. Roman Kossak, *Model Theory for Beginners: 15 Lectures** (College Publications 2021). As the title indicates, the fifteen chapters of this short book – just 138 pages – have their origin in introductory lectures, given to graduate students in CUNY.

   After initial chapters on structures and (first-order) languages, Chapters 3 and 4 are on definability and on simple results such as that ordering is not definable in the language for the integers with addition, $(\mathbb{Z}, +)$. Chapter 5 introduces the notion of 'types', and e.g. gives the back-and-forth proof conventionally attributed to Cantor that countable dense linearly ordered sets without endpoints are always isomorphic to the rationals in their natural order, $(\mathbb{Q}, <)$. Chapter 6 defines relations between structures like elementary equivalence and elementary extension, and establishes the so-called Tarski-Vaught test. Then Chapter 7 proves the compactness theorem, with Chapter 8 using compactness to establish some results about non-standard models of arithmetic and set theory.

   So there is a somewhat different arrangement of initial topics here, compared with books whose first steps in model theory are applications of compactness. The early chapters are indeed nicely done. However, I don't think that Kossak's Chapter 8 will be found an outstandingly clear and helpful first introduction to applications of compactness – it will probably be best read after e.g. Goldrei's nice final chapter in his logic text.

   Chapter 9 is on categoricity – in particular, countable categoricity. (Very sensibly, Kossak wants to keep his use of set theory in this book to a minimum; but he does have a section here looking at κ-categoricity for larger cardinals κ.) And now the book speeds up and requires rather more of its reader, and eventually touches on what I think of as Level 3 topics. Real beginners in model theory without much mathematical background could begin to struggle after the half-way mark in the book.

I mentioned before that some other introductory texts on FOL apart from Goldrei's have sections or chapters beginning model theory. Some topics are briefly touched on in §2.6 of Herbert Enderton's *A Mathematical Introduction to Logic* (Academic Press 1972, 2002), and there is discussion of non-standard analysis in his in §2.8: but this, for our purposes here, is perhaps too little done too fast. So I think the following is better:

6. Dirk van Dalen *Logic and Structure* (Springer, 1980; 5th edition 2012), Chapter 3, which covers rather more model-theoretic material than Enderton and in more detail. You could read §3.1 for revision on the completeness theorem, then tackle §3.2 on compactness, the Löwenheim-

Skolem theorems and their implications, before moving on to the action-packed §3.3 which covers more model theory including non-standard analysis again, and indeed touches on slightly more advanced topics like 'quantifier elimination'.

And there is a nice chapter in another often-recommended text:

7. Richard E. Hodel, *An Introduction to Mathematical Logic** (originally published 1995; Dover reprint 2013). In Chapter 6, 'Mathematics and Logic', §6.1 discusses first-order theories, §6.2 treats compactness and the Löwenheim-Skolem theorem, and §6.3 is on decidable theories. Very clearly done.

Now, thanks to the efforts of the respective authors to write very accessibly, the suggested main path through FOL with Chiswell & Hodges → (part of) Leary & Kristiansen → the beginnings of model theory with (excerpts from) Goldrei → Manzano/Kirby is not at all a hard road to follow. Yet we end up at least in the foothills of model theory. We can climb up to the same foothills by routes involving rather tougher scrambles, taking in some additional side-paths and new views along the way. Here is a suggestion for the more mathematical reader:

8. Shawn Hedman, *A First Course in Logic* (OUP, 2004). This covers a surprising amount of model theory. Ch. 2 tells you about structures and about relations between structures. Ch. 4 starts with a nice presentation of a Henkin completeness proof, and then pauses (as Goldrei does) to fill in some background about infinite cardinals etc., before going on to prove the Löwenheim-Skolem theorems and compactness theorems. Then the rest of Ch. 4 and the next chapter covers more introductory model theory, though already touching on a number of topics beyond the scope of Mansion's book (we are already at Level 2.5, perhaps!). Hedman so far could therefore serve as a rather tougher alternative to Manzano's treatment.

   Then Ch. 6 takes the story on a lot further, quite a way beyond what I'd regard as elementary model theory. For more, see [hedmanbook].

Last but certainly not least, philosophers will certainly want to tackle (parts of) the following quite recently published book, which strikes me as a very impressive achievement:

9. Tim Button and Sean Walsh, *Philosophy and Model Theory** (OUP, 2018). This book both explains technical results in model theory, and also explores the appeals to model theory in various branches of philosophy, particularly philosophy of mathematics, but in metaphysics more generally (recall 'Putnam's model-theoretic argument'), the philosophy of science, philosophical logic and more. So that's a very scattered literature that is being expounded, brought together, examined, inter-related,

criticized and discussed. Button and Walsh don't pretend to be giving the last word on the many and varied topics they discuss; but they are offering us a very generous helping of first words and second thoughts. It's a large book because it is to a significant extent self-contained: model-theoretic notions get defined as needed, and many of the more significant results are proved.

The philosophical discussion is done with vigour and a very engaging style. And the expositions of the technical stuff are usually exemplary (the authors have a good policy of shuffling some extended proofs into chapter appendices). They also say more about second-order logic and second-order theories than is usual.

But I still suspect that an amount of the material is more difficult that the authors fully realize: we soon get to tangle with some Level 3 model theory, and quite a lot of other technical background is presupposed. The breadth and depth of knowledge brought to the enterprise is remarkable: but it does make of a bumpy ride even for those who already know quite a lot. Philosophical readers of this Guide will probably find the book challenging, then, but should find at least the earlier parts fascinating. And indeed, with judicious skimming/skipping – the signposting in the book is excellent – mathematicians with an interest in some foundational questions should find a great deal of interest here too.

And that might already be about as far as many philosophers may want or need to go in this area. Many mathematicians, however, will want to take the story about model theory rather further: so the story resumes in Part III.

7.4 A little history?

The last book we mentioned has an historical appendix contributed by a now familiar author:

10. Wilfrid Hodges, 'A short history of model theory', in Button and Walsh, pp. 439–476.

However, a lot of this refers to model theoretic topics a level up from our current more elementary concerns, so won't be very accessible at this stage. The following focuses on topics more from the beginning of model theory:

11. R. L Vaught, 'Model theory before 1945' in L. Henkin et al, eds, *Proceedings of the Tarski Symposium* (American Mathematical Society, 1974), pp. 153–172.

You'll still probably have to skip parts, but it will give you some idea of the early developments, if you want.

But here's something which is much more fun to read. Alfred Tarski was one of the key figures in that early history. And there is a very enjoyable and well-written biography, which vividly portrays the man, and gives a wonderful sense of his intellectual world, but also contains accessible interludes on his logical work:

12. Anita Burdman Feferman and Solomon Feferman, *Alfred Tarski, Life and Logic* (CUP, 2004).

8 Arithmetic, computability, and incompleteness

The standard mathematical logic curriculum, as well as looking at some elementary results about formalized theories and their models in general, looks at two particular instances of non-trivial, rigorously formalized, axiomatic systems. First, there's *arithmetic* (a paradigm theory about finite whatnots); and then there is *set theory* (a paradigm theory about infinite whatnots). We consider set theory in the next chapter. This chapter is about arithmetic and related matters. More specifically, we consider three inter-connected topics:

1. The elementary theory of numerical *computable functions*.
2. Formal *theories of arithmetic* and how they represent computable functions.
3. Gödel's epoch-making proof of the *incompleteness* of any sufficiently nice formal theory that can 'do' enough arithmetical computations.

Before turning to some short topic-by-topic overviews, though, it's worth pausing for a quick general point about why the idea of computability is of such very central concern to formal logic.

8.1 Logic and computability

(a) The aim of regimenting informal arguments and informal theories into formalized versions is to eliminate ambiguities and to make everything entirely determinate and transparently clear (even if it doesn't always seem that way to beginners!). So, for example, we want it to be entirely determinate what is and what isn't a formal sentence of a given theory, what is and what isn't an axiom of the theory, and what is and what isn't a formal proof in the theory, with absolutely no room left for doubt or dispute.

Let's say, as an initial rough characterization:

> A property P is *effectively decidable* if and only if there is an *algorithm* (a finite set of instructions for a deterministic computation) for settling in a finite number of steps, whether a relevant object has property P.

> Relatedly, the answer to a question Q is effectively decidable if and only if there is an algorithm which gives the answer, again by a deterministic computation, in a finite number of steps.

In other words, a property is effectively decidable just when there's a step-by-step mechanical routine for settling whether an object of the relevant kind has property P, such that a suitably programmed deterministic computer could in principle implement the routine (idealizing away from practical constraints of time, etc.). Similarly, the answer to a question is effectively decidable just when a suitably programmed computer could deliver the answer (in principle, in a finite time).

Two initial examples from propositional logic: the property of being a tautology is effectively decidable (by a truth-table test); and we can effectively decide what is the main connective of a sentence (by bracket counting).

And the point we just made is that we will want it to be effectively decidable e.g. whether a given string of symbols is a well-formed formula of a certain formal language, whether a formula is an axiom of a given formal theory, and whether an array of formulas is a correctly formed proof of the theory. In other words, we will want to set up a formal deductive theory so that a computer could, in principle, mindlessly check e.g. the credentials of a purported proof by deciding whether each step of the proof is indeed in accordance with the official rules of the theory.

(b) NB: It is one thing to be able to effectively decide whether a purported proof of P really is a proof in a given formal theory T. It is another thing entirely to be able to decide in advance whether P has a proof in T.

You'll soon enough find out that, e.g., in a properly set up formal theory of arithmetic T we can effectively check whether a supposed proof of P indeed conforms to the rules of the game. But there will be no way of deciding in advance whether a T-proof of P exists. Such a theory T is said to be undecidable.

It's nice when a theory *is* decidable, i.e. when a computer can tell us whether a given proposition follows from the theory. But few interesting theories are decidable in this sense: so mathematicians aren't going to be put out of business!

(c) Now, in our initial rough definition of the notion of effective decidability, we invoked the idea of what an idealized computer could (in principle) do by implementing some algorithm. This idea surely needs further elaboration.

1. As a preliminary step, we can narrow our focus and just consider the decidability of *arithmetical* properties. Why? Because we can represent e.g. facts about finite whatnots like formulas and proofs by using numerical codings. (Recall how computers work on all kinds of data via binary codings!)

2. And as a second step, we can trade in questions about the effective decidability of arithmetical *properties* for questions about the algorithmic computability of numerical *functions*. Why? Because for any numerical property P we can define a corresponding numerical function (its so-called 'characteristic function') c_P such that if n has the property P, $c_P(n) = 1$

and if n doesn't the have property P, $c_P(n) = 0$. Think of '1' as coding for truth, and '0' for falsehood. Then the question (i) 'can we effectively decide whether a number has the property P?' becomes the question (ii) 'is the numerical function c_P effectively computable by an algorithm?'.

So, in this way, we do quickly get from wanting to know more about the notion of an effectively decidable property to wanting to know more about the notion of a computable numerical function.

8.2 Computable functions: an overview

(a) We will as before use 'S' for the function that maps a number to its successor. Consider, then, the following pairs of equations:

$$x + 0 = x$$
$$x + Sy = S(x + y)$$

$$x \times 0 = 0$$
$$x \times Sy = (x \times y) + x$$

$$x^0 = S0$$
$$x^{Sy} = (x^y \times x)$$

These pairs of equations should be very familiar (at least when written with postfix '+1' instead of prefix 'S'): they in turn define addition, multiplication and exponentiation for the natural numbers.

Take the initial pair of equations. The first of them fixes the result of adding zero to a given number. The second fixes the result of adding the successor of y in terms of the result of adding y. Hence applying and re-applying the two equations, they together tell us how to add $0, S0, SS0, SSS0, \ldots$, i.e. they tell us how to add *any* natural number to a given number x. Similarly, the first of the equations for multiplication fixes the result of multiplying by zero. The second equation fixes the result of multiplying by Sy in terms of the result of multiplying by y and doing an addition. Hence the two pairs of equations together tell us how to multiply a given number x by any of $0, S0, SS0, SSS0, \ldots$. Similarly of course for the pair of equations for exponentiation.

And now note that the six equations taken together not only *define* exponentiation, but they combine to give us an *algorithm* for computing x^y for any natural numbers x, y – they tell us how to compute x^y by doing repeated multiplications, which we in turn compute by doing repeated additions, which we compute by repeated applications of the successor function. That is to say, the chain of equations amounts to a set of instructions for a deterministic step-by-step computation which will output the value of x^y in a finite number of steps. Hence, exponentiation is indeed an effectively computable function.

(b) In each of our pairs of equations, the second one fixes the value of the defined function for argument Sy by invoking the value of the *same* function for argument y. A procedure where we evaluate a function for one input by calling the *same* function for some smaller input(s) is standardly termed 'recursive' – and the particularly simple type of procedure we've illustrated three times is called, more precisely, primitive recursion. Now – arm-waving more than a bit! – consider *any* function which can be defined by a chain of equations similar to the chain of equations giving us a definition of exponentiation. Suppose that, starting from trivial functions like the successor function, we can build up the function's definition by using primitive recursions and composition of functions. Such a function is said to be *primitive recursive.*

Generalizing from the case of exponentiation, any primitive recursive function – p.r. function, for short – is similarly effectively computable; in other words, there is an algorithmic step-by-step procedure for computing its value for any given input(s).

However, we quickly have the following basic result:

> Not all effectively computable functions are primitive recursive.

And there's a very neat abstract argument which proves the point.[1]

(c) The obvious next question is: what further ways of defining functions – in addition to primitive recursion – also give us effectively computable functions?

Here's a pointer. The definition of (say) x^y by primitive recursion in effect tells us to start from x^0, then loop round applying the recursion equation to

[1]It is another nice example of a so-called *diagonalization* argument (compare §4.1). So, start by imagining definitions for all the p.r. functions f_0, f_1, f_2, ... being listed off in some standard specification language in 'dictionary order' (we can do this). Then consider the table

	0	1	2	3	...
f_0	$\underline{f_0(0)}$	$f_0(1)$	$f_0(2)$	$f_0(3)$	...
f_1	$f_1(0)$	$\underline{f_1(1)}$	$f_1(2)$	$f_1(3)$	...
f_2	$f_2(0)$	$f_2(1)$	$\underline{f_2(2)}$	$f_2(3)$	...
f_3	$f_3(0)$	$f_3(1)$	$f_3(2)$	$\underline{f_3(3)}$	...
...	...	...	...	...	$\searrow$

Down the table we list the p.r. functions f_0, f_1, f_2, ... An individual row then gives the values of a particular f_n for each argument. Now look down the diagonal, and define the corresponding *diagonal* function, by putting $\delta(n) = f_n(n) + 1$.

To compute $\delta(n)$, we list off our recipes for p.r. functions until we get to the recipe for f_n. We follow the instructions in that recipe to evaluate that function for the argument n. We then add one. Each step is entirely mechanical. So we can evaluate δ for the given input using a step-by-step algorithmic procedure. Hence δ is an effectively computable function.

By construction, however, the function δ can't be primitive recursive. For suppose otherwise. Then δ must appear somewhere in the enumeration of p.r. functions, i.e. must be the function f_d for some index number d. Now ask what the value of $\delta(d)$ is. By hypothesis, the function δ is none other than the function f_d, so $\delta(d) = f_d(d)$. But by the initial definition of the diagonal function, $\delta(d) = f_d(d) + 1$. Contradiction.

So we have, as they say, 'diagonalized out' of the class of p.r. functions to define a new function δ which is still effectively computable but not primitive recursive.

compute x^1, then x^2, then x^3, ..., keeping going until we reach x^y. In all, we have to loop around y times. In some standard computer languages, implementing this procedure involves using a 'for' loop (which tells us to iterate some procedure, counting as we go, and to do this for cycles numbered 1 to y). In this case, *the number of iterations is given in advance as we enter the loop.* But of course, standard computer languages also have programming structures which implement unbounded searches – they allow open-ended 'do until' loops (or equivalently, 'do while' loops). In other words, they allow some process to be iterated until a given condition is satisfied, *where no prior limit is put on the number of iterations to be executed.*

This suggests that one way of expanding the class of computable functions beyond the primitive recursive functions will be to allow computations employing *open-ended searches.* So let's suppose we do this (there's a standard device for this, but let's not worry about the details now). Functions – more precisely, *total* functions that deliver an output for any numerical input – which can be computed by a chain of applications of primitive recursion and/or open-ended searches are called (simply) *recursive.*

(d) Predictably enough, the next question is: have we *now* got all the effectively computable functions? [2]

The claim that the recursive functions are indeed just the intuitively computable total functions is *Church's Thesis*, and is very widely believed to be true (or at least, it is taken to be an entirely satisfactory working hypothesis). Why? For a start, there are quasi-empirical reasons: no one has found a function which is incontrovertibly computable by a finite-step deterministic algorithmic procedure but which isn't recursive. But there are also much more principled reasons for accepting the Thesis.

Consider, for example, Turing's approach to the notion of effective computation. He famously aimed to analyse the idea of a step-by-step computation procedure down to its very basics, which led him to the concept of computation by a Turing machine (a minimalist computer). And what we can call Turing's Thesis is that the effectively computable (total) functions are just the functions which are computable by some suitably programmed Turing machine.

So do we now have two *rival* claims, Church's and Turing's, about the class of computable functions? Not at all! For it turns out to be quite easy to prove the technical result that a function is recursive if and only if is Turing computable. And so it goes: every other attempt to give an exact characterization of the class of effectively computable functions turns out to locate just the *same* class of functions. That's remarkable, and this is a key theme you will want to explore in a first encounter with the theory of computable functions.

(e) It is fun to find out more about Turing machines, and even to learn to write a few elementary programs (in effect, it is learning to write in a 'machine code').

[2] Technical note: this time, we can't 'diagonalize out', because we can't mechanically list off the recipes involving those open-ended searches that will always terminate and so define total functions.

And there is a beautiful early result that you will soon encounter:

> There is no mechanical decision procedure which can determine whether Turing machine number e, fed a given input n, will ever halt its computation (so there is no general decision procedure which can tell whether Turing machine e in fact computes a total function).

How do we show that? It's another easy diagonalization argument.

But no more spoilers! – I leave it to you to read up on the 'undecidability of the halting problem', and its many weighty implications.

8.3 Formal arithmetic: an overview

(i) The elementary theory of computation really is a lovely area, where accessible Big Results come thick and fast! But now we must turn to consider formal theories of arithmetic.

We standardly focus on *First-order Peano Arithmetic*, PA. It will be no surprise to hear that this theory has a first-order language and logic! It has a built-in constant $\mathsf{0}$ to denote zero, has symbols for the successor, addition and multiplication functions (to keep things looking nice, we still use a prefix S, and infix $+$ and $\times$), and its quantifiers run over the natural numbers. Note, we can form the sequence of numerals $\mathsf{0}, \mathsf{S0}, \mathsf{SS0}, \mathsf{SSS0}, \ldots$ (we will use $\overline{\mathsf{n}}$ to abbreviate the result of writing n occurrences of S before $\mathsf{0}$, so $\overline{\mathsf{n}}$ denotes n).

PA has the following three pairs of axioms governing the three built-in functions:

$$\forall \mathsf{x}\, \mathsf{0} \neq \mathsf{Sx}$$
$$\forall \mathsf{x} \forall \mathsf{y}(\mathsf{Sx} = \mathsf{Sy} \to \mathsf{x} = \mathsf{y})$$

$$\forall \mathsf{x}\, \mathsf{x} + \mathsf{0} = \mathsf{x}$$
$$\forall \mathsf{x} \forall \mathsf{y}\, \mathsf{x} + \mathsf{Sy} = \mathsf{S}(\mathsf{x} + \mathsf{y})$$

$$\forall \mathsf{x}\, \mathsf{x} \times \mathsf{0} = \mathsf{0}$$
$$\forall \mathsf{x} \forall \mathsf{y}\, \mathsf{x} \times \mathsf{Sy} = (\mathsf{x} \times \mathsf{y}) + \mathsf{x}$$

The first pair of axioms specifies that distinct numbers have distinct successors, and that the sequence of successors never circles round and ends up with zero again: so the numerals, as we want, must denote a sequence of distinct numbers, zero and all its eventual successors. The other two pairs of axioms formalize the equations defining addition and multiplication which we have met before.

And then, crucially, there is also an arithmetical Induction Rule. Informally, the familiar idea is that, if zero has property P and if a number's having P implies that its successor has P too, then *all* numbers have property P. Formally, in first-order PA, we have the rule:

> If $\varphi(\mathsf{x})$ is a formula with x free, then from $\varphi(\mathsf{0})$ and $\forall \mathsf{x}(\varphi(\mathsf{x}) \to \varphi(\mathsf{Sx}))$ we can infer $\forall \mathsf{x} \varphi(\mathsf{x})$.

Alternatively but obviously equivalently, we can stipulate that

> Any wff of the form $(\{\varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(Sx))\} \rightarrow \forall x\varphi(x))$ is an axiom.

You need to get at least some elementary familiarity with the workings of this theory.

(ii) But why concentrate on first-order PA? We've emphasized in §6.2 that our informal induction principle is most naturally construed as involving a second-order generalization – *for any arithmetical property P*, if zero has P, and if a number which has P always passes it on to its successor, then every number has P. And when Richard Dedekind (1888) and Giuseppe Peano (1889) gave their axioms for what we can call Dedekind-Peano arithmetic, they correspondingly gave a second-order formulation for their versions of the induction principle. Put it this way: Dedekind and Peano's principle quantifies over *all* properties of numbers, while in first-order PA our induction principle rather strikingly only deals with those properties of numbers which can be expressed by open formulas of its restricted language. Why go for the weaker first-order principle?

Well, as we have seen, first-order logic is in some respects much better behaved than second-order logic (for a start, you can't capture full second-order logic in a nice complete deductive system). And some would say that second-order quantifications over properties should be thought of as being quantifications over property-extensions, i.e. over *sets* of the objects – hence Quine's quip that second-order logic is set theory in sheep's clothing. Dedekind's version of arithmetic indeed explicitly deals with sets. So, the argument goes, if we want a pure theory of arithmetic, one whose logic can be formalized, we should stick to a first-order formulation just quantifying over numbers. And then something like PA's induction rule (or the suite of axioms of the form we described) is the best we can do.

(iii) But still, even if we have decided to stick to a first-order theory, why restrict ourselves to the impoverished resources of PA, with only three function-expressions built into its language? Why not have an expression for e.g. the exponential functions as well, and add to the theory the two defining axioms for that function? Indeed, why not add expressions for other recursive functions too, and then also include appropriate axioms for *them* in our formal theory?

Good question. The answer is to be found in a neat technical observation first made by Gödel. Once we have successor, addition and multiplication available, plus the usual first-order logical apparatus, we can in fact *already* express any other computable (i.e. recursive) function. To take the simplest sort of case, suppose f is a one-place recursive function: then there will be a two-place expression of PA's language which we can abbreviate $F(x, y)$ such that $F(\overline{m}, \overline{n})$ is true if and only if $f(m) = n$. Moreover, when $f(m) = n$, PA can *prove* $F(\overline{m}, \overline{n})$, and when $f(m) \neq n$, PA can *prove* $\neg F(\overline{m}, \overline{n})$. In this way, PA as it were already 'knows' about all the recursive functions, and it would in fact be redundant to add new vocabulary for them.

So PA is expressively a lot richer than you might initially suppose. And indeed, it turns out that even a induction-free subsystem of PA known as Robinson Arithmetic (often called simply Q) can express the recursive functions.

And this key fact puts you in a position to link up your investigations of PA with what you know about computability. For example, there is a fairly straightforward proof that there is no mechanical procedure that a computer could implement which can decide whether a given arithmetic sentence is a theorem of PA (or even a theorem of Q).

(iv) On the other hand, despite its richness, PA is a first-order theory with infinite models, so – applying results from elementary model theory (see the previous chapter) – this first order arithmetic will have non-standard models, i.e. will have models whose domains contain more than a zero and its successors. It is worth knowing at an early stage just something about what some of these non-standard models can look like. And you will also want to further investigate the contrast with second-order versions of arithmetic which are categorical (i.e. don't have non-standard models).

8.4 Gödel's theorems: an overview

(i) Now for our third related topic: Gödel's incompleteness theorems.

First-order PA, we said, turns out to be a very rich theory. Is it rich enough to settle every question that can be raised in its language? No! In 1931, Kurt Gödel proved that PA is *negation incomplete* – meaning that we can form a sentence G in its language such that PA proves neither G nor $\neg$G.[3] How does he do the trick?

(ii) It's fun to give an outline sketch, which I hope will intrigue you enough to leave you wanting to find out more! So:

G1. Gödel introduces a *Gödel-numbering scheme* for a formal theory like PA, which is a simple way of coding expressions of PA – and also sequences of expressions of PA – using natural numbers. The scheme gives an algorithm for sending an expression (or sequence of expressions such as a proof) to a number; and a corresponding algorithm for undoing the coding, sending a code number back to the expression (sequence of expressions) it codes. Relative to a choice of scheme, the code number for an expression (or a sequence of expressions) is its unique *Gödel number*.

G2. Fix on such a numbering scheme. Then, relative to that scheme we can define relations like Prf, where $Prf(m, n)$ holds if and only if m is the Gödel number of a PA-proof of the sentence with code number n. So Prf is a numerical relation which, so to speak, 'arithmetizes' the syntactic relation between a sequence of expressions (proof) and a particular sentence (con-

[3] More accurately: he gives an argument which applies to a different theory, but one which can be directly carried over to apply to PA instead.

clusion). True enough, this isn't kind of arithmetic relation we are familiar with. But it is perfectly well-defined.

G3. And now note that *Prf* is not only well-defined but it is an effectively decidable arithmetical relation – i.e. there's a procedure for computing, given numbers m and n, whether $Prf(m, n)$ holds. Informally, we just decode m (that's an algorithmic procedure). Now check whether the resulting sequence of expressions – if there is one – is a well-constructed PA-proof according to the rules of the game (proof-checking is another algorithmic procedure). If that sequence is a proof, check whether it ends with a sentence with the code number n (that's another algorithmic procedure).

G4. It is easy to extend the result that PA can express any computable *function* to the claim that it can express any computably decidable *relation*. So there will in particular be a formal expression in the language of PA which we can abbreviate $\mathsf{Prf}(\mathsf{x}, \mathsf{y})$ which expresses the relation *Prf*. This means that $\mathsf{Prf}(\overline{\mathsf{m}}, \overline{\mathsf{n}})$ is true if and only if m codes for a PA proof of the sentence with Gödel number n.

G5. Now define $\mathsf{Prov}(\mathsf{x})$ to be the expression $\exists \mathsf{z}\mathsf{Prf}(\mathsf{z}, \mathsf{x})$. Then $\mathsf{Prov}(\overline{\mathsf{n}})$, i.e. $\exists \mathsf{z}\mathsf{Prf}(\mathsf{z}, \overline{\mathsf{n}})$, is true iff some number Gödel-numbers a PA-proof of the wff with Gödel-number n, i.e. is true just if the wff with code number n is a theorem of PA. Therefore $\mathsf{Prov}(\mathsf{x})$ is naturally called a *provability predicate.*

G6. Next, with only a little bit of cunning, we construct a *Gödel sentence* G in the language of PA with the following property: G is true if and only if $\neg\mathsf{Prov}(\overline{\mathsf{g}})$ is true, where g is the code number of G.

Don't worry for the moment about how we do this cunning construction (it is surprisingly easy). Just note that G is true on interpretation if and only if the sentence with Gödel number g is not a PA-theorem, i.e. if and only if G is not a PA-theorem.

In short, G *is true if and only if it isn't a* PA-*theorem.* So, rather stretching a point, it is rather as if G 'says' *I am unprovable in* PA.

G7. Now, suppose G were provable in PA. Then, since G is true if and only if it isn't a PA-theorem, G would be false. So PA would have a false theorem. Hence assuming PA is *sound* and only has true theorems, then it can't prove G. Hence, since it is not provable, G is indeed true. Which means that ¬G is false. Hence, still assuming PA is sound, it can't prove ¬G either.

So, in sum, assuming PA is sound, it can't prove either of G or ¬G. As announced, PA is negation incomplete.

Wonderful!

(iii) Now the argument generalizes to formal theories of arithmetic other than PA. Suppose T is *any* nicely axiomatized sound theory which can express enough arithmetic ('nicely' is a placeholder for a story about how we can effectively decide what's an axiom etc. – let's not fret about the details here). Then, the claim is, we can use the same sort of cunning construction to find a true G_T such that T can prove neither G_T nor $\neg\mathsf{G}_T$. Let's be really clear: this doesn't, repeat *doesn't*, say that G_T is 'absolutely unprovable', whatever that could mean. It just says that G_T and its negation are unprovable-in-T.

Ok, you might well ask, why don't we simply 'repair the gap' in T by adding the true sentence G_T as a new axiom? Well, consider the theory $U = T + \mathsf{G}_T$ (to use an obvious notation). Then (i) U is still sound, since the old T-axioms are true and the added new axiom is true. (ii) U is still a nicely axiomatized formal theory given that T is. (iii) U can still express enough arithmetic. So we can find a sentence G_U such that U can prove neither G_U nor $\neg\mathsf{G}_U$.

And so it goes. Keep throwing more and more additional true axioms at T and our theory will remain negation-incomplete (unless it stops counting as nicely axiomatized). So here's the key take-away message: any sound nicely axiomatized theory T which can express enough arithmetic will not just be incomplete but in a good sense T will be *incompletable*.

"Hold on! When discussing model theory in §7.1 we mentioned the theory T_{true}, i.e. *true arithmetic*, the set of *all* true sentences of the language of PA. This theory is trivially complete – for any sentence S of the language, it (trivially) proves the true one out of S and ¬S!" True enough. But what follows is that T_{true} can't be nicely axiomatized. And this is, indeed, one reason why the incompleteness theorem matters. It sabotages the logicist project of Frege and Russell which (putting it crudely) aims to show that we can set down some logical axioms and definitions from which we can deduce, for a start, all the truths of basic arithmetic (so giving us a negation-complete theory for such truths). That can't be done. But then, if even the truths of the first-order language of PA can't be captured in that way, if the basic truths of arithmetic outstrip what is settled by logic and definitions of arithmetical vocabulary, then just what *is* the status of such truths?

(iv) Now, note that our first sketch of an incompleteness theorem made the *semantic* assumption that we are dealing with a *sound* theory. But we can in fact prove incompleteness on the basis of a *syntactic* assumption too, namely that we are dealing with a *consistent* theory which proves no contradiction (well, it is a little bit more complicated than that, but again we won't fret about the fine print here). The crucial point is that our incompleteness theorem comes in two flavours, depending whether we appeal to semantic or syntactic assumptions. And it is in fact the second, syntactic, version that is usually called *Gödel's First Incompleteness Theorem*: roughly, a consistent theory which can prove enough arithmetic can't be negation complete. A proof of this version is still quite elementary, but it requires more preparatory work showing that theories like PA can not only express but prove relevant facts about recursive functions.

(v) There's a Second Incompleteness Theorem too. The syntactic version of the First Theorem for PA tells us that *if* PA *is consistent,* G *is unprovable in* PA. But, using the Gödel-numbering scheme we used when proving this Theorem, this fact is itself expressible in PA! Why?

Well, the claim that PA is consistent is equivalent to the arithmetic claim that no number codes for a proof of absurdity in PA; and, using the provability predicate, *that* arithmetical claim can be expressed by an arithmetic sentence we can abbreviate Con. While the claim that G is unprovable in PA is expressed in PA by G itself. So the First Theorem can be coded up in PA by the arithmetic sentence $\mathsf{Con} \to \mathsf{G}$.

But there's more. We said that the proof of the First Incompleteness Theorem is elementary. And in fact, it is so elementary that PA itself can prove the formal expression of the Theorem, i.e. it can prove $\mathsf{Con} \to \mathsf{G}$ (though it is tedious to check this). But since we know that if PA is consistent, it can't prove G, it follows that if PA is consistent it can't prove Con.

Generalize, and we get the Second Incompleteness Theorem: roughly, a consistent theory T containing enough arithmetic can't prove the corresponding sentence Con_T which codes the claim that T is consistent.

(vi) Why is this interesting? You might say: assuming T is consistent, Con_T is just another truth like G_T which T can't prove. However, there's an important corollary.

If T can't prove itself consistent, then it won't be able to prove that a *stronger* theory which contains it is consistent. For example, assuming PA is consistent, it can't prove that PA is consistent: and that means it can't prove the consistency of a stronger theory like the set theory ZFC in which we can implement Peano arithmetic. And that observation sabotages what's called Hilbert's programme – the hopeful programme of seeking to produce consistency proofs in relatively safe weak theories (like PA) of stronger theories (like ZFC). Or so the story goes.

8.5 Main recommendations on arithmetic, etc.

I hope those arm-waving overviews piqued your interest! But it is time to get down to details. Though if you want a more expansive overview of the territory before you get down to working through full-blown textbooks, then you can look at one of

1. Robert Rogers, *Mathematical Logic and Formalized Theories* (North-Holland, 1971), Chapter VIII, 'Incompleteness, Undecidability' (still quite discursive, very clear).
2. Robert S. Wolf, *A Tour Through Mathematical Logic* (Mathematical Association of America, 2005), Chapter 3, 'Recursion Theory and Computability'; and Chapter 4, 'Gödel's Incompleteness Theorems' (more detailed, requiring more of the reader, though some students do really like this book).

Or if you just want a quick introduction to Gödel in particular, together with a demolition job on some of the wilder conclusions drawn by those with a quarter-understanding of his incompleteness theorem, this is excellent:

3. Torkel Franzén, *Gödel's Theorem: An Incomplete Guide to its Use and Abuse* (A. K. Peters, 2005). John Dawson (who we'll meet again below) writes "Among the many expositions of Gödel's incompleteness theorems written for non-specialists, this book stands apart. With exceptional clarity, Franzén gives careful, non-technical explanations both of what those theorems say and, more importantly, what they do not. No other book aims, as his does, to address in detail the misunderstandings and abuses of the incompleteness theorems that are so rife in popular discussions of their significance. As an antidote to the many spurious appeals to incompleteness in theological, anti-mechanist and post-modernist debates, it is a valuable addition to the literature." Invaluable, in fact!

But now turning to textbooks, how to approach the area? Gödel's 1931 proof of his incompleteness theorem actually uses only facts about the primitive recursive functions. As we noted, these functions are only a subclass of the effectively computable numerical functions. A more general treatment of the effectively computable functions was developed a few years later (by Gödel, Turing and others), and this in turn throws more light on the incompleteness phenomenon. So there's a choice to be made. Do you look at things in roughly the historical order, first introducing just the primitive recursive functions, explaining how they get represented in theories of formal arithmetic, and then learning how to prove initial versions of Gödel's incompleteness theorem – and only then move on to deal with the general theory of computable functions? Or do you explore the general theory of computation first, only turning to the incompleteness theorems later?

My own Gödel books take the first route. But I also recommend alternatives taking the second route.

4. Peter Smith, *Gödel Without (Too Many) Tears**: freely downloadable from logicmatters.net/igt. This is a very short book – just 130 pages – which, after some general introductory chapters, and a little about formal arithmetic, explains the idea of primitive recursive functions, explains the arithmetization of syntax, and then proves Gödel's First Theorem pretty much as Gödel did, with a minimum of fuss. There follow a few chapters on closely related matters and on the Second Theorem. *GWT* is, I hope, very clear and accessible, and is perhaps all you need for a first foray into this area if you don't want (yet) to tangle with the general theory of computation.

However, the more mathematical reader can perhaps jump straight to …

5. Peter Smith, *An Introduction to Gödel's Theorems*** (2nd edition CUP, 2013: now downloadable from logicmatters.net/igt), is three times the length of *GWT* and ranges more widely. It starts by informally exploring various ideas such as effective computability, and then it proves two correspond-

ingly informal versions of the first incompleteness theorem. The next part of the book gets down to work talking about formal arithmetics, developing some of the theory of primitive recursive functions, and explaining the 'arithmetization of syntax'. Then it establishes more formal versions of Gödel's first incompleteness theorem and goes on discuss the second theorem, all in more detail than *GWT*. The last part of the book widens out the discussion to explore the idea of recursive functions more generally, discussing Turing machines and the Church-Turing thesis, and giving further proofs of incompleteness (e.g. deriving it from the 'recursive unsolvability' of the halting problem for Turing machines).

6. Richard Epstein and Walter Carnielli, *Computability: Computable Functions, Logic, and the Foundations of Mathematics* (Wadsworth 2nd edn. 2000: Advanced Reasoning Forum 3rd edn. 2008) is an excellent introductory book on the standard basics, particularly clearly and attractively done. Part I, on 'Fundamentals', covers some background material, e.g. on the idea of countable sets (many readers will be able to speed-read through these initial chapters). Part II, on 'Computable Functions', comes at them two ways: first via Turing Machine computability, and then via primitive recursive and then partial recursive functions, ending with a proof that the two approaches define the same class of effectively computable functions. Part III, 'Logic and Arithmetic', turns to formal theories of arithmetic and the way that the representable functions in a formal arithmetic like Robinson's Q turn out to be the recursive ones. Formal arithmetic is then shown to be undecidable, and Gödelian incompleteness derived. The shorter Part IV has a chapter on Church's Thesis (with more discussion than is often the case), and finally a chapter on constructive mathematics. There are many interesting historical asides along the way.

Those three books should be very accessible to those without much mathematical background: but even more experienced mathematicians should appreciate the careful introductory orientation which they provide. Then next, taking perhaps half-a-step up in mathematical sophistication, we arrive at a quite delightful book:

7. George Boolos and Richard Jeffrey, *Computability and Logic* (CUP 3rd edn. 1990). This is a modern classic, wonderfully lucid and engaging, admired by generations of readers. Indeed, looking at it again in revising this Guide, I couldn't resist some re-reading! It starts with a exploration of Turing machines, 'abacus computable' functions, and recursive functions (showing that different definitions of computability end up characterizing the same class of functions). And then it moves on discuss logic and formal arithmetic (with interesting discussions ranging beyond what is covered in my book or E&C).

   There are in fact two later editions – heavily revised and considerably expanded – with John Burgess as a third author. But I know that I am not the only one to think that these later versions (good though they are) do lose something of the original book's famed elegance and individuality and distinctive flavour. Still, whichever edition comes to hand, do read it! – you will learn a great deal in an enjoyable way.

One comment: none of these books – including my longer one – gives a full proof of Gödel's Second Incompleteness Theorem. That's because it is tediously fiddly to show that the so-called 'derivability conditions' obtain which allow e.g. PA to prove the formal version of the First Theorem, i.e. $\mathsf{Con} \to \mathsf{G}$. If you *really* want all the details, see one of the relevant recommendations in Part III.

8.6 Some parallel/additional reading

Let's start with group of three books at about the same level as those mentioned in the previous section. First, the Open Logic Project now has a good volume on our topics:

8. Jeremy Avigad and Richard Zach, *Incompleteness and Computability: An Open Introduction to Gödel's Theorems***, available at [icomp-open]. Chapters 1 to 5 are on computability and Gödel, covering a good deal in just 120 *very* sparsely printed pages. Avigad and Zach are admirably clear as far as they go – though inevitably, given the length, they have to go pretty briskly. But this could be enough for those who want a short first introduction. And others could well find this very useful revision material, highlighting some basic main themes.

Still, I'd certainly recommend taking a slower tour through more of the sights by following the recommendations in the previous section, or by reading the following excellent book that could well have been an alternative main recommendation:

9. Herbert E. Enderton's relatively short book *Computability Theory: An Introduction to Recusion Theory* (Associated Press, 2011). This is written with attractive zip and lightness of touch (this is a notably more relaxed book than his earlier *Logic*). The first chapter is on the informal Computability Concept. There are then chapters on general recursive functions and on register machines (showing that the register-computable functions are exactly the recursive ones), and a chapter on recursive enumerability. Chapter 5 makes 'Connections to Logic' (including proving Tarski's theorem on the undefinability of arithmetical truth and a semantic incompleteness theorem). The final two chapters push on to say something about 'Degrees of Unsolvability' and 'Polynomial-time Computability'. This is all very nicely and accessibly done.

This book, then, makes an excellent alternative to Epstein & Carnielli in particular: it is, however, a little more abstract and sophisticated, which why I have on balance recommended E&C for many readers. The more mathematical might well prefer Enderton. (By the way, staying with Enderton, I should mention that Chapter 3 of his earlier *A Mathematical Introduction to Logic* (Academic Press 1972, 2002) gives a good brisk treatment of different strengths of formal theories of arithmetic, and then proves the incompleteness theorem first for a formal arithmetic with exponentiation and then – after touching on other issues – shows how to use the β-function trick to extend the theorem to apply to arithmetic without exponentiation. Not the best place to start, but this chapter too could be very useful revision material.)

Thirdly, I have already warmly recommended the following book for its coverage of first-order logic:

10. Christopher Leary and Lars Kristiansen's *A Friendly Introduction to Mathematical Logic*** [friendlylogic]. Chapters 4 to 7 now give a very illuminating double treatment of matters related to incompleteness (you don't have to have read the previous chapters in this book to follow the later ones, other than noting the arithmetical system N introduced in their §2.8). In headline terms that you'll only come fully to understand in retrospect:

    a) L&K's first approach doesn't go overtly via computability. Instead of showing that certain syntactic properties are primitive recursive and showing that all primitive recursive properties can be 'represented' in theories like N (as I do in *IGT*), L&K rely on more directly showing that some key syntactic properties can be represented. This representation result then leads to, inter alia, the incompleteness theorem.

    b) L&K follow this, however, with a general discussion of computability, and then use the introductory results they obtain to prove various further theorems, including incompleteness again.

    This is all presented with the same admirable clarity as the first part of the book on FOL.

There are, of course, many other more-or-less introductory treatments covering aspects of computability and/or incompleteness, and we will return to the topic in Part III of this Guide. For now, I will mention just three further, and rather more individual, books.

First, of the relevant texts in American Mathematical Society's 'Student Mathematical Library', by far the best is

11. A. Shen and N. K. Vereshchagin, *Computable Functions*, (AMA, 2003). This is a lovely, elegant, little book, which can be recommended for giving a differently-structured quick tour through some of the Big Ideas. Well worth reading as a follow-up to a more conventional text.

Next we come to a stand-out book that you should certainly tackle at some point (though I rather suspect that many readers will appreciate it more if they come to it after reading one or more of the main recommendations in the previous section):

12. Raymond Smullyan, *Gödel's Incompleteness Theorems*, Oxford Logic Guides 19 (Clarendon Press, 1992). This is delightully short – under 140 pages – proving some rather beautiful, slightly abstract, versions of the incompleteness theorems. This is a modern classic which anyone with a taste for mathematical elegance will find very rewarding.

To introduce the third book, the first thing to say is that it presupposes *very* little knowledge about sets, despite the title. If you are familiar with the idea that the natural *numbers* can be identified with (implemented as) *finite sets* in a standard way, and with a few other low-level ideas, then you can dive in without further ado to

13. Melvin Fitting's, *Incompleteness in the Land of Sets** (College Publications, 2007). This is a very engaging read, approaching the incompleteness theorem and related results in an unusual but illuminating way. From the book's blurb: "Russell's paradox arises when we consider those sets that do not belong to themselves. The collection of such sets cannot constitute a set. Step back a bit. Logical formulas define sets (in a standard model). Formulas, being mathematical objects, can be thought of as sets themselves – mathematics reduces to set theory. Consider those formulas that do not belong to the set they define. The collection of such formulas is not definable by a formula, by the same argument that Russell used. This quickly gives Tarski's result on the undefinability of truth. Variations on the same idea yield the famous results of Gödel, Church, Rosser, and Post.

    This book gives a full presentation of the basic incompleteness and undecidability theorems of mathematical logic in the framework of set theory. Corresponding results for arithmetic follow easily, and are also given. Gödel numbering is generally avoided, except when an explicit connection is made between set theory and arithmetic. The book assumes little technical background from the reader. One needs mathematical ability [and] a general familiarity with formal logic ..."

And, finally, if only because I've been asked about it such a large number of times, I suppose I should end by also mentioning the (in)famous

14. Douglas Hofstadter, *Gödel, Escher, Bach** (Penguin, first published 1979). When students enquire about this, I helpfully say that it is the sort of book that you will probably really like if you like this kind of book, and you won't if you don't. It is, to say the very least, quirky, idiosyncratic and entirely distinctive. However, as I far as I recall, the parts of the book which touch on techie logical things are in fact pretty reliable and won't lead you astray.

Which is a great deal more than can be said about many popularizing treatments of Gödel's theorems!

8.7 A little history?

If you haven't already done so, *do* read

15. Richard Epstein's brisk and very helpful 28 page 'Computability and Undecidability – A Timeline' which is printed at the very end of Epstein & Carnielli, listed in §8.5.

This will really give you the headline news you initially need. Enthusiasts can find a little more detail about the early days in e.g. Rod Adams's *An Early History of Recursive Functions and Computability** (Docent Press, 2011). But it is a lot more interesting to read

16. John Dawson, *Logical Dilemmas: The Life and Work of Kurt Gödel* (A. K. Peters, 1997).

Not, perhaps, as lively as the Fefermans' biography of Tarski which I mentioned in §7.4 – but then Gödel was such a very different man. Fascinating, though!

9 Beginning set theory, less naively

In Chapter 4, we touched on some elementary concepts and constructions involving sets. We now go further into set theory, though still not beyond the beginnings that any logician really ought to know about. In Part III of the Guide we will return to cover more advanced topics like 'large cardinals', proofs of the consistency and independence of the Continuum Hypothesis, and a lot more besides. But this present chapter concentrates on some core basics.

9.1 Elements of set theory: an overview

Even more, perhaps, than previous overviews, this one may well fall squarely between two stools, being too elementary for mathematicians and not explanatory enough for those readers whose background is purely in philosophy. So if, for one reason or the other, you find that these preliminary remarks aren't particularly helpful for you, then do simply skip on to the next section which gives the main reading recommendations.

(a) If you've not already done so, you now want to get a really firm grip on the key facts about the 'algebra of sets' (concerning unions, intersections, complements and how they interact). You also need to know, inter alia, the basics about powersets, about encoding pairs and other finite tuples using unordered sets, and about Cartesian products, the extensional treatment of relations and functions, the idea of equivalence classes, and how to treat infinite sequences as sets (see Chapter 4).

(b) Moving on, one fundamental early role for set theory was "putting the theory of real numbers, and classical analysis more generally, on a firm foundation". What does this involve?

Natural numbers are finite objects, in the sense that it only takes a finite amount of data to fully specify a particular natural number. Similarly for integers and rational numbers. But not so for real numbers. As is very familiar, a real can rendered e.g. by an infinite sequence of ever-closer rational approximations which need neither terminate. So in theorizing about real numbers we are tangling with the infinite. Set theory gives us a framework for reasoning about such non-finite objects. How?

Assume for the moment, then, that we already have the rational numbers to hand, and let's define the idea of a sequence of ever-closer rational approximations more carefully. A *Cauchy sequence*, then, is an infinite sequence of rationals $s_1, s_2, s_3, \ldots$ which *converges* – i.e. the differences $|s_m - s_n|$ are as small as we want, once we get far enough along the sequence. More carefully, take any $\epsilon > 0$ however small, then for some k, $|s_m - s_n| < \epsilon$ for all $m, n > k$. Now say that two Cauchy sequences $s_1, s_2, s_3, \ldots$ and $s'_1, s'_2, s'_3, \ldots$ are *equivalent* if their members eventually get arbitrarily close – i.e. when we take any $\epsilon > 0$ however small, then for some k, $|s_n - s'_n| < \epsilon$ for all $n > k$. Cauchy identifies real numbers with equivalence classes of Cauchy sequences. So, for Cauchy, $\sqrt{2}$ would be the equivalence class containing sequences of rationals like 1.4, 1.41, 1.414, 1.4142, 1.41421, …, i.e. rationals whose squares approach 2.

Alternatively, dropping the picture of sequential approach, we can identify a real number with a *Dedekind cut*, defined as a (proper, non-empty) subset C of the rationals which (i) is downward closed – i.e. if $q \in C$ and $q' < q$ then $q' \in C$ – and (ii) has no largest member. For example, take the negative rationals together with the positive ones whose square is less than two: these form a cut. Dedekind (more or less) identifies the positive irrational $\sqrt{2}$ with the cut we just defined.

Assuming some set theory, we can now show that – whether defined as cuts on the rationals or defined as equivalence classes of Cauchy sequences of rationals – the real numbers have the desired properties either way. Assuming our set theory is consistent, the resulting theory of the reals can be shown to be consistent too.

We can then go on define functions between real numbers in terms of sets of ordered tuples of reals. But I won't spell this out further here. However, you should get to know something of how the overall story goes, and also get some sense of what assumptions about sets are needed for the story to work to give us a basis for reconstructing classical real analysis. (You will need a number of levels of sets: sets of rationals, and sets of sets of rationals, and sets of sets of sets, and up a few more levels depending on the details.)

(c) Now, as far as construction of the reals and the foundations of analysis are concerned, we could take the requisite set theory – the apparatus of infinite sets, infinite sequences, equivalence classes and the rest – as describing a *superstructure* sitting on top of a given prior basic universe of rational numbers already governed by a prior suite of numerical laws. However, we don't need to do this. For we can in fact *already* construct the rationals and simpler number systems within set theory itself.

For the naturals, pick any set you like and call it '0'. And then consider e.g. the sequence of sets $0; \{0\}; \{\{0\}\}; \{\{\{0\}\}\}; \ldots$. Or alternatively, consider the sequence $0; \{0\}; \{0, \{0\}\}; \{0, \{0\}, \{0, \{0\}\}\}; \{0, \{0\}, \{0, \{0\}\}, \{0, \{0\}, \{0, \{0\}\}\}\}; \ldots$ where at each step after the first we extend the sequence by taking the set of all the sets we have so far. Either sequence then has the structure of the natural-number series. There is a first member; every member has a unique successor (which is distinct from it); different members have different successors; the sequence never circles around and starts repeating. So such a sequence of sets will do as a representation, implementation, or model of the natural numbers (call it what you will).

Let's not get hung up about the best way to describe the situation; we will simply say we have constructed a natural number sequence. And elementary reasoning about sets will show that the familiar arithmetic laws about natural numbers apply to numbers as just constructed (including e.g. the principle of arithmetical induction).

Once we have a natural number sequence we can go on to construct the integers from it in various ways. Here's one. Informally, any integer equals $m - n$ for some natural numbers m, n (to get a negative integer, take $n > m$). So, first shot, we can treat an integer as an ordered pair of natural numbers. But since $m - n = m' - n'$ for lots of m', n', choosing a particular pair of natural numbers to represent an integer involves an arbitrary choice. So, a neater second shot, we can treat an integer as an equivalence class of ordered pairs of natural numbers (where the pairs $\langle m, n\rangle$ and $\langle m', n'\rangle$ are equivalent in the relevant way when $m + n' = m' + n$). Again the usual laws of integer arithmetic can then be proved from basic principles about sets.

Similarly, once we have constructed the integers, we can construct rational numbers in various ways. Informally, any rational equals p/q for integers p, q, with $q \neq 0$. So, first shot, we can treat a rational numbers as a particular ordered pair of integers. Or to avoid making a choice between equivalent renditions, we can treat a rational as an equivalence class of ordered pairs of integers.

We again needn't go further into the details here, though you will want to see them worked through enough to confirm that these can constructions can indeed all be done. The point we want to emphasize now is simply this: once we have chosen an initial object to play the role of 0 – the empty set is the conventional choice! – and once we have a set-building operation which we can iterate sufficiently often, and once we can form equivalence classes from among sets we have already built, we can construct sets to do the work of natural numbers, integers and rationals in standard ways. Hence, we don't need a theory of the rationals prior to set theory before we can go on to construct the reals: *the whole game can be played inside pure set theory.*

(d) Another theme. It is an elementary idea that two sets are equinumerous (have the same cardinality) just if we can match up their members one-to-one, i.e. when there is a one-to-one correspondence, a bijection, between the sets. It is easy to show that the set of even natural numbers, the set of primes, the set of integers, the set of rationals are all *countably* infinite in the sense of being equinumerous with the set of natural numbers.

By contrast, as we showed in §4.1 using a simple diagonal argument, the set of infinite binary strings is not countably infinite. Two corollaries from our diagonal argument:

1. An infinite binary string can be thought of as representing a set of natural numbers, namely the set which contains n if and only if the n-th digit in the string is 1; and different strings represent different sets of naturals. Hence the powerset of the natural numbers, i.e. the set of subsets of the naturals, is also not countably infinite.

2. An infinite binary string can equally well be thought of as representing a real number between 0 and 1 in binary; and different strings represent different reals. So the set of real numbers between 0 and 1 is not countably infinite either – hence neither is the set of all the real numbers.

And now a famous question arises – easy to ask, but (it turns out) extraordinarily difficult to answer. Take an infinite collection of real numbers. It could be equinumerous with the set of natural numbers (like, for example, the set of *real* numbers 0, 1, 2, ...). It could be equinumerous with the set of all the real numbers (like, for example, the set of irrational numbers). But are there any infinite sets of reals of intermediate size (so to speak)? – can there be an infinite subset of real numbers that can't be put into one-to-one correspondence with the natural numbers and can't be put into one-to-one correspondence with all the real numbers either? Cantor conjectured that the answer is 'no'; and this negative answer is known as the Continuum Hypothesis.

Efforts to confirm or refute the Continuum Hypothesis were a major driver in early developments of set theory. We now know the problem is a profound one – the standard axioms of set theory can't settle the hypothesis one way or the other. Is there some attractive and natural additional axiom which will settle the matter? I'll not give a spoiler here! – but exploration of this question takes us way beyond the initial basics of set theory.

(e) The argument that the power set of the naturals isn't equinumerous with the set of naturals can be generalized. Cantor's Theorem tells us that a set is *never* equinumerous with its powerset.[1]

Note, there *is* a bijection between the set A and the set of singletons of elements of A; in other words, there *is* a bijection between A and *part* of its powerset $\mathcal{P}(A)$. But we've just seen that there is no bijection between A and the *whole* of $\mathcal{P}(A)$. Intuitively then, A is smaller in size than $\mathcal{P}(A)$, which will in turn be smaller than $\mathcal{P}(\mathcal{P}(A))$, etc. We now want to develop this intuitive idea of one set's having a smaller cardinal size than another into a general theory about relative cardinal size.

(f) Let's pause at this point to consider the emerging picture.

Starting perhaps from some given urelements – elements which don't themselves have members – we can form sets of them, and then sets of sets, sets of sets of sets, and so on and on: and at each new level, we accumulate more and more sets formed from the urelements and/or the sets formed at earlier levels. At each level, more and more sets are formed. In particular, once we have an infinite number of entities at one level, we get an even greater infinity of entities at the next as we form powersets, and so on up.

[1] Why? Suppose there *is* a bijection f between a set A and its powerset $\mathcal{P}(A)$ (the set of subsets of A). So if x is an element of A, $f(x)$ is an element of $\mathcal{P}(A)$, i.e. is a subset of A. Hence, we can sensibly ask whether x is, as it happens, a member of $f(x)$, and it makes sense to define the set $\{x \in A \mid x \notin f(x)\}$. Being a subset of A, this must be $f(c)$ for some $c \in A$, given that f is a bijection. Now ask: is $c \in f(c)$? We immediately have $c \in f(c)$ if and only if $c \in \{x \in A \mid x \notin f(x)\}$ if and only $c \notin f(c)$. Contradiction.

Now, if we want to be able to apply set-theoretic apparatus in talking about e.g. widgets or wombats or (more seriously!) space-time points, then it might seem that we will want the base level of non-membered elements to be populated with those widgets, wombats or space-time points as the case might be. But for purely mathematical purposes such as reconstructing analysis, it seems that at the base level we only need a single non-membered base-level entity, and it is tidy to think of this as the empty set. So if our concerns are purely mathematical, we can take the whole universe of sets to contain only 'pure' sets (when we look at the members of members of ... members of sets, we never find widgets, wombats or space-time points!). And in fact, given that we can for relevant purposes code for widgets, wombats or space-time points using some kind of numbers-as-sets, our set-theory-for-applications can *still* involve only pure sets. That's why typical introductions to set theory either explicitly restrict themselves to talking about pure sets, or after officially allowing the possibility of urelements promptly ignore them.

(g) Lots of questions arise. Here are two:

1. First, how far can we iterate the 'set of' operation – how high do these levels upon levels of sets-of-sets-of-sets-of-...stack up? Once we have the natural numbers in play, we only need another dozen or so more levels of sets in which to reconstruct 'ordinary' mathematics: but now we are embarked on set theory for its own sake, how far can we go up the hierarchy of levels?

2. Second, at a particular level, how many sets do we get at that level? And a prior question, how do we 'count' the members of infinite sets?

   With finite sets, we not only talk about their relative sizes (larger or smaller), but actually count them and give their absolute sizes by using finite cardinal numbers. These finite cardinals are the natural numbers, which we have learnt can be identified with particular sets. We now want similarly to have a story about the infinite case; we not only want an account of relative infinite sizes but also a theory about infinite cardinal numbers apt for giving the size of infinite collections. Again these infinite cardinals will be identified with particular sets. But how can this story go?

It turns out that to answer both these questions, we need a new notion, the idea of infinite ordinal numbers. We can't say very much about this here, but some arm-waving pointers might be useful.

(h) Let's start rather naively. Here are the familiar natural numbers, but re-sequenced with the evens in their usual order before the odds in *their* usual order:

$$0, 2, 4, 6, \ldots, 1, 3, 5, 7, \ldots.$$

If we use '$\sqsubset$' to symbolize the order-relation here, then $m \sqsubset n$ just in case either (i) m is even and n is odd or else (ii) m and n have the same parity and $m < n$. Note that $\sqsubset$ is a *well-ordering* in the standard sense that it is a linear order and, for any numbers we take, one will be the $\sqsubset$-least.

Now, if we march through the naturals in their new $\sqsubset$-ordering, checking off the first one, the second one, the third one, etc., where does the number 7 come in the order? Plainly, we cannot reach it in any finite number of steps: it comes, in a word, *transfinitely* far along the $\sqsubset$-sequence. So if we want a position-counting number (officially, an *ordinal* number) to tally how far along our well-ordered sequence the number 7 is located, we will need a transfinite ordinal. We will have to say something like this: We need to march through all the even numbers, which here occupy positions arranged exactly like all the natural numbers in their natural order. And then we have to go on another 4 steps. Let's use 'ω' to indicate the length of the sequence of natural numbers in their natural order, and we'll call a sequence structured like the naturals in their natural order an ω-sequence. The evens in their natural order can be lined up one-to-one with the naturals in order, so form another ω-sequence. Hence, to indicate how far along the re-sequenced numbers we find the number 7, it is then tempting to say that it occurs at $\omega + 4$-th place.

And what about the whole sequence, evens followed by odds? How long is it? How might we count off the steps along it, starting 'first, second, third, ...'? After marching along as many steps as there are natural numbers in order to treck through the evens, then – pausing only to draw breath – we have to march on through the odds, again going through positions arranged like all the natural numbers in their natural ordering. So, we have two ω-sequences, put end to end. It is very natural to say that the positions in the whole sequence are tallied by a transfinite ordinal we can denote $\omega + \omega$.

Here's another example. There are familiar maps for coding ordered pairs of natural numbers by a single natural: take, for example, the function which maps m, n to $[m, n] = 2^m(2n + 1) - 1$. And consider the following ordering on these 'pair-numbers' $[m, n]$:

$$[0,0], [0,1], [0,2], \ldots, [1,0], [1,1], [1,2], \ldots, [2,0], [2,1], [2,2], \ldots, \ldots$$

If we now use '$\prec$' to indicate this order, then $[m, n] \prec [m', n']$ just in case either (i) $m < m'$ or else (ii) $m = m'$ and $n < n'$. (This type of ordering is standardly called *lexicographic*: in the present case, compare the dictionary ordering of two-letter words drawn from an infinite alphabet.) Again, $\prec$ is a well-ordering on the natural numbers.

Where does $[5, 3]$ come in this sequence? Before we get to this 'pair' there are already five blocks of the form $[m, 0], [m, 1], [m, 2], \ldots$ for fixed m, each as long as the naturals in their usual order, first the block with $m = 0$, then the block with $m = 1$, and three more blocks, each ω long; and then we have to count another four steps along, tallying off $[5, 0], [5, 1], [5, 2], [5, 3]$. So it is inviting to say we have to count along to the $\omega \cdot 5 + 4$-th step in the sequence to get to the 'pair' $[5, 3]$.

And what about the whole sequence of 'pairs'? We have blocks ω long, with the blocks themselves arranged in a sequence ω long. So this time it is tempting to say that the positions in the whole sequence of 'pairs' are tallied by a transfinite ordinal we can indicate by $\omega \cdot \omega$.

We can continue. Suppose we re-arrange the natural numbers into a new well-ordering like this: take all the numbers of the form $2^l \cdot 3^m \cdot 5^n$, ordered by ordering the triples $\langle l, m, n\rangle$ lexicographically, followed by the remaining naturals in their normal order. We tally positions in *this* sequence by the transfinite ordinal $\omega \cdot \omega \cdot \omega + \omega$. And so it goes.

Note by the way that we have so far been considering just (re)orderings of the familiar set of natural numbers – the sequences are equinumerous, and have the same infinite *cardinal* size; but the well-orders are tallied by different infinite *ordinal* numbers. Or so we want to say.

But is this sort of naive talk of transfinite ordinals really legitimate? Well, it was one of Cantor's great and lasting achievements to show that we can indeed make perfectly good sense of all this.

Now, in Cantor's work the theory of transfinite ordinals is already entangled with his nascent set theory. Von Neumann later cemented the marriage by giving the canonical treatment of ordinals in set theory. And it is via this treatment that students now typically first encounter the arithmetic of transfinite ordinals, some way into a full-blown course about set theory. This approach can, unsurprisingly, give the impression that you have to buy into quite a lot of set theory in order to understand even the basics about ordinals and their arithmetic. However, not so. Our little examples so far are of recursive (re)orderings of the natural numbers – i.e. a computer can decide, given two numbers, which way round they come in the ordering. There is a whole theory of recursive ordinals which talks about how to tally the lengths of such (re)orderings of the naturals, which has important applications e.g. in proof theory. And these tame beginnings of the theory of transfinite ordinals needn't entangle us with the kind of rather wildly infinitary and non-constructive ideas characteristic of modern set theory.

(i) However, here we *are* concerned with set theory, and so our next topic will naturally be von Neumann's very elegant implementation of ordinals in set theory as the 'hereditarily transitive sets'. The basic idea is to define a particular well-ordered sequence of sets – call them the ordinals$_{\mathrm{vN}}$ – and show that any well-ordered collection of objects, however long the ordering, will have the same type of ordering as an initial segment of these ordinals$_{\mathrm{vN}}$. So we can use the ordinals$_{\mathrm{vN}}$ as a universal measuring scale against which to tally the length of any well-ordering.

And at this point, I'll have to leave it to you to explore the details of the construction of the ordinals$_{\mathrm{vN}}$ in the recommended readings. But once we have them available, we can say more about the way that the universe of sets is structured; we can take the levels to be indexed by ordinals$_{\mathrm{vN}}$ (and then assume that for every ordinal there is a corresponding level of the universe).

We can also now define a scale of cardinal size. We noted that well-orderings of different ordinal length can be equinumerous; different ordinals$_{\mathrm{vN}}$ can have the same cardinality. So von Neumann's next trick is to define a cardinal number to be the first ordinal (in the well-ordered sequence of ordinals) in a family of equinumerous ordinals. Again this neat idea we'll have leave for the moment for later exploration. However – and this is an important point – to get this to all work out as we want, in particular to ensure that we can assign any two non-equinumerous sets respective cardinalities κ and λ such that either $\kappa < \lambda$ or $\lambda < \kappa$, we will need the Axiom of Choice. (This is something to keep looking out for in beginning set theory: where do we start to need to appeal to some Choice principle?)

(j) We are perhaps already rather past the point where scene-setting remarks at this level of arm-waving generality can be very helpful. Time to dive into the details! But one final important observation before you start.

The themes we have been touching on can and perhaps should initially be presented in a relatively informal style. But something else that also belongs here near the beginning of your first forays into set theory is an account of the development of axiomatic ZFC (Zermelo-Fraenkel set theory with Choice) as the now standard way of formally regimenting set theory. As you will see, different books take different approaches to the question of just *when* it is best to start getting more rigorously axiomatic, formalizing our set-theoretic ideas.

Now, there's a historical point worth noting, which explains something about the shape of the standard axiomatization. You'll recall from the remarks in §4.1(b) that a set theory which makes the assumption that *every* property has an extension will be inconsistent. So Zermelo set out in an epoch-making 1908 paper to lay down what he thought were the basic assumptions that mathematicians actually needed about sets, while not overshooting and falling into such contradictions. His axiomatization was not, it seems, initially guided by a positive conception of the universe of sets so much as by the desire to keep safe and not assume too much. But in the 1930s, Zermelo himself and especially Gödel came to develop the conception of sets as a hierarchy of levels (with new sets always formed from objects at lower levels, so never containing themselves, and with no end to the levels where we form more sets from what we have accumulated so far, so we never get to a paradoxical set of all sets). This cumulative hierarchy is described and explored in the standard texts. Once this conception is in play, it does invite a more direct and explicit axiomatization as a story about levels and sets formed at levels: however, it was only much later that this positively motivated axiomatization gets spelt out, particularly in what has come to be called Scott-Potter set theory. Most text books stick for their official axioms to the Zermelo approach, hence giving what looks to be a rather unmotivated selection of axioms whose attraction is that they all look reasonably modest and separately in keeping with the hierarchical picture, so unlikely to get us into trouble. In particular the initial recommendations below take this conventional line.

9.2 Main recommendations on set theory

This present chapter is, as advertised, just about the basics of set theory. Even here, however, there are is a very large number of books to choose from, so an annotated Guide will (I hope!) be particularly welcome.

But first, if you want a more expansive 35pp. overview of basic set theory, with considerably more mathematical detail and argument, I think the following chapter (the best in the book?) works pretty well:

1. Robert S. Wolf, *A Tour Through Mathematical Logic* (Mathematical Association of America, 2005), Ch. 2, 'Axiomatic Set Theory'.

And now let's return to a couple of books which I mentioned in §4.2, which you may have skipped past then.

2. Cambridge lecture notes by Tim Button have become incorporated into *Set Theory: An Open Introduction*** (2019) [opensettheory], and this short book is one of the most successful outputs from the Open Logic Project. Its earlier chapters in particular are extremely good, and are very clear on the conceptual motivation for the iterative conception of sets and its relation to the standard ZFC axiomatization. However, things get a bit patchier as the book progresses: later chapters on ordinals, cardinals, and choice, get rather tougher, and might work better (I think) as parallel readings to the more expansive main recommendations I'm about to make. But very well worth looking at.

Also worth mentioning again is a famous 'bare minimum' book (only 104 pp. long), and which could well be extremely useful for someone making a start on exploring fundamental concepts and wanting a short but discursive introduction:

3. Paul Halmos, *Naive Set Theory** (1960: republished by Martino Fine Books, 2011).

However, Halmos doesn't cover all of what I'm counting as belonging to the elements of set theory, and Button can't get into enough detail in his notes, so most readers will want to look instead at one or other of the first two of the following admirable 'entry level' treatments which cover rather more in a bit more depth but still very accessibly:

4. Derek Goldrei, *Classic Set Theory* (Chapman & Hall/CRC 1996). The author taught at the Open University, and wrote specifically for students engaged in remote learning: his book has the friendly subtitle 'For guided independent study'. The result as you might expect – especially if you looked at Goldrei's FOL text mentioned in §5.3 – is exceptionally clear, and it is indeed admirably well-structured for independent self-teaching. Moreover, it is rather attractively written (as set theory books go!). The coverage is very much as as outlined in our overview. And one particularly nice feature is the way the book (unusually?) spends enough time motivating the idea of transfinite ordinal numbers before turning to their now conventional implementation in set theory.

5. Herbert B. Enderton's, *The Elements of Set Theory* (Academic Press, 1977) forms a trilogy along with the author's *Logic* and *Computability* which we have already mentioned in earlier chapters.

   This book again has exactly the coverage we need at this stage. But more than that, it is particularly clear in marking off the informal development of the theory of sets, cardinals, ordinals etc. (guided by the conception of sets as constructed in a cumulative hierarchy) from the ensuing formal axiomatization of ZFC. It is also particularly good and non-confusing about what is involved in (apparent) talk of classes which are too big to be sets – something that can mystify beginners. It is written with a certain lightness of touch and proofs are often presented in particularly well-signposted stages. The last couple of chapters perhaps do get a bit tougher, but overall this really is quite exemplary exposition.

Also starting from scratch, we find two further excellent books which are rather less conventional in style:

6. Winfried Just and Martin Weese, *Discovering Modern Set Theory I: The Basics* (American Mathematical Society, 1996). This covers overlapping ground to Goldrei and Enderton, but perhaps more zestfully and with a little more discussion of conceptually interesting issues. At some places, it is more challenging – the pace can be a bit uneven.

   I like the style a lot, and think it works very well. I don't mean the occasional (slightly laboured?) jokes: I mean the in-the-classroom feel of the way that proofs are explored and motivated, and also the way that teach-yourself exercises are integrated into the text. The book is evidently written by enthusiastic teachers, and the result is very engaging. (The story continues in a second volume.)
7. Yiannis Moschovakis, *Notes on Set Theory* (Springer, 2nd edition 2006). This also takes a slightly more individual path through the material than Goldrei and Enderton, with occasional bumpier passages, and with glimpses ahead. But to my mind, this is very attractively written, and again nicely complements and reinforces what you'll learn from the more conventional books.

Of these two pairs of books, I'd rather strongly advise reading one of the first pair and then one of the second pair.

I will add two more firm recommendations at this level. The first might come as a bit of surprise, as it is something of a 'blast from the past'. But we shouldn't ignore old classics – they can have a lot to teach us even when we have read the more recent books, and this is very illuminating:

8. Abraham Fraenkel, Yehoshua Bar-Hillel and Azriel Levy, *Foundations of Set-Theory* (North-Holland, originally 1958; but you want the revised 2nd edition 1973): Chapters 1 and 2 are the immediately relevant ones.

Both philosophers and mathematicians should appreciate the way this puts the development of our canonical ZFC set theory into some context, and also discusses alternative approaches. Standard textbooks can present our canonical theory in a way that makes it seem that ZFC has to be the One True Set Theory, so it is worth understanding more about how it was arrived at and where some choice points are. This book really is attractively readable, and should be very largely accessible at this early stage. I'm not myself an enthusiast for history for history's sake: but it is very much worth knowing the stories that unfold here.

Now, as I noted in the initial overview section, one thing that every set-theory novice now acquires is the picture of the universe of sets as built up in a hierarchy of stages or levels, each level containing all the sets at previous levels plus new ones (so the levels are cumulative). It is significant that, as Fraenkel et al. makes clear, the picture wasn't firmly in place from the beginning. But the hierarchical conception of the universe of sets is brought to the foreground in

9. Michael Potter, *Set Theory and Its Philosophy* (OUP, 2004). For philosophers and for mathematicians concerned with foundational issues this surely is a 'must read', a unique blend of mathematical exposition (mostly about the level of Enderton, with a few glimpses beyond) and extensive conceptual commentary. Potter is presenting not straight ZFC but a very attractive variant due to Dana Scott whose axioms more directly encapsulate the idea of the cumulative hierarchy of sets. It has to be said that there are passages which are harder going, sometimes because of the philosophical ideas involved, and sometimes because of occasional expositional compression. However, if you have already read a set theory text from the main list, you should have no problems.

9.3 Some parallel/additional reading on standard ZFC

There are so many good set theory books with different virtues, many by very distinguished authors, that I should certainly pause to mention some more. Here then are four more general introductory books, listed in order of publication; each has many things to recommend it to beginners. Browse through to see which might suit your interests and mathematical level.

10. D. van Dalen, H.C. Doets and H. de Swart, *Sets: Naive, Axiomatic and Applied* (Pergamon, 1978). The first chapter covers the sort of elementary (semi)-naive set theory that any mathematician needs to know, up to an account of cardinal numbers, and then takes a first look at the paradox-avoiding ZF axiomatization. This is very attractively and illuminatingly done. (Or at least, the conceptual presentation is attractive – sadly, and a sign of its time of publication, the book seems to have been photo-typeset from original pages produced on electric typewriter, and the result is visually not attractive at all.)

The second chapter carries on the presentation axiomatic set theory, with a lot about ordinals, and getting as far as talking about higher infinities, measurable cardinals and the like. The final chapter considers some applications of various set theoretic notions and principles. Well worth seeking out, if you don't find the typography off-putting.

11. Karel Hrbacek and Thomas Jech, *Introduction to Set Theory* (Marcel Dekker, 3rd edition 1999). This eventually goes a bit further than Enderton or Goldrei (more so in the 3rd edition than earlier ones), and you could – on a first reading – skip some of the later material. Though do look at the final chapter which gives a remarkably accessible glimpse ahead towards large cardinal axioms and independence proofs. Recommended if you want to consolidate your understanding by reading a second presentation of the basics and want then to push on just a bit.

    Jech is a major author on set theory whom we'll encounter again, and Hrbacek once won a AMA prize for maths writing. So, unsurprisingly, this is a very nicely put together book.

12. Keith Devlin, *The Joy of Sets* (Springer, 1979: 2nd edn. 1993). The opening chapters of this book are remarkably lucid and attractively written. The opening chapter explores 'naive' ideas about sets and some set-theoretic constructions, and the next chapter introduces axioms for ZFC pretty gently (indeed, non-mathematicians could particularly like Chs 1 and 2, omitting §2.6). Things then speed up a bit, and by the end of Ch. 3 – some 100 pages into the book – we are pretty much up to the coverage of Goldrei's much longer first six chapters, though Goldrei says more about (re)constructing classical maths in set theory. Some will prefer Devlin's fast-track version. (The rest of the book then covers non-introductory topics in set theory, of the kind we take up again in Part III.)

13. Judith Roitman, *Introduction to Modern Set Theory*** (Wiley, 1990: a 2011 version is available at [roitmanset]. This relatively short, and very engagingly written, book manages to cover quite a bit of ground – we've reached the constructible universe by p. 90 of the downloadable pdf version, and there's even room for a concluding chapter on 'Semi-advanced set theory' which says something about large cardinals and infinite combinatorics. A few quibbles aside, this could make excellent revision material as Roitman is particularly good at highlighting key ideas without getting bogged down in too many details.

Those four books all aim to cover the basics in some detail. The next two books are much shorter, and are differently focused.

14. A. Shen and N. K. Vereshchagin, *Basic Set Theory* (American Mathematical Society, 2002). This is just over 100 pages, and mostly about ordinals. But it is very readable, with 151 'Problems' as you go along to test your understanding. Potentially *very* helpful by way of revision/consolidation.

15. Ernest Schimmerling, *A Course on Set Theory* (CUP, 2011) is perhaps slightly mistitled, if 'course' suggests a comprehensive treatment. This is just 160 pages long, starting off with a brisk introduction to ZFC, ordinals, and cardinals. But then the author explores applications of set theory to other areas of mathematics such as topology, analysis, and combinatorics, in a way that will be particularly interesting to mathematicians. An engaging supplementary read at this level.

Applications of set theory to mathematics are also highlighted in a book in the LMS Student Text series which is worth mentioning at this level:

16. Krzysztof Ciesielski, *Set Theory for the Working Mathematician* (CUP, 1997). This eventually touches on advanced topics in the set theory. But the earlier chapters introduce some basic set theory, which is then put to work in e.g. constructing some strange real functions. So this might well appeal to mathematicians who know some analysis, who could indeed tackle Chs 6 to 8 on the basis of other introductions.

9.4 Further conceptual reflection on set theories

(a) Michael Potter's *Set Theory and Its Philosophy* must be the starting point for philosophical reflections about set theory. In particular, he gives a good account of how our standard set theory emerges from a certain hierarchical conception of the universe of sets as built up in stages. There is also now an excellent more recent exploration of the conceptual basis of set theory in

17. Luca Incurvati, *Conceptions of Set and the Foundations of Mathematics* (CUP, 2020). Incurvati gives more by way of a careful defence of the hierarchical conception of sets and also an unusually sympathetic critique of some rival conceptions and the set theories which they motivate. Engaging, knowledgeable and readable.

Rather differently, if you haven't tackled their book in working on model theory, you will want to look at

18. Tim Button and Sean Walsh's *Philosophy and Model Theory** (OUP, 2018). Now see especially §1.B (on first-order vs second-order ZFC), Ch. 8 (on models of set theory), and perhaps Ch. 11 (more on Scott-Potter set theory).

(b) I will leave further philosophical commentary until the Part III chapter on more advanced set theory, except to mention a short piece by Penelope Maddy, which takes us right back to our starting point when we introduced set theory as giving us a 'foundation' for real analysis. But what does that really mean? Maddy starts by noting "It's more or less standard orthodoxy these days that set theory … provides a foundation for classical mathematics. Oddly enough, it's less clear what 'providing a foundation' comes to." Her opening pages then give a particularly clear and crisp account of what might be meant by talk of foundations in this context. It is *very* well worth reading for orientation:

19. Penelope Maddy, 'Set-theoretic foundations', in A. Caicedo et al., eds., *Foundations of Mathematics* (AMS, 2017), available at [maddy-found] See §1 in particular.

9.5 A little more history?

As already shown in the recommended book by Fraenkel, Bar-Hillel and Levy, the history of set theory is a long and tangled story, fascinating in its own right and conceptually illuminating too. José Ferreirós has an impressive book *Labyrinth of Thought: A History of Set Theory and its Role in Modern Mathematics* (Birkhäuser 1999). But that's more than most readers are likely to want. But you will find some of the headlines here, worth chasing up especially if you didn't read the book by Fraenkel et al.:

20. José Ferreirós, 'The Early Development of Set Theory', *The Stanford Encyclopaedia of Philosophy*, available at [sep-devset].

This article has references to many more articles, like Kanimori's fine piece on 'The mathematical development of set theory from Cantor to Cohen'. But you might to need to be on top of rather more set theory before getting to grips with *that*.

9.6 Postscript: Other treatments

What else is there? There is a classic book by Azriel Levy with the inviting title *Basic Set Theory** (Springer 1979, republished by Dover 2002). However, while this is 'basic' in the sense of not dealing with topics like forcing, this *is* quite an advanced-level treatment of the set-theoretic fundamentals. So let's return to it in Part III.

András Hajnal and Peter Hamburger have a book *Set Theory* (CUP, 1999) which is also in the LMS Student Text series. They nicely bring out how much of the basic theory of cardinals, ordinals, and transfinite recursion can be developed in a semi-informal way, before introducing a full-fledged axiomatized set theory. But I think Enderton or van Dalen et al. do this better. The second part of this book is on more advanced topics in combinatorial set theory.

George Tourlakis's *Lectures in Logic and Set Theory, Volume 2: Set Theory* (CUP, 2003) has been recommended to me a number of times. Although this is the second of two volumes, it is a stand-alone text. Indeed Tourlakis goes as far as giving a 100 page outline of the logic covered in the first volume as the long opening chapter in this volume. Assuming you have already studied FOL, you can initially skip this chapter, consulting if/when needed. That still leaves over 400 pages on basic set theory, with long chapters on the usual axioms, on the Axiom of Choice, on the natural numbers, on order and ordinals, and on cardinality. (The final chapter on forcing should be omitted at this stage, and strikes me as less clear than what precedes it.)

As the title suggests, Tourlakis aims to retain something of the relaxed style of the lecture room, complete with occasional asides and digressions. And as the page length suggests, the pace is quite gentle and expansive, with room to pause over questions of conceptual motivation etc. However, there is a certain quite excessive and unnecessary formalism that many (most?) will find off-putting, and which slows things right down. Simple constructions and results therefore take a *very* long time to arrive. We don't meet the von Neumann ordinals for three hundred pages, and we don't get to Cantor's theorem on the uncountability of $\mathcal{P}(\omega)$ until p. 455! So while this book might be worth dipping into for some of the motivational explanations, I can't myself recommend it overall.

Finally here, I'll mention another more recent text from the same publisher, Daniel W. Cunningham's *Set Theory: A First Course* (CUP, 2016). But this doesn't strike me as a particularly friendly introduction. As the book progresses, it turns into pages of old-school Definition/Lemma/Theorem/Proof with rather too little commentary; key ideas seem often to be introduced in a phrase, without much discursive explanation. Readers who care about the logical niceties will also raise their eyebrows at the author's over-causal way with use and mention, or e.g. the too-typically hopeless passage about replacing variables with values on p. 14). And this isn't just being pernickety: what exactly are we to make of the claim on p. 31 that a class is "any collection of the form $\{x\colon \varphi(x)\}$"? Not recommended to logicians of a sensitive disposition!

Index