Beginning Mathematical Logic: A Study Guide

There are many wonderful introductory texts on mathematical logic, but there are also many not-so-useful books. So how do you find your way around the very large literature old and new, and how do you choose what to read? *Beginning Mathematical Logic* provides the necessary guide. It introduces the core topics and recommends the best books for studying these topics enjoyably and effectively. This will be an invaluable resource both for those wanting to teach themselves new areas of logic and for those looking for supplementary reading before or during a university course.

PETER SMITH was formerly Senior Lecturer in Philosophy at the University of Cambridge, and taught logic for more years than he cares to remember. His books include *Explaining Chaos* (1998), *An Introduction to Formal Logic* (2003; 2020), *An Introduction to Gödel's Theorems* (2007; 2013), and *Gödel Without (Too Many) Tears* (2020). He was also editor of *Analysis* for a dozen years.

The page intentionally left blank

Beginning Mathematical Logic

A Study Guide

Peter Smith

LOGIC MATTERS

Published by Logic Matters, Cambridge

A paperback copy of this book is available by print-on-demand from Amazon: ISBN 978-1-91690633-4.

Additional resources are available at www.logicmatters.net/tyl where the latest version of this PDF will always be found.

This version: January 5, 2024.

Since this book is not produced by a publisher with a marketing department, your university librarian will not get to know about it in the usual way. You will therefore need to ask them to order a printed copy for the library. Please do so, if you find this book useful.

Contents

The page intentionally left blank

Preface

This is not another textbook on mathematical logic: it is a Study Guide, a book mostly about textbooks on mathematical logic. Its purpose is to enable you to locate the best resources for teaching yourself various areas of logic, at a fairly introductory level. Inevitably, given the breadth of its coverage, the Guide is rather long: but don't let that scare you off! There is a great deal of signposting and there are also explanatory overviews to enable you to pick your way through and choose the parts which are most relevant to you.

Beginning Mathematical Logic is a descendant of my much-downloaded *Teach Yourself Logic*. The new title highlights that the Guide focuses mainly on the core mathematical logic curriculum. It also signals that I do not try to cover advanced material in any detail.

The first chapter says more about who the Guide is intended for, what it covers, and how to use it. But let me note straightaway that most of the main reading recommendations do indeed point to published *books*. True, there are quite a few relevant sets of lecture-notes that university teachers have made available online. Some of these are excellent. However, they do tend to be terse, and often *very* terse (as entirely befits material originally intended to support a lecture course). They are therefore usually not as helpful as fully-worked-out book-length treatments, at least for students needing to teach themselves.

So where can you find the titles mentioned here? I suppose I ought to pass over the issue of downloading books from certain very well-known and extremely well-stocked copyright-infringing PDF repositories. That's between you and your conscience (though almost all the books are available to be sampled there). Anyway, many do prefer to work from physical books. Most of these titles 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 I'm not assuming that you will be buying the recommended books, I have *not* made cost or being currently in print a significant consideration. However, I have marked with a star* books that are available new or second-hand relatively inexpensively (or at least are unusually good value for the length and/or importance of the book). When e-copies of books are freely and legally available, links are provided. Where journal articles or encyclopaedia entries have been recommended, these can almost always be freely downloaded, and again I give links.

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 extraordinarily lucky I was. In writing this Study Guide and making it readily available, I am trying to give a little back by way of heartfelt thanks. I hope you find it useful.[1]

A note on this reprint I have corrected more than forty minor but annoying misprints. I have also quietly tidied Ch. 2 and elsewhere rephrased a few infelicitous sentences. However, I haven't yet significantly revised either my overviews of the various topics or the corresponding reading recommendations: that's work for a second edition.

Perhaps, though, I should mention two books published since the original printing in 2022, substantial texts aimed at students which promised to be of particular interest to readers of this Study Guide. Joseph Mileti's *Modern Mathematical Logic* (CUP, 2023) ranges widely over first-order logic, some model theory, set theory, arithmetic and computability. Despite the title, however, the book's approach is rather conventional, and its chapters on the various topic aren't to be preferred to the existing recommendations, though some could make useful supplementary reading: for more, see tinyurl.com/mileti-mml. Jeremy Avigad's *Mathematical Logic and Computation* (CUP, 2023) on the other hand is much more interesting, and I more warmly recommend a number of sections in this book note: tinyurl.com/avigad-mlc.

[1] I owe much to the kindness of strangers: many thanks, then, to all those who commented on earlier versions of *Teach Yourself Logic* and *Beginning Mathematical Logic* over more than a decade, far too many to list here. I am particularly grateful though to Rowsety Moid for all his suggestions over the years, and for a lengthy set of comments which led to many last-minute improvements.

Further comments and suggestions for a possible revised edition of this Guide will always be most welcome.

Athena's familiar at the very end of the book is borrowed from the final index page of the 1794 Clarendon Press edition of Aristotle's *Poetics*, with thanks to McNaughtan's Bookshop, Edinburgh.

1 The Guide, and how to use it

Who is this Study Guide for? What does it cover? At what level? How should the Guide be used? And what background knowledge do you need, in order to make use of it? This preliminary chapter explains.

1.1 Who is the Guide for?

It is a depressing phenomenon. Relatively few mathematics departments have undergraduate courses on mathematical logic. And serious logic is taught less and less in philosophy departments too.

Yet logic itself remains as exciting and rewarding a subject as it ever was. So how is knowledge to be passed on if there are not enough courses, or if there are none at all? It seems that many will need to teach themselves 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. There's more than a sixty year span of publications which remain relevant, which means that there are hundreds of good books to choose from.

That's why students – whether mathematicians or philosophers – wanting to learn some logic by self-study will need a Guide like this if they are to find their way around the very large literature old and new, with the aim of teaching themselves enjoyably and effectively. And even those fortunate enough to be offered courses might very well appreciate advice on entry-level texts which they can usefully read in preparation or in parallel.

There are other students too who will rightly have interests in areas of logic, e.g. theoretical linguists and computer scientists. But I haven't really kept them much in mind while putting together this Guide.

1.2 The Guide's structure

There is another preliminary chapter after this one, Chapter 2 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). But then we start covering the usual mathematical logic curriculum, at roughly an upper undergraduate level.

The standard menu of core topics has remained fairly fixed ever since e.g. Elliott Mendelson's justly classic *Introduction to Mathematical Logic* (1st edn., 1964), and this menu is explored in Chapters 3 to 7. The following four chapters then look at other logical topics, still at about the same level. The final chapter of the Guide glances ahead at more advanced readings on the core areas, and briefly gestures towards one last topic.

(a) In more detail, then,

Chapter 3 discusses *classical first-order logic* (FOL), which is at the fixed centre of any mathematical logic course.

The remaining chapters all depend on this crucial one and assume some knowledge of it, as we discuss the use of classical FOL in building formal theories, or we consider extensions and variants of this logic.

Now, there is one extension worth knowing just a little about straight away (in order to understand some themes touched on in the next few chapters). So:

Chapter 4 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 start work on the topics of the following three key chapters in whichever order you choose:

Chapter 5 introduces a modest amount of *model theory* which, roughly speaking, explores how formal theories relate to the structures they are about.

Chapter 6 looks at one family of formal theories, i.e. *formal arithmetics*, and explores the theory of computable arithmetical functions. We arrive at proofs of epochal results such as Gödel's incompleteness theorems.

Chapter 7 is on *set theory* proper – starting with constructions of number systems in set theory, then examining basic notions of cardinals and ordinals, the role of the axiom of choice, etc. We then look at the standard formal axiomatization, i.e. first-order ZFC (Zermelo–Fraenkel set theory with the Axiom of Choice), and also nod towards alternatives.

Now, as well as second-order logic, there is another variant of FOL which is often mentioned in introductory mathematical logic texts, and that you will want to know something about at this stage. So

Chapter 8 introduces *intuitionistic logic*, which drops the classical principle that, whatever proposition we take, either it or its negation is true. But why might we want to do that? What differences does it make?

And this topic can't really be sharply separated from another whole area of logic which can be under-represented in many textbooks; that is why

Chapter 9 takes a first look at *proof theory*. OK, this is a pretty unhelpful label given that most areas of logic deal with proofs! – but it conventionally points to a cluster of issues about the structure of proofs and the consistency of theories, etc.

(b) Now, a quick glance at e.g. the entry headings in *The Stanford Encyclopedia of Philosophy* reveals that philosophers have been interested in a wide spectrum of other logics, ranging far beyond classical and intuitionistic versions of FOL and their second-order extensions. And although this Guide – as its title suggests – is mainly focussed on core topics in mathematical logic, it is worth pausing to consider just a few of those variant types of logic.

First, in looking at intuitionist logic, you will already have met a new way of thinking about the meanings of the logical operators, using so-called 'possible-world semantics'. We can now usefully explore this idea further, since it has many other applications. So:

Chapter 10 discusses *modal logics*, which deploy possible-world semantics, initially to deal with various notions of necessity and possibility. In general, these modal logics are perhaps of most interest to philosophers. However, there is one particular variety which any logician should know about, namely *provability logic*, which (roughly speaking) explores the logic of operators like 'it is provable in formal arithmetic that . . . '.

Second, standard FOL (classical or intuitionistic) can be criticized in various ways. For example, (1) it allows certain arguments to count as valid even when the premisses are irrelevant to the conclusion; (2) it is not as neutral about existence assumptions as we might suppose a logic ought to be; and (3) it can't cope naturally with terms denoting more than one thing like 'Russell and Whitehead' and 'the roots of the quintic equation E'. It is worth saying something about these supposed shortcomings. So:

Chapter 11 discusses so-called relevant logics (where we impose stronger requirements on the relevance of premisses to conclusions for valid arguments), free logics (logics free of existence assumptions, where we no longer presuppose that e.g. names in an interpreted formal language always actually name something), and plural logics (where we can e.g. cope with plural terms).

For reasons I'll explain, the first two of these variant logics are mostly of concern to philosophers. However, any logician interested in the foundations of mathematics should want to know more about the pros and cons of dealing with talk about pluralities by using set theory vs second-order logic vs plural logic.

(c) How are these chapters from Chapter 3 onwards structured?

Each starts with one or more overviews of its topic area(s). These overviews are not full-blown tutorials or mini encyclopedia-style essays – they are simply intended to give some preliminary orientation, with some rough indications of what the individual chapters are about. They should enable you to choose which topics you want to pursue.

I don't pretend that the level of coverage in the overviews is uniform. And if you already know something of the relevant topic, or if these necessarily brisk remarks sometimes mystify, feel very free to skim or skip as much you like.

Overviews are then followed by the key section, giving a list of main recommended texts for the chapter's topic(s), put into what strikes me as a sensible reading order.

I next offer some suggestions for alternative/additional reading at about the same level or at only another half a step up in difficulty/sophistication.

And because it can be quite illuminating to know just a little of the background history of a topic, most chapters end with a few suggestions for reading on that.

(d) This is primarily a Guide to *beginning* mathematical logic. So the recommended introductory readings in Chapters 1 to 11 won't take you *very* far. But they should be more than enough to put you in a position from which you can venture into rather more advanced work under your own steam. Still, I have added a final chapter which looks ahead:

Chapter 12 offers suggestions for those who want to delve further into the topics of some earlier core chapters, in particular looking again at model theory, computability and arithmetic, set theory, and proof theory. Then I add a final section on a new topic, type theories and the lambda calculus, a focus of much recent interest.

Very roughly, if the earlier chapters at least begin at undergraduate level, this last one is definitely at graduate level.

1.3 Strategies for self-teaching from logic books

(a) As I said in the Preface, one major reason for the length of this Guide is its breadth of coverage. But there is another significant reason, connected to a point which I now want to highlight:

> I very strongly recommend tackling a new area of logic by reading a variety of texts, ideally a series of books which *overlap* in level (with the next one in the series covering some of the same ground and then pushing on from the previous one).

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 number of different treatments, perhaps at rather different levels. The multiple overlaps in coverage in the reading lists in later chapters, which contribute to making 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 as to 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, choosing books that work for you. But do also cultivate the crucial 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 and levels of approach.

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

> Mathematics is, as they say, not 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.

1.4 Choices, choices

How have I decided which texts to recommend?

An initial point. If I were choosing a textbook around which to shape a lecture course on some area of mathematical logic, I would no doubt be looking at many of the same books that I mention later; but my preference-rankings could well be rather different. So, to emphasize, the main recommendations in this Guide are for books which I think should be particularly good for *self-studying* logic, without the benefit of expansive classroom introductions and additional explanations.

Different people find different expository styles congenial. What is agreeably discursive for one reader might be irritatingly slow-moving for another. For myself, I do particularly like books that are good at explaining the ideas behind the various formal technicalities while avoiding needless early complications, excessive hacking through routine detail, or misplaced 'rigour'. So I prefer a treatment that highlights intuitive motivations and doesn't rush too fast to become too abstract: this is surely what we particularly 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 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 be very widely agreed to have significant virtues (even if other logicians would have different favourites).

1.5 So what do you need to bring to the party?

There is no specific knowledge you need before tackling the main recommended books on FOL. And in fact none of the more introductory books recommended in other chapters except the last requires very much 'mathematical maturity'. So mathematics students from mid-year undergraduates up should be able to just dive in and explore.

What about philosophy students without any mathematical background? It will certainly help to have done an introductory logic course based on a book at the level of my own *Introduction to Formal Logic** (2nd edition, CUP, 2020; now freely downloadable from logicmatters.net/ifl), or Nicholas Smith's excellent *Logic: The Laws of Truth* (Princeton UP 2012). And non-mathematicians could very usefully broaden their informal proof-writing skills by also looking at this much-used and much-praised book:

> 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; at least skip and skim through the book, taking what you need (perhaps paying special attention to the chapter on mathematical induction). [1]

1.6 Two notational conventions

Finally, let me highlight two points of notation.

First, it is helpful to adopt here the following convention for distinguishing two different uses of letters as variables:

[1] For a much less conventional book than Velleman's, with a different emphasis, some might also be both instructed and entertained by Joel David Hamkins, *Proof and the Art of Mathematics** (MIT Press, 2020). This is attractively written, though it is occasionally uneven in level and tone. Readers with very little pre-existing mathematical background could enjoy dipping into this. Lots of striking and memorable examples.

> *Italic letters*, as in A, F, n, x, will always be used just as part of our informal *logicians' English*, typically as place-holders or in making generalizations. Occasionally, *Greek capital letters* will also be used equally informally for sets (in particular, for sets of sentences).
>
> Sans-serif letters by contrast, as in $\mathsf{P}, \mathsf{F}, \mathsf{n}, \mathsf{x}$, are always used as symbols belonging to some particular formal language, an artificial language cooked-up by logicians.

For example, we might talk in logician's English about a logical formula being of the shape $(A \vee B)$, using the italic letters as place-holders for formal sentences. And then $(\mathsf{P} \vee \mathsf{Q})$, a formula from a particular logical language, could be an instance of that form, with these sans-serif letters being sentences of the relevant language. Similarly, $x + 0 = x$ might be an equation of ordinary informal arithmetic, while $\mathsf{x} + \mathsf{0} = \mathsf{x}$ will be an expression belonging to a formal theory of arithmetic.

Our second convention, just put into practice, is that we will *not* in general be using quotation marks when mentioning symbolic expressions. Logicians can get *very* pernickety, and insist on the use of quotation marks in order to make it extra clear when we are mentioning an expression of, say, formal arithmetic in order to say something about that expression itself as opposed to using it to make an arithmetical claim. But in the present context it is unlikely you will be led astray if we just leave it to context to fix whether a symbolic expression is being mentioned rather than put to use (though I do put mentioned single lower-case letters in quotes when it seems helpful, just for ease of reading).

2 A very little informal set theory

Notation, concepts and constructions from entry-level set theory are very often presupposed in elementary mathematical texts – including some of the introductory logic texts mentioned in the next few chapters, even before we get round to officially studying set theory itself. If the absolute basics aren't already familiar to you, it is worth pausing to get initially acquainted at an early stage.

In §2.1, then, I note the little that you should ideally know about sets here at the outset. It really isn't much! And for now, we proceed 'naively' – i.e. we proceed quite informally, and will just assume that the various constructions we talk about are permitted: §2.2 says a bit more about this naivety. §2.3 then gives recommended readings on basic informal set theory for those who need them.

Finally, in §2.4 I note that, while the use of set-talk in elementary contexts is conventional, in many cases it can in fact be eliminated without significant loss.

2.1 Sets: a very quick checklist

(i) Let's assume familiarity with basic notation. We can specify a small set just by listing its members, as in $\{a, b, c, d\}$; but otherwise we use so-called set-builder notation, as in $\{x \mid \varphi(x)\}$ which denotes the set of things which satisfy the condition φ.

Crucially, we must distinguish set-membership from the subset relation (notationally, $\in$ vs $\subseteq$). So, for example, $a \in \{a\}$ but not $a \subseteq \{a\}$.

We also need the idea of the union and intersection of two sets, together with the notion of the powerset of A, i.e. the set of all subsets of A.

(ii) Sets are in themselves unordered; but we often need to work with ordered pairs, ordered triples, etc.

Use '$\langle a, b\rangle$' – or simply '(a, b)' – for the ordered pair, first a, then b. We can implement ordered pairs using unordered sets in various ways: all we need is some definition which ensures that $\langle a, b\rangle = \langle a', b'\rangle$ if and only if $a = a'$ and $b = b'$. The following is standard: $\langle a, b\rangle =_{\text{def}} \{\{a, b\}, \{a\}\}$.

Once we have ordered pairs available, we can use them to implement ordered triples: for example, define $\langle a, b, c\rangle$ as $\langle\langle a, b\rangle, c\rangle$. We can similarly define quadruples, and n-tuples for larger n.

(iii) 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. Hence $A \times B$ is $\{\langle x, y\rangle \mid x \in A \;\&\; y \in B\}$. Cartesian products of n sets are defined as sets of n-tuples in the obvious way.

Next, we add to these basics the standard set-theoretic treatment of relations and functions:

(iv) 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 $A \times B$.

Similarly, 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 many purposes, it is mostly harmless to simply *identify* a property or relation with its extension-as-a-set.

(v) 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 $y \in B$) such that $f(x) = y$. Similarly for n-place functions. Again, for many purposes, we can harmlessly identify a function with its graph.

(vi) Two sets are equinumerous 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. A set is countably infinite if and only if it is equinumerous with the natural numbers.

It is almost immediate that *there are infinite sets which are not countably infinite*. A simple example is the set of infinite binary strings. Why so? If we take any countably infinite list of such strings, we can always define another infinite binary string which 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.

This is just the beginning of a story about how sets can have different infinite 'sizes' or cardinalities. Cantor's Theorem tells us that the power set of A is always bigger than A (we can't put the members of the powerset of A in one-one correspondence with the members of A). But at this stage you need to know little more than that bald fact: further elaboration can wait.

So far, so very elementary. But there's another idea that you should also meet sooner rather than later, so that you recognize any passing references to it:

(vii) The Axiom of Choice, in one version, says that, given an infinite family of sets, there is a corresponding choice function – i.e. a function which 'chooses' a single member from each set in the family. 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 specifying its members.

At this stage you 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.

2.2 A note about naivety

Evidently, the set of musketeers {Athos, Porthos, Aramis} is not another musketeer and so isn't a member of itself. Likewise, the set of prime numbers isn't itself a prime number, so again isn't a member of itself. We'll say that a set which is similarly not a member of itself is *normal*. Now we ask: is there a set R whose members are all and only the normal sets?

No. For if there were, it would be normal if and only if it was a member of itself and hence wasn't normal – contradiction! The putative set R is, in some sense, 'too big' to exist. Hence, if we overshoot and naively 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 (this is the upshot of 'Russell's paradox').

Now, some people use 'naive set theory' to mean, specifically, a theory which makes that simple but hopeless assumption that *any* property at all has a set as its extension. As we've just seen, naive set theory in *this* sense is inconsistent.

But for many others, 'naive set theory' just means set theory developed informally, without rigorous axiomatization, but guided by unambitious low-level principles (such as that we can always form intersections or powersets from given sets). And in this different second sense, a modicum of naive set theory is exactly what you need here at the outset. When we turn to set-theory proper in Chapter 7 we will proceed less naively!

2.3 Recommendations on informal basic set theory

If you are a mathematics student, then the ideas on our checklist ought already to be *very* familiar, e.g. from those introductory chapters or appendices you so often find in mathematics texts. A particularly good example which you can use to consolidate ideas if you are a bit rusty is

1. James R. Munkres, *Topology* (Prentice Hall, 2nd edition, 2000). Chapter 1, 'Set Theory and Logic'. This tells you very clearly about basic set-theoretic concepts, up to countable vs uncountable sets and the axiom of choice (plus a few other things worth knowing about).

If on the other hand you are a philosophy student who hasn't been very well brought up, you could find the following very helpful. It is expressly written for non-mathematicians and is extremely accessible:

2. David Makinson, *Sets, Logic and Maths for Computing* (Springer, 3rd edn 2020), Chapters 1 to 3.

   Chapter 1 reviews basic facts about sets. Chapter 2 is on relations. Chapter 3 is on functions. This too can be warmly recommended – though you might want to supplement it by following up the reference to Cantor's Theorem.

But the stand-out recommendation for those who need it is:

3. Tim Button, *Set Theory: An Open Introduction* (Open Logic Project), Chapters 1–5. Available at tinyurl.com/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 very warmly recommended.

Note, Makinson doesn't mention the Axiom of Choice at all. And while Button does eventually get round to Choice in his Chapter 16, 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:

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

2.4 Virtual classes, real sets

An afterword. According to Cantor, a set is a unity, a single thing in itself over and above its members. But if *that* is the guiding idea, then it is worth noting that *a good deal of elementary set talk in mathematics can in effect be treated as just a handy façon de parler*. Yes, it is a useful and familiar idiom for talking about many things at once; but in elementary contexts apparent talk of a set of Xs is often not really intended to carry any serious commitment to there being any *additional* object, a set, over and above those Xs. On the contrary, in such contexts, apparent talk about a *set of Fs* can very often be paraphrased away into more direct talk about those Fs themselves, without any loss of content.

Here is just one example from elementary logic. It is usual to say something like this: (1) "A set of formulas Γ logically entails the formula A if and only if any valuation which makes every member of Γ true makes A true too". Don't worry for now about the talk of valuations: just note that the reference to a *set* of formulas and its *members* is arguably doing no real work here. It would do just as well to say (2) "The formulas G logically entail A if and only if every valuation which makes those formulas G all true makes A true too" (swapping the plurally referring term 'G' for the singular term 'Γ'). The set version (1) adds nothing relevantly important to the plural version (2).

When set talk can be paraphrased away like this, we are only dealing with – as they say – mere *virtual classes*.

One source for this terminology 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 eventually 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.

The distinction being made here is an old one. Here is Paul Finsler, writing 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 hangs on the particular terminology, 'classes' vs 'sets'. What matters (or will eventually matter in at least some cases) is the distinction between non-committal, eliminable, talk – talk of merely virtual sets/classes – and uneliminable talk of sets as entities in their own right. And here's a general suggestion: when you encounter talk of a set of Fs (outside set theory itself) it is well worth asking yourself: is the idiom doing any serious work here or does it just provide us with a brisk way of talking about many things at once, the Fs?

3 First-order logic

Now let's get down to business!

This chapter begins with a two-stage overview in §§3.1, 3.2 of classical first-order logic, FOL, which is the starting point for any mathematical logic course. (Why 'classical'? Why 'first-order'? All will eventually be explained!)

At this level, the most obvious difference between various treatments of FOL is in the choice of proof-system: so §3.3 comments on two main options.

Then §3.4 highlights the main self-study recommendations. These are followed by some suggestions for parallel and further reading in §3.5. And after the short historical §3.6, this chapter ends with §3.7, a postscript commenting on some other books, mostly responding to frequently asked questions.[1]

3.1 Propositional logic

(a) 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 (including the ordinary language of informal mathematics) these logical operators work in surprisingly complex ways, introducing the kind of obscurities and possible ambiguities we certainly 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 simplified, tightly-disciplined, languages. For Aristotle, his regimented language was a fragment of very stilted Greek; for us, our regimented languages are entirely artificial formal constructions. But either way, the plan is that we tackle a stretch of reasoning by reformulating it in a suitable regimented language with much 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. First, we clarify the intended structure of the original

[1] A note to philosophers. If you *have* carefully read a substantial introductory logic text for philosophers such as Nicholas Smith's, or even my own, you will already be familiar with (versions of) a fair amount of the material covered in this chapter. However, in following up the readings for this chapter, 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.

argument by rendering it into an unambiguous simplified/formalized language. Second, there's the separate business of assessing the validity of the resulting regimented argument.

In exploring FOL, then, we will use appropriate formal languages which contain, in particular, tidily-disciplined surrogates for the propositional connectives *and*, *or*, *if*, *not* (standardly symbolized $\wedge$, $\vee$, $\to$, $\neg$), 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 helpful to develop FOL in two main stages:

(I) We start by introducing languages whose built-in logical apparatus comprises just the propositional connectives, and then 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.

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

So let's have a little more detail about stage (I) in this section, and then we'll turn to stage (II) in the next section.

(b) We first look, then, at the *syntax* of propositional languages, defining what count as the well-formed formulas (wffs) of such languages.

We start with a supply of propositional 'atomic' wffs, as it might be $\mathsf{P}, \mathsf{Q}, \mathsf{R}, \ldots$, and a supply of logical operators, typically $\wedge$, $\vee$, $\to$ and $\neg$, plus perhaps the always-false absurdity constant $\bot$. We then have rules for building 'molecular' wffs, such as *if A and B are wffs, so is* $(A \to B)$.

If you have already encountered languages of this kind, you now need to get to know how to prove various things about them that seem obvious and that you perhaps previously took for granted – for example, that 'bracketing works' to block ambiguities like $\mathsf{P} \vee \mathsf{Q} \wedge \mathsf{R}$, so every well-formed formula has a unique unambiguous parsing.

(c) On the *semantic* side, we need the idea of a *valuation* for a propositional language.

We start with an assignment of truth-values, *true* vs *false*, to the atomic formulas, the basic building blocks of our languages. We now appeal to the 'truth-functional' interpretation of the connectives: we have rules like $(A \to B)$ *is true if and only if A is false or B is true or both* which determine the truth-value of a complex wff as a function of the truth-values of its constituents. And these rules then fix that any wff – however complex – is determined to be either definitely true or definitely false (one or the other, but not both) on any particular valuation of its atomic components. This core assumption is distinctive of *classical two-valued semantics*.

(d) Even at this early point, questions arise. For example, how satisfactory is the representation of an informal conditional *if P then Q* by a formula $\mathsf{P} \to \mathsf{Q}$ which uses the truth-functional arrow connective? And why restrict ourselves to just a small handful of truth-functional connectives?

You don't want to get too entangled with the first question, though you do need to find out why we represent the conditional in FOL in the way we do. As for the second question, it's an early theorem that every truth-function can in fact be expressed using just a handful of connectives.

(e) Now a crucial pair of definitions (we start using 'iff' as standard shorthand for 'if and only if'):

> A wff A from a propositional language is a *tautology* iff it is true on any assignment of values to the relevant atoms.
>
> A set of wffs Γ *tautologically entails* A iff any assignment of values to the relevant atoms which makes all the sentences in Γ true makes A true too.

So the notion of tautological entailment aims to regiment the idea of an argument's being logically valid in virtue of the way the connectives appear in its premisses and conclusion.

You will need to explore some of the key properties of this semantic entailment relation. And note that in this rather special case, we can mechanically determine whether Γ entails A, e.g. by a 'truth table test' (at least so long as there are only finitely many wffs in Γ, and hence only finitely many relevant atoms to worry about).

(f) Different textbook presentations filling out steps (b) to (e) can go into different levels of detail, but the basic story remains much the same. However, now the path forks. For the usual next topic will be a *formal deductive system* in which we can construct step-by-step derivations of conclusions from premisses in propositional logic. There is a variety of such systems to choose from, and I'll mention no less than five main types in §3.3.

Different proof systems for classical propositional logic will (as you'd expect) be equivalent – meaning that, given some premisses, we can derive the same conclusions in each system. However, the systems do differ considerably in their intuitive appeal and user-friendliness, as well as in some of their more technical features. Note, though: apart from 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.

In due course, the educated logician will want to learn at least a little about the various types of proof system – at the minimum, you should eventually get a sense of how they respectively work, and come to appreciate the interrelations between them. But here – as is usual when starting out on mathematical logic – we look in particular at *axiomatic* logics and one style of *natural deduction* system.

(g) At this point, then, we will have two quite different ways of defining what makes for a deductively good argument in propositional classical logic:

> We said that a set of premisses Γ *tautologically entails the conclusion* A iff every possible valuation which makes Γ all true makes A true. (That's a semantically defined idea.)
>
> We can now also say that Γ *yields the conclusion* A *in your chosen proof-system* S iff there is an S-type derivation of the conclusion A from premisses in Γ. (This is a matter of there being a proof-array with the right syntactic shape.)

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 A from Γ, then A really *is* tautologically entailed by Γ. We also would like our favoured proof-system S to be *complete* – we want it to capture all the correct semantic entailment claims. In other words, if A is tautologically entailed by the set of premisses Γ, then there is indeed some S-derivation of A from premisses in Γ.

So, in short, 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 need hold no terrors! However, in establishing soundness and completeness for propositional logics you will encounter some useful strategies which can later be beefed-up to give soundness and completeness results for stronger logics.

3.2 FOL basics

(a) Having warmed up with propositional logic, we turn to full FOL so we can also deal with arguments whose validity depends on their quantificational structure (starting with the likes of our old friend 'Socrates is a man; all men are mortal; hence Socrates is a mortal').

We need to introduce appropriate formal languages with quantifiers (more precisely, with first-order quantifiers, running over a fixed domain of objects: the next chapter explains the contrast with second-order quantifiers). So *syntax* first.

The simplest atomic formulae now have some internal structure, being built up from names (typically mid-alphabet lower case letters) and predicates expressing properties and relations (typically upper case letters). So, for example, *Socrates is wise* might be rendered by Ws, and *Romeo loves Juliet* by Lrj – the predicate-first syntax is conventional but without deep significance.

Now, we can simply replace the name in the English sentence *Socrates is wise* with the quantifier expression *everyone* to give us another sentence (i) *Everyone is wise*. Similarly, we can simply replace the second name in *Romeo loves Juliet* with the quantifier expression *someone* to get the equally grammatical (ii) *Romeo loves someone*. In FOL, however, the formation of quantified sentences is a smidgin more complicated. So (i) will get rendered by something like $\forall \mathsf{x}\mathsf{W}\mathsf{x}$ (roughly, *Everyone x is such that x is wise*). Similarly (ii) gets rendered by something like $\exists \mathsf{x}\mathsf{L}\mathsf{r}\mathsf{x}$ (roughly, *Someone x is such that Romeo loves x*).[2]

Generalizing a bit, the basic syntactic rule for forming a quantified wff is roughly this: if $A(\mathsf{n})$ is a formula containing some occurrence(s) of the name 'n', then we can swap out the name on each occurrence for some particular variable, and then prefix a quantifier to form quantified wffs like $\forall \mathsf{x} A(\mathsf{x})$ and $\exists \mathsf{y} A(\mathsf{y})$.

But what is the rationale for this departure from the syntactic patterns of ordinary language and this use of the apparently more complex 'quantifier/variable' syntax in expressing generalizations? The headline point is that in our formal languages we crucially need to avoid the kind of structural ambiguities that we can get in ordinary language when there is more than one logical operator involved. Consider for example the ambiguous 'Everyone has not arrived'. Does that mean 'Everyone is such that they have not arrived' or 'It is not the case that everyone has arrived'? Our logical notation will distinguish $\forall \mathsf{x}\neg \mathsf{A}\mathsf{x}$ and $\neg\forall \mathsf{x}\mathsf{A}\mathsf{x}$, with the relative 'scopes' of the generalization and the negation now made fully transparent by the structure of the formulas.

(b) Turning to *semantics*: the first key idea we need is that of a *model structure*, a (non-empty) domain of objects equipped with some properties, relations and/or functions. And here we treat properties etc. extensionally. 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. (Compare our remarks on naive set theory in §2.1.)

Then, crucially, you need to grasp the idea of an *interpretation* of an FOL language in such a structure. Names are interpreted as denoting 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); a two-place predicate gets assigned a binary relation; and so on. Similarly, function-expressions get assigned suitable extensions.

Such an interpretation of the elements of a first-order language then generates a valuation (a unique assignment of truth-values) for every sentence of the interpreted language. How does it do that? Well, for a start, a simple predicate-name sentence like Ws will be true just if the object denoted by 's' is in the extension of W; a sentence like Lrj is true if the ordered pair of the objects denoted by 'r' and 'j' is in the extension of L; and so on. That's easy, and extending the story to cover sentences involving function-expressions is also straightforward. The propositional connectives continue to behave basically as in propositional logic.

But extending the formal semantic story to explain how the interpretation of a language fixes the valuations of more complex, quantified, sentences requires a new Big Idea. Roughly, the thought is:

[2]The notation with the rotated 'A' for *for all*, the universal quantifier, and rotated 'E' for *there exists*, the existential quantifier, is now standard, though bracketing conventions vary. But older texts used simply '(x)' instead of '$(\forall x)$' or '$\forall x$'.

1. $\forall$xWx is true just when Wn is true, no matter what the name 'n' might pick out in the domain.

This first version is evidently along the right lines: however, trying to apply it more generally can get problematic if the name 'n' is already has already been recruited for use with a *fixed* interpretation in the domain. So, on second thoughts, it will be better to use some other symbol to play the role of a new *temporary* name. Some FOL languages are designed to have a supply of special symbols for just this role. But a common alternative is allow a variable like 'x' to do double duty, and to act as a temporary name when it isn't tied to a preceding quantifier. Then we put

2. $\forall$xWx is true just when Wx is true, no matter what 'x' picks out when treated as a temporary name.

Compare: *Everything is W* is true just when *that is W* is true whatever the demonstrative 'that' might pick out from the relevant domain. And more generally, if 'A(x)' stands in for some wff with one or more occurrences of 'x',

3. $\forall$xA(x) is true just when A(x) is true, no matter what 'x' picks out when treated as a temporary name.

And then how do we expand this sort of story to treat sentences governed by more than one quantifier? We'll have to get more than one temporary name into play – and there are somewhat different ways of doing this. We needn't pursue this further here: but you do need to get your head round the details of one fully spelt-out story.

(c) We can now introduce the idea of a *model* for a set of sentences, i.e. an interpretation which makes all the sentences true together. And we can then again define a semantic relation of *entailment*, this time for FOL sentences:

> A set of FOL sentences Γ semantically entails A iff any interpretation which makes all the sentences in Γ true also makes the sentence A true – i.e., when any model for Γ is a model for A.

You'll again need to know some of the basic properties of this entailment relation.

For one important example, note that if Γ has no model, then – on our definition – Γ semantically entails A for any A at all, including any contradiction.

(d) Unlike the case of tautological entailment, this time there is no general procedure for mechanically testing whether Γ semantically entails A when quantified wffs are in play. So the use of proof systems to warrant entailments now really comes into its own.[3]

[3]A comment about this whose full import will only emerge later.

As we'll note in §6.1, one essential thing that we care about in building a proof system is that we can mechanically check whether a purported proof really obeys the rules of the system. And if we *only* care about that, we could e.g. allow a proof system to count every instance of a truth-table tautology as axiom, since it can be mechanically checked what's

You can again encounter five main types of proof system for FOL, with their varying attractions and drawbacks. And to repeat, you'll want at some future point to find out at least something about all these styles of proof. But, as before, we will principally be looking here at axiomatic systems and at one kind of natural deduction.

As you will see, whichever form of proof system you take, some care is require in handling inferences using the quantifiers in order to avoid fallacies. And we will need extra care if we don't use special symbols as temporary names but allow the same variables to occur both 'bound' by quantifiers and 'free'. You do need to tread carefully hereabouts!

(e) As with propositional logic, we will next want to show that our chosen proof system for FOL is sound and doesn't overshoot (so giving us false positives) and is complete and doesn't undershoot (leaving us unable to derive some semantically valid entailments).

In other words, if S is our FOL proof system, Γ a set of sentences, and A a particular sentence, we need to show:

> If there is an S-proof of A from premisses in Γ, then Γ does indeed semantically entail A. (Soundness)
>
> If Γ semantically entails A, then there is an S-proof of A from premisses in Γ. (Completeness)

There's some standard symbolism. $\Gamma \vdash A$ says that there is a proof of A from Γ; $\Gamma \vDash A$ says that A is semantically entailed by Γ. So to establish soundness and completeness is to prove $\Gamma \vdash A$ if and only if $\Gamma \vDash A$.

Now, as will become clear, it is important that the completeness theorem actually comes in two versions. There is a *weake*r version where Γ is restricted to having only finitely many members (perhaps zero). And there is a crucial *stronge*r version which allows Γ to be infinite.

And it is at *this* point, proving strong completeness, that the study of FOL becomes mathematically really interesting.

(f) Later chapters will continue the story along various paths; here though I should quickly mention just one immediate corollary of completeness.

Proofs in formal systems are always only finitely long; so a proof of A from Γ can only call on a *finite* number of premisses in Γ. But the strong completeness theorem for FOL allows Γ to have an *infinite* number of members. This combination of facts immediately implies the *compactness* theorem for sentences of FOL languages:

an instance of a tautology. Such a system obviously wouldn't illuminate why all tautologies can be seen as following from a handful of more basic principles. But suppose we don't particularly care about doing *that*. Suppose, for example, that our prime concern is to get clear about the logic of quantifiers. Then we might be content to, so to speak, let the tautologies look after themselves, and just adopt every tautology as an axiom, and then add quantifier rules against this background.

Some treatments of FOL, as you will see, do exactly this.

4. If every finite subset of Γ has a model, so does Γ.[4]

This compactness theorem, you will discover, has numerous applications in model theory.

3.3 A little more about types of proof-system

I've often been struck, answering queries on an internet forum, by how many students ask variants of "how do you prove X in first-order logic?", as if they have never encountered the idea that there is no single deductive system for FOL! So I do think it is worth emphasizing here at the outset that there *are* various styles of proof-system – and moreover, for each general style, there are many different particular versions.

This isn't the place to get into too many details with lots of examples. Still, some quick headlines could be very helpful for orientation.

(a) Let's have a mini-example to play with. Consider the argument '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'. Intuitively valid, of course. After all, just suppose for a moment that Jack *did* miss the train: then he'll be late; and hence we'll need to reschedule. Which shows that *if* he missed the train, we'll need to reschedule.

Using the obvious translation manual to render the argument into a formal propositional language, we'll therefore want to be able to show that – in our favoured proof system – we can correspondingly argue from the premisses $(\mathsf{P} \to \mathsf{Q})$ and $(\mathsf{Q} \to \mathsf{R})$ to the conclusion $(\mathsf{P} \to \mathsf{R})$.

(b) You will be familiar with the general idea of an axiomatized theory. We are given some axioms and some deductive apparatus is presupposed. Then the theorems of the theory are whatever can be derived from the axioms. Similarly:

> In an *axiomatic* logical system, we adopt some basic logical truths as axioms. And then we explicitly specify the allowed rules of inference: usually these are just very simple ones such as the *modus ponens* rule for the conditional which we will meet in a moment.
>
> A proof from some given premisses to a conclusion then has the simplest possible structure. It is just a sequence of wffs – each of which is either (i) one of the premisses, or (ii) one of the logical axioms, or (iii) follows from earlier wff in the proof by one of the rules of inference – with the whole sequence ending with the target conclusion.

[4] That's equivalent to the claim that if (i) Γ *doesn't have a model*, then there is a finite subset $\Delta \subseteq \Gamma$ such that (ii) Δ *has no model*. Suppose (i). This implies that Γ semantically entails a contradiction. So by completeness we can derive a contradiction from Γ in your favourite proof system. That proof will only use a finite collection of premisses $\Delta \subseteq \Gamma$. But if Δ proves a contradiction, then by soundness, Δ semantically entails a contradiction, which can only be the case if (ii).

And a logical theorem of the system is then a wff that can be proved from the logical axioms alone (without appeal to any further premisses).

Now, a standard axiomatic system for FOL (such as in Mendelson's classic book) will include as axioms all wffs of the following shapes:

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

More carefully, all instances of those two schemas – where we systematically replace letters like A, B, etc. with wffs (simple or complex) – will count as axioms. And among the rules of inference for our system will be the modus ponens rule:

MP. From A and $(A \to B)$ you can infer B.

With this apparatus in place, we can then construct the following formal derivation, arguing as wanted from $(\mathsf{P} \to \mathsf{Q})$ and $(\mathsf{Q} \to \mathsf{R})$ to $(\mathsf{P} \to \mathsf{R})$.

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!

(c) Informal deductive reasoning, however, is not relentlessly linear like this. We do not require that each proposition in a proof (other than a given premiss or a logical axiom) has to follow from what's gone before. Rather, 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 A is true'; we go on to show that, given what we've already established, this extra supposition leads to a contradiction; we then drop or 'discharge' the temporary supposition and conclude that *not-A*. That's how one sort of reductio ad absurdum argument works. For another example, we may again say 'Suppose that A is true'; this time we go on to show that we can now derive C; we then again discharge the temporary supposition and conclude that *if A, then C*. That's how we often argue for a conditional proposition: in fact, this is exactly what we did in the informal reasoning we gave to warrant the argument about Jack at the beginning of this section.

That motivates our using a more flexible kind of proof-system:

> A *natural-deduction* system of logic aims to formalize patterns of reasoning now including those where we can argue by making and then later discharging temporary assumptions. Hence, for example,

as well as the simple modus ponens (MP) rule for the conditional '$\to$', there will be a *conditional proof* (CP) rule along the lines of 'if we can infer B from the assumption A, we can drop the assumption A and conclude $A \to B$'.

Now, in a natural-deduction system, we will evidently need *some* way of keeping track of which temporary assumptions are in play and for how long. Two particular ways of doing this are commonly used:

(i) A multi-column layout was popularized by Frederick Fitch in his classic 1952 logic text, *Symbolic Logic: an Introduction*. Here's a proof in this style, from the same premisses to the same conclusion as before:

1.	$(\mathsf{P} \to \mathsf{Q})$		premiss
2.	$(\mathsf{Q} \to \mathsf{R})$		premiss
3.		P	supposition for the sake of argument
4.		Q	by MP from 3, 1
5.		R	by MP from 4, 2
6.	$(\mathsf{P} \to \mathsf{R})$		by CP, given the 'subproof' 3–5

So the key idea is that the line of proof snakes from column to column, moving a column to the right (as at line 3) when a new temporary assumption is made, and moving back a column to the left (as at line 6) when the assumption heading the column is dropped or discharged. This mode of presentation really comes into its own when multiple temporary assumptions are in play, and makes such proofs very easy to read and follow. And, compared with the axiomatic derivation, this regimented line of argument does indeed seem to warrant being called a 'natural deduction'!

(ii) However, the layout for natural deductions favoured for proof-theoretic work was first introduced Gerhard Gentzen in his doctoral thesis of 1933. He sets out the proofs as trees, with premisses or temporary assumptions at the top of branches and the conclusion at the root of the tree – and he uses a system for explicitly tagging temporary assumptions and the inference moves where they get discharged.

Let's again argue from the same premisses to the same conclusion as before. We will build up our Gentzen-style proof in two stages. First, then, take the premisses $(\mathsf{P} \to \mathsf{Q})$ and $(\mathsf{Q} \to \mathsf{R})$ and the additional supposition P, and construct the following proof of R using modus ponens twice:

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

The horizontal lines, of course, signal inference moves.

OK: so we've shown that, assuming P, we can derive R, by using the other assumptions. Hence, moving to the second phase of the argument, we will next discharge the assumption P while keeping the other assumptions in play, and apply conditional proof (CP), in order to infer $(\mathsf{P} \to \mathsf{R})$. We'll signal that the assumption P is no longer in play by now enclosing it in square brackets. So applying (CP) turns the previous proof into this:

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

For clarity, we tag both the assumption which is discharged and the corresponding inference line where the discharging takes place with matching labels, in this case '(1)'. (We'll need multiple labels when multiple temporary assumptions are put into play and then dropped.)

In this second proof, then, just the unbracketed sentences at the tips of branches are left as 'live' assumptions. So this is our Gentzen-style proof from those remaining premisses $(\mathsf{P} \to \mathsf{Q})$ and $(\mathsf{Q} \to \mathsf{R})$ to the conclusion $(\mathsf{P} \to \mathsf{R})$.

(d) There is *much* more to be said of course, but that's enough by way of some very introductory remarks about the first half of the following list of commonly used types of proof system:

1. Old-school axiomatic systems.
2. (i) Natural deduction done Gentzen-style.
   (ii) Natural deduction done Fitch-style.
3. 'Semantic tableaux' or 'truth trees'.
4. Sequent calculi.
5. Resolution calculi.

So next, a *very* brief word about semantic tableaux, which are akin to Gentzen-style proof trees turned upside down.

The key idea is this. Instead of starting from some premisses Γ and arguing towards an eventual conclusion A, we begin instead by assuming the premisses are all true while the wanted conclusion is false. And then we 'work backwards' from the assumed values of these typically complex wffs, aiming to uncover a valuation v of the atoms for the relevant language which indeed makes Γ all true and A false. If we succeed, and actually find such a valuation v, then that shows that A doesn't follow from Γ. But if our search for such a valuation v gets completely entangled in contradiction, that tells us that there is no such valuation: in other words, on any valuation, if Γ are all true, then A has to be true too.

Note however that assuming e.g. that a wff of the form $(A \vee B)$ is true doesn't tell us which of A and B is true too: so as we try to 'work backwards' from the values of more complex wffs to the values of their components we will typically have to explore branching options, which are most naturally displayed on a downward-branching tree. Hence 'truth trees'.

The details of a truth-tree system for FOL are elegantly simple – which is why the majority of elementary logic books for philosophers introduce either (2.ii) Fitch-style natural deduction or (3) truth trees, or both. And it is well worth getting to know about tree systems at a fairly early stage because they can be adapted rather nicely to dealing with logics other than FOL. However, introductory mathematical logic textbooks do usually focus on either (1) axiomatic systems or (2.i) Gentzen-style proof systems, and those will remain our initial main focus here too.

As for (4) the sequent calculus, in its most interesting form this really comes into its own in more advanced work in proof theory. While (5) resolution calculi are perhaps of particular concern to computer scientists interested in automating theorem proving.

(e) I should stress, though, that even once you've picked your favoured general *type* of proof-system to work with from (1) to (5), there are many more choices to be made before landing on a specific system of that type. For example, F. J. Pelletier and Allen Hazen published a useful survey of logic texts aimed at philosophers which use natural deduction systems (tinyurl.com/pell-hazen). They note that no less than thirty texts use a variety of Fitch-style system (2.ii): and, rather remarkably, *no two of these have exactly the same system of rules for FOL!*

Moral? Don't get too hung up on the finer details of a particular textbook's proof-system; it is the overall guiding ideas that matter, together with the Big Ideas underlying proofs *about* the chosen proof-system (such as the soundness and completeness theorems).

3.4 Basic recommendations for reading on FOL

A preliminary reference. In my elementary logic book I do carefully explain the 'design brief' for the languages of FOL, spelling out the rationale for the quantifier-variable notation. For some, 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.

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

2. Ian Chiswell and Wilfrid Hodges, *Mathematical Logic* (OUP, 2007), up to §7.6.

This is very approachable. It is written by mathematicians primarily for mathematicians, yet it is only one notch up in actual difficulty from some introductory texts for philosophers like mine or Nick Smith's. However – as its title might suggest – it does have a notably more mathematical 'look and feel'. Philosophers can skip over a few of the more mathematical illustrations; while depending on background, mathematicians should be able to take this book at pace.

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, tinyurl.com/CHbooknote.

Very warmly recommended, then. For the moment, you only *need* read up to and including §7.6. 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). Downloadable at tinyurl.com/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 beginning of §3.3 on compactness, 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 (by just helping themselves to all instances of tautologies as axioms). But this rather 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.

Some illustrations of ideas can presuppose a smattering of background mathematical knowledge (the authors are mathematicians); 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, and say more about this admirable book in another book note, tinyurl.com/LKbooknote.

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

4. Derek Goldrei, *Propositional and Predicate Calculus: A Model of Argument* (Springer, 2005). This book is explicitly designed for self-study and works very well. 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 for self-teaching: 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 his book with a little reading on 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.

3.5 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 additional or alternative reading. Then in the final section of the chapter, I will mention some other books I've been asked about.

I'll begin a notch or two down in level from the texts we have looked at so far, with a book written by a philosopher for philosophers. It should be particularly accessible to non-mathematicians who haven't done much formal logic before, and could help ease the transition to coping with the more mathematical style of the books recommended in the last section.

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 usefully introduces you to semantic tableaux *and* an Hilbert-style axiomatic proof system *and* natural deduction *and* even a sequent calculus as well. Indeed, though written for non-mathematicians, 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.

To repeat, unlike our main recommendations, Bostock does give a brisk but very clear presentation of tableaux ('truth trees'), and he proves completeness for tableaux in particular, which I always think makes the needed construction seem particularly natural. If you are a philosopher, you may well have already encountered these truth trees in your introductory logic course. If not, at some point you will want to find out about them. As an alternative to Bostock, my elementary introduction to truth trees for propositional logic available at tinyurl.com/proptruthtrees will quickly give you the basic idea in an accessible way. Then you can dip into my introduction to truth trees for quantified logic at tinyurl.com/qtruthtrees.

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

6. 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.

Next, here's a much-used text which has gone through multiple editions; 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 on further topics, so it could be worth making early acquaintance with

7. 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 – one reason why I have recommended beginning with C&H instead. For more on this book, see tinyurl.com/dalenlogic.

As a follow up to C&H, I just recommended L&K's *Friendly Introduction* which uses an axiomatic system. As an alternative to that, here is an older (and, in its day, much-used) text:

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

   This also focuses on an axiomatic system, and is often regarded as a classic of exposition. However, it does strike me as somewhat less approachable than L&K, so I'm not surprised that students do quite often report finding this book a bit challenging *if used by itself as a first text*.

   However, this is an admirable and very reliable piece of work which most readers should be able to cope with well if used as a supplementary second text, e.g. after you have tackled C&H. And stronger mathematicians might well dive into this as their first preference.

   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 tinyurl.com/enderlogicnote.

I should also certainly mention the outputs from the Open Logic Project. This is an entirely admirable, collaborative, open-source, enterprise inaugurated by Richard Zach, and continues to be work in progress. You can freely download the latest full version and various sampled 'remixes' from tinyurl.com/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: the mix of chapters on propositional and quantificational logic in the following selection has been expanded and developed considerably, and the result is much more book-like:

9. Richard Zach and others, *Sets, Logic, Computation** (Open Logic: downloadable at tinyurl.com/slcopen).

   There's a lot to like here (Chapters 5 to 13 are the immediately relevant ones for the moment). In particular, Chapter 11 could make for very useful supplementary reading on natural deduction. Chapter 10 tells you about a sequent calculus (a slightly odd ordering!). And Chapter 12 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 what you have learnt elsewhere.

So much, then, for reading on FOL running on more or less parallel tracks to the main recommendations in the preceding section. I'll finish this section by recommending two books that push the story on a little. First, an absolute classic, short but packed with good things:

10. 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 real delight for the mathematically minded.

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

11. Melvin Fitting, *First-Order Logic and Automated Theorem Proving* (Springer, 1990, 2nd end. 1996).

    A wonderfully lucid book by a renowned expositor. Yes, at a number of places in the book there are illustrations of how to implement algorithms in Prolog. But either you can easily pick up the very small amount of background knowledge that's needed to follow everything that is going on (and that's quite fun) or you can in fact just skip lightly over 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 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 fifth style of proof!

This book's approach, then, is rather different from most of the other recommended books. 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.)

3.6 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 of anything like the contemporary kind – Frege's system in his *Begriffsschrift* of 1879 – allows higher-order quantification in the sense explained in the next chapter (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, English translation 1950 – and still very worth reading) that FOL is highlighted under the label 'the restricted predicate calculus'. Those three books all give axiomatic presentations of logic (though notationally very different from each other): axiomatic systems similar enough to the third are still often called 'Hilbert-style systems'

(b) 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 certain subject matter (for Frege, 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 particular kinds of constituents of the world and their properties; for Russell, logic tells us absolutely general truths. 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* in the sense of lacking real content (in the way that a repetitious claim like 'Brexit is Brexit' lacks real content). 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: logical truths 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*. But now consider this little two-step derivation:

> Suppose for the sake of argument that *P and Q*; then we can derive *P* – by the first rule which partly fixes the meaning of 'and'.
>
> And given that little suppositional inference, the rule of conditional proof, which partly gives the meaning of 'if', entitles us to drop the supposition and conclude *if P and Q, then P*.

Or presented as a Gentzen-style proof we have

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

In short, the inference rules governing 'and' and 'if' enable us to derive that logical truth 'for free' (from no remaining assumptions): it's a theorem of a formal system with those rules.

If this 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 truths fall out just as byproducts of the inference rules whose applicability is, in some sense, built into the very meaning of the connectives and the quantifiers.

It is a nice question how far we should buy that sort of de-mystifying story about the nature of logical truth. But whatever your eventual judgement on this, there surely *is* something odd about thinking with Frege and Russell that a systematized logic is 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? And hence, shouldn't a formalized logic highlight *rules of valid proof-building* (perhaps as in a natural deduction system) rather than stressing *logical truths* (as logical axioms)?

(c) Back to the history of the technical development of logic. An obvious starting place is with the clear and judicious

12. William Ewald, 'The emergence of first-order logic', *The Stanford Encyclopaedia*, tinyurl.com/emergenceFOL.

If you want rather more, the following is also readable and very helpful:

13. José Ferreirós, 'The road to modern logic – an interpretation', *Bulletin of Symbolic Logic* 7 (2001): 441–484, tinyurl.com/roadtologic.

And for a longer, though rather bumpier, read – you'll probably need to skim and skip! – you could also try dipping into this more wide-ranging piece:

14. 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): tinyurl.com/developlogic.

3.7 Postscript: Other treatments?

I will end this chapter by responding – often rather brusquely – to a variety of Frequently Asked Questions raised in response to earlier versions of the Guide (often questions of the form "But why haven't you recommended X?"). So, in what follows,

(a) I quickly mention a handful of books aimed at philosophers (but only one will be of interest to us at this point).
(b) Next, I consider four deservedly classic books, now more than fifty years old.
(c) Then I look at eight more recent mathematical logic texts (I again highlight one in particular).
(d) Finally, for light relief, I look at some fun extras from an author whom we have already met.

(a) The following five books are very varied in style, level and content, but are all designed with philosophers particularly in mind.

(a1) Richard Jeffrey, *Formal Logic: Its Scope and Limits* (McGraw Hill 1967, 2nd edn. 1981).
(a2) Merrie Bergmann, James Moor and Jack Nelson, *The Logic Book* (McGraw Hill 1980; 6th edn. 2013).
(a3) John L. Bell, David DeVidi and Graham Solomon, *Logical Options: An Introduction to Classical and Alternative Logics* (Broadview Press 2001).
(a4) Theodore Sider, *Logic for Philosophy** (OUP, 2010).
(a5) Jan von Plato, *Elements of Logical Reasoning** (CUP, 2014).

Quick comments: Sider's book (a4) falls into two halves, and the second half is quite good on modal logic; but the first half of the book, the part which is relevant to us now, is very poor. Only the first two chapters of *Logical Options* (a3) are on FOL, and not at the level we really want. Von Plato's *Elements* (a5) is good but better regarded, I think, as an introduction to proof theory and we will return to it in Chapter 9.

The Logic Book (a2) is over 550 pages, starting at about the level of my introductory book, and going as far as results like a full completeness proof for FOL, so its coverage overlaps considerably with the main recommendations of §3.4. 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.

So that leaves Richard Jeffrey's lovely book. This is relatively short, and the first half on propositional logic is mostly at a very introductory level, which is why I haven't mentioned it before. But if you know a little about trees for propositional logic — as e.g. explained in the reading reference (6) in §3.5 – then you could start at Chapter 5 and read the rest of the book with enjoyment and illumination. For this gives a gentle yet elegant introduction to the undecidability of FOL and a very nice proof of completeness for trees.

(b) Next, four classic books, again listed in order of publication. All of them are worth visiting sometime, even if they are not now the first choices for beginners.

(b1) Elliott Mendelson, *Introduction to Mathematical Logic* (van Nostrand 1964; Chapman and Hall/CRC, 6th edn. 2015).
(b2) Joseph R. Shoenfield, *Mathematical Logic* (Addison Wesley, 1967).
(b3) Stephen C. Kleene, *Mathematical Logic* (John Wiley 1967; Dover Publications 2002).
(b4) Geoffrey Hunter, *Metalogic* (Macmillan 1971; University of California Press 1992).

Perhaps the most frequent question I used to get asked in response to early versions of the Guide was 'But what about Mendelson, Chs 1 and 2'? Well, (b1) 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 when the world was a lot younger!

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 working through. But quite unsurprisingly, over fifty years on, there are many rather more accessible and more amiable alternatives for *beginning* serious logic. Mendelson's book is a landmark well worth visiting one day, but I can't recommend starting here (especially for self-study). For a little more, see tinyurl.com/mendelsonlogic.

Shoenfield's (b2) is really aimed at graduate mathematicians, and is not very reader-friendly. Maybe take a look one day, particularly at the final chapter on set theory; but not yet! For a little more, see tinyurl.com/schoenlogic.

Kleene's (b3) – not to be confused with his hugely influential earlier *Introduction to Metamathematics* – goes much more gently than Mendelson: it takes almost twice as long to cover propositional and predicate logic, so Kleene has much more room for helpful discursive explanations. This was in its time a rightly much admired text, and still makes excellent supplementary reading.

But if you *do* want an old-school introduction from the same era, you might most enjoy the somewhat less renowned book by Hunter, (b4). This is not 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 6. Unfortunately, the typography – from pre-LaTeX days – isn't very pretty to look at. But in fact the treatment of an axiomatic system of logic is extremely clear and accessible.

(c) We now turn to a number of more recent texts in mathematical logic that have been suggested as candidates for this Guide. As you will see, the most interesting of them – which almost made the cut to be included in §3.5's list of additional readings – is the idiosyncratic book by Kaye.

(c1) H.-D. Ebbinghaus, J. Flum and W. Thomas, *Mathematical Logic* (Springer, 2nd edn 1994, 3rd edn. 2021).
(c2) René Cori and Daniel Lascar, *Mathematical Logic, A Course with Exercises: Part I* (OUP, 2000).
(c3) Shawn Hedman, *A First Course in Logic* (OUP, 2004).
(c4) Peter Hinman, *Fundamentals of Mathematical Logic* (A. K. Peters, 2005).
(c5) Wolfgang Rautenberg, *A Concise Introduction to Mathematical Logic* (Springer, 2nd edn. 2006).
(c6) Richard Kaye, *The Mathematics of Logic* (CUP 2007).
(c7) Harrie de Swart, *Philosophical and Mathematical Logic* (Springer, 2018)
(c8) Martin Hils and François Loeser, *A First Journey Through Logic* (AMS Student Mathematical Library, 2019).

I have added the last two to the list in response to queries. But while the relevant Chapters 2 and 4 of (c7) are quite attractively written, and have some interest, there also are a number of presentation choices I'd quibble with. You can do better. While (c8) just isn't designed to be a conventional mathematical logic text. It does have a fast-track introduction to FOL, but this is done far *too* fast to be of much use to anyone. We can ignore it.

So going back to earlier texts, Ebbinghaus, Flum and Thomas's (c1) 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 in Germany.

There is a lot of material here, often covered well. But I can't find myself wanting to recommend it as a good place to *start*. 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 failing to capture the elegance that using a sequent calculus can bring. You can freely download the old second edition at tinyurl.com/EFTlogic. For more on this book, see tinyurl.com/EFTbooknote.

Chapters 1 and 3 of Cori and Lascar's (c2) could appeal to the more mathematical reader. Chapter 1 is on semantic aspects of propositional logic, and is done clearly. Also, an unusually good feature of the book, there are – as with other chapters – interestingly non-trivial exercises, with expansive answers given at the end. Chapter 2, I would say, jumps to a significantly more demanding level, introducing Boolean algebras (and really, you should probably know a bit of algebra and topology to fully appreciate what is going on – we'll return to this in §12.1). Chapter 3 gets back on track with the syntax and semantics of predicate languages, plus a smidgin of model theory too. Not perhaps, the place to start for a first introduction to this material, but worth reading. Then Chapter 4, the last in the book, is on proof systems, but probably not so helpful.

Shawn Hedman's (c2) 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, however. 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 §5.3 in this Guide). But Hedman offers a peculiarly ugly not-so-natural deductive system. By contrast though – as already noted – he *is* good on so-called resolution proofs. For more about what does and what doesn't work in Hedman's book, see tinyurl.com/hedmanbook.

Peter Hinman's (c3) 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 (his concerns are more model-theoretic). And most readers will find this book pretty tough going. This is certainly not, then, the place to start with FOL. However, the first three chapters of the book do contain some supplementary material that could be very interesting once you have got hold of the basics from elsewhere, and could particularly appeal to mathematicians. For more about what does and what doesn't work in Hinman's book, see tinyurl.com/hinmanbook.

The first three chapters of Wolfgang Rautenberg's (c4) are on FOL and have some nice touches. But I suspect these hundred pages are rather *too* concise to serve most readers as an initial introduction; and the preferred formal system is not a 'best buy' either. Can be recommended as good revision material, though.

Finally, Richard Kaye is the author of a particularly attractively written 1991 classic on models of Peano Arithmetic (we will meet this in §12.3). So I had high hopes for his later *The Mathematics of Logic* (c5). "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. ... 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 very unusual route work as an *introduction*? I am not at all convinced. It seems to me that the journey is made too bumpy and the road taken is far too uneven in level for this to be appealing as an early trip through first-order logic. However, if you *already* know a fair amount of this material from more conventional presentations, the different angle of approach in this book linking topics together in new ways could well be very interesting and illuminating.

(d) I have already strongly recommended Raymond Smullyan's 1968 *First-Order Logic*. Smullyan went on to write some absolutely classic texts on Gödel's theorem and on 'diagonalization' arguments, which we'll be mentioning later. But as well as these, he also wrote many 'puzzle'-based books aimed at a wider audience, including e.g. the rightly renowned *What is the Name of This Book?** (Dover Publications reprint of 1981 original, 2011) and *The Gödelian Puzzle Book** (Dover Publications, 2013).

Smullyan has also written *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, 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 given 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. To be sure, 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 – a very entertaining and illuminating supplement.

(Smullyan's later *A Beginner's Guide to Mathematical Logic**, Dover Publications, 2014, is rather 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.)

4 Second-order logic, quite 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 can leave the exploration of non-classical logics to later chapters, starting with Ch. 8. I will, however, say a little about second-order logic straight away, in this chapter. Why?

Theories expressed in first-order languages with a first-order logic turn out to have their limitations – that's a theme that will recur when we look at model theory (Ch. 5), theories of arithmetic (Ch. 6), and set theory (Ch. 7). And you will occasionally find explicit contrasts being drawn with richer theories expressed in second-order languages with a second-order logic. So, although it's a judgement call, I think it is worth getting to know just a bit about second-order logic quite early on in order to understand the contrasts being drawn.

But first, . . .

4.1 A preliminary note on many-sorted logic

(a) As you will now have seen from the core readings, 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 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 understood as running just over scalars, and with the other two quantifiers running just over vectors. Or we could explicitly declare which domain a quantified variable is running over by using a notation like $(\forall a \colon S)$ to assign a to scalars and $(\forall \mathbf{v} \colon V)$ to assign $\mathbf{v}$ to vectors: mathematicians often do this informally. (And in some formal 'type theories', this kind of notation becomes the official policy: see §12.7.)

It might seem really rather strange, 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 therefore have to render the rule for scalar multiplication along the lines of

(3) $$\forall x \forall y \forall z((Sx \wedge Vy \wedge Vz) \rightarrow x(y + z) = xy + xz),$$

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

(b) 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. So, 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?

In fact, it isn't hard to set up such a revised version of FOL (it *is* first-order, as the quantifiers are still of the now familiar basic type, running over objects in the relevant domains – compare §4.2). The syntax and semantics of a many-sorted language can be defined quite easily. Syntactically, we will 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 (for example, the vector-addition function can only be applied to terms for vectors). 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 will have the same nice technical properties as standard one-sorted FOL; crucially, you can prove soundness and 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 need to render propositions like (2) by using restricted quantifications like (3) – but get the benefit of a slightly-simpler-in-practice logic.[1]

So you pays your money and you takes your choice. For many (most?) purposes, logicians prefer the second option, sticking to standard single-sorted FOL. That's because, at the end of the day, we care rather less about elegance when regimenting this or that theory than about having a simple-but-powerful logical system.

[1] Note though that we do also get some added flexibility on the second option. The use of a sorted quantifier $\forall aFa$ with the usual logic presupposes that there is at least one thing in the relevant domain for the variable a. But a corresponding restricted quantification $\forall x(Ax \rightarrow Fx)$, where the variable x quantifies over some wider domain while A picks out the relevant sort which a was supposed to run over, leaves open the possibility that there is nothing of that sort.

4.2 Second-order logic

(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:

(Ind 1) Take *any* numerical property X; if (i) zero has property 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 $0'$ or $0''$ or $0'''$ or ..., where the prime '$'$' is a standard sign for the function that maps a number to its successor). There are no stray numbers outside that sequence, so a property that percolates down the sequence eventually applies to any number at all.

There is no problem about expressing some particular *instances* of the induction principle in a first-order language. Suppose P is a formal one-place predicate expressing some particular arithmetical property: then we can express the induction principle for this property by writing

(Ind 2) $$(\mathsf{P0} \wedge \forall \mathsf{x}(\mathsf{Px} \to \mathsf{Px}')) \to \forall \mathsf{x}\, \mathsf{Px}$$

where the small-'x' quantifier runs over the natural numbers and again the prime expresses the successor function. But how can we state the *general* principle of induction in a formal language, the principle that applies to *any* numerical property? The natural candidate is something like this:

(Ind 3) $$\forall \mathsf{X}((\mathsf{X0} \wedge \forall \mathsf{x}(\mathsf{Xx} \to \mathsf{Xx}')) \to \forall \mathsf{x}\, \mathsf{Xx}).$$

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 (Ind 2) by the predicate 'P', and the expressed generalization is intended to run over all properties of numbers, so that (Ind 3) indeed formally renders (Ind 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 do want to stick with a theory framed in a first-order arithmetical language L which just quantifies over numbers, the best we can do to render the induction principle is to use a template or schema and say something like

(Ind 4) For any arithmetical L-predicate $A(\ )$, simple or complex, the corresponding wff of the form $(A(\mathsf{0}) \wedge \forall \mathsf{x}(A(\mathsf{x}) \to A(\mathsf{x}')) \to \forall \mathsf{x}\, A(\mathsf{x})$ is an axiom.

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

(b) Another interesting issue to think about. Start with a definition:

> Suppose R is a binary relation. Define R^n (for $n > 0$) to be the relation that holds between a and b when there are n R-related intermediaries between them – i.e. when there are objects $x_1, x_2, \ldots x_n$ such that Rax_1, Rx_1x_2, Rx_2x_3, …, Rx_nb. And take R^0 just to be R.
>
> Then R^*, the *ancestral* of R, is the relation that holds between a and b just when there is some $n \geq 0$ such that R^nab – i.e. just when there is a finite chain of R-related steps from a to b.

Example: if R is the relation *is a parent of*, then R^* is the relation *is a (direct) ancestor of*. Which explains 'ancestral'! An arithmetical example: if S is the relation *is the successor of*, then S^*nm holds when there is a sequence of successors starting with m and finishing with n. And n is a natural number just if S^*n0.

Now four easy observations:

(i) First note that given a relational predicate R expressing the relation R, we can of course define complex expressions, which we might abbreviate $\mathsf{R^n}$, to express the corresponding relations R^n. For example, we just put

$$\mathsf{R^3ab} =_{\text{def}} \exists \mathsf{x_1} \exists \mathsf{x_2} \exists \mathsf{x_3}(\mathsf{Rax_1} \wedge \mathsf{Rx_1x_2} \wedge \mathsf{Rx_2x_3} \wedge \mathsf{Rx_3b}).$$

(ii) Now suppose we *can* also construct an expression $\mathsf{R^*}$ for the ancestral of the relation expressed by R. And then consider the infinite set of wffs

$$\{\neg\mathsf{Rab}, \neg\mathsf{R^1ab}, \neg\mathsf{R^2ab}, \neg\mathsf{R^3ab}, \ldots, \neg\mathsf{R^nab}, \ldots, \mathsf{R^*ab}\}$$

Then (X) every finite collection of these wffs has a model (let n be the largest index appearing, and consider the case where a is the R-ancestor of b more than n generations removed). But obviously (Y) the whole infinite set of sentences doesn't have a model (a can't be an R-ancestor of b without there being some n such that R^nab).

(iii) Now, if we stay first-order, then we know that the compactness theorem holds: i.e. if every finite subset of some set of sentences has model, then so does the whole set. That means for first-order wffs we can't have both (X) and (Y). Which shows that we can't after all construct an expression $\mathsf{R^*}$ from R and first-order logical apparatus. In short, we can't define the ancestral of a relation in first-order logic.

(iv) On the other hand, a little reflection shows that a stands in the ancestral of the R-relation to b just in case b inherits every property that is had by any immediate R-child of a and which is then always preserved by the R relation.[2] Hence Frege could famously define the ancestral using *second-*

[2] Why? For one direction: if b is an eventual R-descendant of a, then b will evidently inherit any property which is passed all the way down an R-related chain starting from an R-child of a. For the other direction: if b inherits any R-transmitted property from an R-child of a, it will in particular inherit a's property of being an R-descendant of a.

order apparatus like this:

$$\mathsf{R}^*\mathsf{ab} =_{\text{def}} \forall \mathsf{X}[(\forall \mathsf{x}(\mathsf{Rax} \to \mathsf{Xx}) \land \forall \mathsf{x}\forall \mathsf{y}(\mathsf{Xx} \land \mathsf{Rxy} \to \mathsf{Xy})) \to \mathsf{Xb}]$$

And note that, because we *can* construct a second-order expression R^* for the ancestral of the relation expressed by R, then – because (X) and (Y) are true together – compactness must fail for second-order logic.

In sum, we *can't* define the ancestral of a relation in first-order logic (and hence can't define equivalent notions like transitive closure either). But we *can* do so in second-order logic. So we see that – as with induction – allowing quantification into predicate position increases the expressive power of our language in a mathematically very significant way.

(c) And it isn't difficult to extend the syntax and semantics of first-order languages to allow for second-order quantification. Start with simple cases.

The required added *syntax* is unproblematic.

Recall how we can take a formula $A(\mathsf{n})$ 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 $\forall \mathsf{x} A(\mathsf{x})$.

We just need now to add the analogous rule that we can take a formula $A(\mathsf{P})$ 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 of the form $\forall \mathsf{X} A(\mathsf{X})$.

Fine print apart, that's straightforward.

The standard *semantics* is equally straightforward. We interpret names, predicates and functions just as before, and likewise for the connectives and first-order quantifiers. And again we model the story about the novel second-order quantifiers on the account of first-order quantifiers. So first fix a domain of quantification.

Recall that, roughly, $\forall \mathsf{x} A(\mathsf{x})$ is true on a given interpretation of its language just when $A(\mathsf{n})$ remains true, however we vary the object in the domain which is assigned to the name 'n' as its interpretation.

Similarly then, $\forall \mathsf{X} A(\mathsf{X})$ is true on an interpretation just when $A(\mathsf{P})$ remains true, however we vary the subset of the domain which is assigned to the unary predicate 'P' as *its* interpretation (i.e. however we vary 'P's extension).

Again, there's fine print; but you get the general idea.

We'll now also 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 extra 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 A just in case every interpretation that makes all of Γ true makes A true.

(d) So, in bald summary, the situation is this. There are quite a few familiar mathematical claims like the arithmetical induction principle, and familiar mathematical constructions like forming the ancestral or forming the closure of a relation, which 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 (Ind 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, as you will see in due course, a first-order theory of arithmetic which has to rely on a limited induction principle like (Ind 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 extras in their domains).

The obvious question which arises from all this, then, is *why is it the standard modern practice to privilege FOL*? Why not adopt a second-order logic from the outset as our preferred framework for regimenting mathematical arguments? – after all, as noted in §3.6, early formal logics like Frege's allowed more than first-order quantifiers.

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

There can be sound but incomplete deductive systems S_2 for a language including second-order quantifiers. So we can have the one-way conditional that, whenever there is an S_2-proof from premisses in Γ to the conclusion A, then Γ semantically entails A. But the converse fails. We can't have a respectable formal system S_2 (where it is decidable what's a proof, etc.) such that, whenever Γ semantically entails A, there is an S_2-proof from premisses in Γ to the conclusion A. *Once second-order sentences (with their standard interpretation) are in play, we can't fully capture the relation of semantic entailment in a formal deductive system.*

(f) Let's pause 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 quite independently – fixing the domain of one doesn't fix the domain of the other. And because each sort of quantifier, as it were, stands alone, the familiar first-order logic continues to apply to each separately.

But in second-order logic it is entirely different. For note that on the standard semantic story, it is now the *same* domain which fixes the interpretation 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 (e.g. *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; it is this which blocks the possibility of a complete deductive system for second-order logic.

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

(g) Of course, it's not supposed to obvious at the outset that we *can't* have a complete deductive system for second-order logic with the standard semantics, any more than it is obvious at the outset that we *can* have a complete deductive system for first-order logic!

True, we *have* now shown in (b) that compactness fails in the second-order case, and that is enough to show that we can't have a *strongly* complete deductive system for second-order logic with standard semantics (just recycle the ideas of §3.2, fn. 4). However, it does take much more work to show that we can't have even a *weakly* complete proof system: the usual argument relies on Gödel's incompleteness theorem which we haven't yet met.

And it isn't obvious either what exactly the significance of this failure of completeness might be. In fact, the whole question of the status of second-order logic leads to some tangled debates.

Let's briefly touch on one disputed issue. 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 (for example) our second-order induction principle

$$\text{(Ind 3)} \qquad \forall \mathsf{X}((\mathsf{X0} \wedge \forall \mathsf{x}(\mathsf{Xx} \to \mathsf{Xx'})) \to \forall \mathsf{x}\, \mathsf{Xx}).$$

in the form

$$\text{(Ind 5)} \qquad \forall \mathsf{X}((0 \in \mathsf{X} \wedge \forall \mathsf{x}(\mathsf{x} \in \mathsf{X} \to \mathsf{x}' \in \mathsf{X}) \to \forall \mathsf{x}\, \mathsf{x} \in \mathsf{X}),$$

making it explicit that the big-'X' variable is running over sets of numbers? Well, we *can* do that. Though if (Ind 5) is to replicate the content of (Ind 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 (Ind 3) can be rewritten as (Ind 5), this just goes to show that in using second-order quantifiers we are straying into the realm of set theory. 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 §2.4; and see the papers by Boolos mentioned below). So this means that (Ind 5) in fact dresses up the induction principle (Ind 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 §4.4 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 can shelve such issues: you just need to grasp a few basic technical facts about second-order logic.

4.3 Recommendations on many-sorted and second-order logic

First, for something on the formal details of many-sorted first-order languages and their logic:

What little you need for present purposes is covered in four clear pages by

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

There is, however, a bit more that can be fussed over here, and some might be interested in looking at e.g. Hans Halvorson's *The Logic in Philosophy of Science* (CUP, 2019), §§5.1–5.3.

Turning now to second-order logic:

For a brief review, saying only a little more than my overview remarks, see

2. Richard Zach and others, *Sets, Logic, Computation** (Open Logic) §13.3, slc.openlogicproject.org.

You could then look e.g. at the rest of Chapter 4 of Enderton (1). Or, rather more usefully at this stage, 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. (Don't worry if some points will only become really clear once you've done some model theory and some formal arithmetic.)

To nail down some of the technical basics you can then 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, various ways of developing the semantics of *first-order* logical languages; and then it compares the first-order case with the *second-order* options, both 'full' semantics and 'Henkin' semantics.

For alternative introductory reading you could look at the clear

5. Theodore Sider, 'Crash course on higher-order logic', §§1–3, 5. Available at tinyurl.com/siderHOL.

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

6. 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).

4.4 Conceptual issues

So much for formal details. 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, and is remarkably accessible.

Shapiro, in both his *Handbook* essay and in his earlier book, mentions Boolos's arguments against regarding second-order logic as essentially set-theoretical. Very roughly, the idea is that, instead of interpreting e.g. the second-order quantification in the induction axiom (Ind 3) as in effect quantifying over sets, we should read it along these lines:

(Ind 3′) Whatever numbers we take, if 0 is one of them, and if n' is one of them if n is, then we have all the numbers.

So the idea is that we don't need to invoke *sets* to interpret (Ind 3), just a noncommittal use of *plurals*. For more on this, just because he is so very readable, let me highlight the thought-provoking

7. 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.

Note, however, that the usual semantics for second-order logic and Boolos's proposed alternative do share an assumption – in effect, neither treat *properties* very seriously! Recall, we started off stating the informal induction principle

(Ind 1) in terms of a generalization over *properties of numbers*. But in interpreting its second-order regimentation (Ind 3), we've only spoken of *sets of numbers* (to serve as extensions of properties, the standard story) or spoken even more economically, just about *numbers, plural* (Ind $3'$, Boolos). Where have the properties gone? Philosophers, at any rate, might want to resist reducing higher-order entities (properties, properties of properties) to first-order entities (objects, or sets of objects). Now, this is most certainly not the place to enter into those debates. But for a nice survey with pointers to relevant discussions, see

8. Lukas Skiba, 'Higher order metaphysics', *Philosophy Compass* (2021), tinyurl.com/skibameta.

5 Model theory

The high point of a first serious encounter with FOL is the proof of the completeness theorem. Introductory texts then usually discuss at least a couple of quick corollaries of the proof – the *compactness theorem* (which we've already met) 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 which are recommended as advanced reading in §12.2). 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 foray into mathematical logic, though, you should certainly get your head around Level 1 model theory. Then tackle as much Level 2 theory as grabs your interest.

But what topics can we assign to these first two levels?

5.1 Elementary model theory

(a) Model theory is about mathematical structures and about how to characterize and classify them using formal languages. Put 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 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 group theory or Boolean algebra), theories of various kinds of order, etc., and also 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 need to start by thinking more about structures and the ways they can be interrelated. For example, one structure can be simply a substructure of another, or can extend another. Or we can map one structure to another in a way that preserves some relevant structural features (for a rough analogy, think of the map which sends metro-stations-and-their-relations to points-on-a-diagram-and-*their*-relations in a way that preserves e.g. structural 'between-ness' relations). In particular, we will be interested in structure-preserving maps which send one structure to a copy embedded inside another structure, and cases 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 also want to know something about the interplay between these expansions/reductions of structures and expansions/reductions of 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

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

with the quantifier running over $\mathbb{N}$. 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 take instead the structure $(\mathbb{Z}, <)$, i.e. all the integers, negative and positive, equipped with *their* standard order relation. And consider the corresponding formal language where $<$ gets re-interpreted accordingly. The same formula as before, but with the quantifier now running over $\mathbb{Z}$, also suffices to define the successor relation over the integers. 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 $(\mathbb{Z}, <)$. Rather as you would expect, the ordering relation gives only the relative position of integers, but doesn't fix the zero.

OK, those are indeed trivial 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 (here, we often use 'theory' in a very broad sense that encompasses any set of sentences), we can ask how varied the models of a given theory can be. In many cases, quite different structures for interpreting a given language can be 'elementarily equivalent', meaning that they satisfy just the same sentences of the language. 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 seat of first-order sentences – a first-order theory – for which just *these* structures are the models? Or given a particular structure, and a language for it with the right sort of names, predicates and functional expressions, we can 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 (i.e. the set of sentences true of the structure) *decidable,* in the sense that a computer could work out what sentences belong to 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 §3.2:

(5) *The compactness theorem* (a.k.a. the finiteness theorem). If every finite subset of a set of sentences Γ from a first-order language 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. Look at the details of the proof: it gives an abstract recipe for building the required model. And assuming that we are dealing with normal first-order languages (with a countable vocabulary), you'll find that the recipe delivers a *countable* model – so in effect, our proof shows that a syntactically consistent set of sentences has a model whose domain is just (some or all) the natural numbers. From this observation we get

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

But 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 going via the completeness theorem about 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.[1] In the second-order case we can have (i) and (ii), so that rules out compactness. In the first-order case, we have (i) and (iii); hence

(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. We met a another limitation, similarly proved, in §4.2 when we showed that we cannot define the ancestral of a relation in first-order terms. But now let's note a much more dramatic limitative 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 $'$ (expressing the successor function), and the constant $\mathsf{0}$ (denoting zero). So note that L_A then has a sequence of expressions $\mathsf{0}, \mathsf{0}', \mathsf{0}'', \mathsf{0}''', \ldots$ which can serve as numerals, denoting 0, 1, 2, 3, ...

Now let 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 mul-

[1]Consider the infinite set of sentences

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

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. So compactness fails.

tiplication functions 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]

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 uniquely pin down the wildly infinitary universe of sets – it doesn't even manage to pin down an uncountable universe.

What is emerging then, in these first steps into model theory, are some very considerable and perhaps unexpected(?) expressive limitations of first-order formalized theories (in addition to those we touched on in §4.2). These limitations can be thought of as one of the main themes of Level 1 model theory.

[2] Indulge me! Let me give the proof idea, because it is so very neat. For brevity, write $\overline{\mathsf{n}}$ as short for $\mathsf{0}$ followed by n occurrences of the prime $'$: so $\overline{\mathsf{n}}$ denotes n.

OK: 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 $\Delta \subset T^+_{true}$ has a model, T^+_{true} itself has a model, by compactness. That model, as well as having a zero and its successors, must also have in its domain a 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 note, since the new model must still make true e.g. 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 does not 'look like' the original intended model.

(e) 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); 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 (even if we restrict ourselves to models in the natural numbers, there can be models which give deviant interpretations of successor, addition and multiplication).

How does that last claim square with the proof you often meet early in a maths course that a theory usually called 'Peano Arithmetic' *is* categorical? The answer is straightforward. As already indicated in (3) above, 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 (see Ch 4, and follow up the readings if you didn't do so before.)

(f) Still at Level 2, there are results about which theories are *complete* in the sense of entailing either A or $\neg A$ for each relevant sentence A, 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) As explained in the last footnote, 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. We can now also take a standard first-order of the real numbers and use a similar compactness argument to show that it has a non-standard model with an

element r in the domain such 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 our model will in fact 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 at all). Fascinating!

5.2 Recommendations for beginning first-order model theory

A preliminary point. When exploring model theory you will very quickly encounter talk of different infinite cardinalities, and also occasional references to the Axiom of Choice. You need to be familiar enough with these basic set-theoretic ideas (perhaps from the readings suggested back in Chapter 2).

Let's begin with a more expansive and very helpful overview (though you may not understand everything at this preliminary stage). For a bit more detail about the initial agenda of model theory, it is hard to beat

1. Wilfrid Hodges, 'Model theory', in the *The Stanford Encyclopaedia of Philosophy* at tinyurl.com/sepmodel.

Now, a number of the introductions to FOL that I noted in §3.5 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 nowadays really off-putting to read. What, then, are the more recent options?

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

3. María Manzano, *Model Theory*, Oxford Logic Guides 37 (OUP, 1999).
   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 and compactness theorems again, and then has a sequence of chapters on various core model-theoretic notions and proofs. This should all be tolerably accessibly (especially if not your very first encounter with model theoretic ideas).

It seems to me that Manzano's discussions at some points would have benefitted from rather more informal commentary, motivating various choices, and sometimes the symbolism is unnecessarily heavy-handed. But overall, Manzano's text could work well enough as a follow-up to Goldrei. For more details, see tinyurl.com/manzanobook.

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:

4. 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 *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 tinyurl.com/kirbybooknote.

Finally, we noted that first-order theories behave differently from second-order theories where we have quantifiers running over all 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 §4.3.

5.3 Some parallel and slightly more advanced reading

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 §2.8: but this is perhaps too little done too fast.

So I think the following suits our needs here better:

5. Dirk van Dalen *Logic and Structure* (Springer, 1980; 5th edition 2012), Chapter 3.

   This covers rather more model-theoretic material than Enderton and in greater 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 some slightly more advanced topics.

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

6. 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.

For rather more detail, here is a recent book with an enticing title:

7. 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 very nicely done. However, I don't think that Kossak's Chapter 8 will be found an outstandingly clear 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 starts to require 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 might begin to struggle after the half-way mark. But this is very nice addition to the introductory literature.

Thanks to the efforts of the respective authors to write very accessibly, the suggested main path into the foothills of model theory (from Chiswell & Hodges → Leary & Kristiansen → Goldrei → Manzano/Kirby/Kossack) is not at all a hard road to follow.

Now, 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, then, 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 e.g. Manzano'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, beyond what I'd regard as elementary model theory. For more, see tinyurl.com/hedmanbook.

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

9. Tim Button and Sean Walsh, *Philosophy and Model Theory** (OUP, 2018).

   This book explains technical results in model theory, and explores the appeals to model theory in various branches of philosophy, particularly philosophy of mathematics, but also in metaphysics more generally, 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 needed technical results 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 do rather suspect that, despite their best efforts, an amount of the material is more difficult than 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 with judicious skimming/skipping – the signposting in the book is excellent – many mathematicians 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 go further into model theory; so we pick up the story again in §12.2.

5.4 A little history

The last book we mentioned includes a historical appendix from a now familiar author:

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

Read the first six or so sections. Later sections refer to model theoretic topics a level up from our current more elementary concerns, so won't be very accessible at this stage. For another piece that focuses on topics from the beginning of model theory, you could perhaps try R. L. Vaught's 'Model theory before 1945' in L. Henkin et al, eds, *Proceedings of the Tarski Symposium* (American Mathematical Society, 1974), pp. 153–172. You'll probably have to skim parts, but it will also give you some idea of the early developments.

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:

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

6 Arithmetic, computability, and incompleteness

The standard mathematical logic curriculum, as well as looking at elementary results about formalized theories and their models in general, investigates 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 is well worth pausing for a quick general point about why the idea of computability is of such very central concern to formal logic.

6.1 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 clear 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. We want to be able to settle these things in a way which leaves absolutely no room left for doubt or dispute.

(b) As a step towards sharpening this thought, 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.

To put it only slightly different words, a property P 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 Q 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: we can effectively decide what is the main connective of a sentence (by bracket counting), and the property of being a tautology is effectively decidable (by a truth-table calculation).

And the point we made at the outset in (a) now comes to this: we will want it to be effectively decidable e.g. whether a given string of symbols has the property of being 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.

(c) 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 effectively decide in advance whether P actually *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 in fact conforms to the rules of the game. But once we are dealing with an even mildly interesting T, there will be no way of deciding in advance whether a T-proof of P exists. Such a theory T is said to be (effectively) undecidable.

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

(d) 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 always represent facts about finite whatnots like formulas and proofs by using numerical codings. We can then trade in questions about formulas or proofs for questions about their code numbers.

2. And as a second step, we can also 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, by those two steps, we can quickly move from e.g. the question whether it is effectively decidable whether a string of symbols is a wff to a corresponding question about whether a certain numerical function is computable.

6.2 Computable functions

(a) For convenience, we will now use 'S' for the function that maps a number to its successor (where we previously used a prime). Consider, then, the following pairs of equations:

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

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

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

In some notation or other, these pairs of equations should be very familiar: they in turn define addition, multiplication and exponentiation for the natural numbers.

At the risk of labouring the obvious, let's spell out the point. 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 do so by giving 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 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/or by plugging one function we already know about into another. Such a function is said to be *primitive recursive.*

And generalizing from the case of exponentiation, we have the following observation:

> Any primitive recursive function is effectively computable.

(c) So far, so good. However, it is easy to show that

> Not all effectively computable functions are primitive recursive.

A neat abstract argument proves the point.[1] But this raises an obvious question: 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 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.*

[1] Roughly, we can effectively list off the primitive recursive functions by listing their recipes; so we have an algorithm which gives us f_n, the n-th such function. Then define the function d by putting $d(n) = f_n(n) + 1$. Evidently, d differs from any f_n for the value n, so isn't one of the primitive recursive functions. But it is computable.

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 implementing such searches, using a 'minimization' operator – roughly, μxFx sets us off on a search through increasing values of x and returns the least x which satisfies the condition F. Let's not worry about the details now; though note that since there might not be a value of x which satisfies F, the minimization operator may not return a value, so using it may only define a partial function. However those *total* functions which can be computed by a chain of applications of primitive recursion and/or open-ended searches implemented by the minimization operator are called (simply) *recursive.*

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

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, Alan 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 the claim 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'). 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? Why does it matter? I leave it to you to read up on the 'undecidability of the halting problem', and its many weighty implications.

6.3 Formal arithmetic

(a) 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 principle. As noted in §4.2, in a first-order framework we can stipulate that

> Any wff of the form $(\{A(\mathsf{0}) \wedge \forall \mathsf{x}(A(\mathsf{x}) \to A(\mathsf{Sx}))\} \to \forall \mathsf{x} A(\mathsf{x}))$ is an axiom,

where $A(\ )$ stands in for some suitable expression. Or obviously equivalently, we can formulate the same idea as an inference rule:

> From $A(\mathsf{0})$ and $\forall \mathsf{x}(A(\mathsf{x}) \to A(\mathsf{Sx}))$ we can infer $\forall \mathsf{x} A(\mathsf{x})$.

You need to get some elementary familiarity with the resulting theory.

(b) But why concentrate on first-order PA? We've emphasized in §4.2 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, we have already addressed this in Chapter 4: first-order logic is much better behaved than second-order logic. And some would say that second-order logic is really just a bit of set theory in disguise. So, the argument goes, if we want a theory of pure arithmetic, one whose logic can be formalized, we should stick to a first-order formulation just quantifying over numbers. Then something like PA's induction rule (or the suite of axioms of the form we described) is the best we can do.

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 $\mathsf{F}(\ ,\ )$ such that $\mathsf{F}(\overline{\mathsf{m}}, \overline{\mathsf{n}})$ is true if and only if $f(m) = n$. Moreover, when $f(m) = n$, PA can *prove* $\mathsf{F}(\overline{\mathsf{m}}, \overline{\mathsf{n}})$, and when $f(m) \neq n$, PA can *prove* $\neg\mathsf{F}(\overline{\mathsf{m}}, \overline{\mathsf{n}})$. In this way, PA as it were already has the resources to capture all the recursive functions and can compute their values. Similarly, PA can already capture any algorithmically decidable relation.

So PA is expressively a *lot* richer than you might initially suppose. And it turns out that even an 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, we quickly get 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).

(c) 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 something about what some of these non-standard models can look like (they have a copy of the natural numbers in their domains but also additional 'non-standard numbers'). 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).

6.4 Towards Gödelian incompleteness

(i) Now for our third related topic: Gödel's epoch-making incompleteness theorems. We'll look at the first of the two theorems here.

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 a theory like PA must be *negation incomplete* – meaning that we can form a sentence G in its language such that PA proves neither G nor $\neg\mathsf{G}$. 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 here goes:

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 code number for an expression (or a sequence of expressions) is its unique *Gödel number*.

G2. We can then 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 (its conclusion).

G3. 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. Since PA can express any algorithmically decidable *relation*, there will in particular be a formal expression in the language of PA which we can abbreviate Prf which expresses the effectively decidable 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 $\mathsf{Prov(y)}$ to be the expression $\exists \mathsf{x} \mathsf{Prf}(\mathsf{x}, \mathsf{y})$. Then $\mathsf{Prov}(\overline{\mathsf{n}})$, i.e. $\exists \mathsf{x} \mathsf{Prf}(\mathsf{x}, \overline{\mathsf{n}})$, is true if and only if 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 Prov 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 $\overline{\mathsf{g}}$ is the numeral for g, the code number of G.

Don't worry for the moment about how we do this construction (it involves a so-called 'diagonalization' trick which 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 $\neg\mathsf{G}$ is false. Hence, still assuming PA is sound, it can't prove $\neg\mathsf{G}$ either.

So, in sum, assuming PA is sound, it can't prove either of G or $\neg\mathsf{G}$. As announced, PA is negation incomplete.

Wonderful!

(iii) Now the argument generalizes to other nicely axiomatized sound theories T which can express enough arithmetical truths. 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.*

(iv) Now, we haven't quite arrived at what's usually called the First Incompleteness Theorem. For that, we need an extra step Gödel took, which enables us to drop the semantic assumption that we are dealing with a *sound* theory T for a weaker consistency requirement. But I'llnow leave you to explore the (not very difficult) details, and also to find out about the Second Theorem.

It really is time to start reading!

6.5 Main recommendations on arithmetic, etc.

I hope those overviews were enough to pique your interest. But if you want a more expansive introduction to the territory, then you can very usefully 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).

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 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. First, then, there is

3. Peter Smith, *Gödel Without (Too Many) Tears** (Logic Matters, 2020): 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 it perhaps gives 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, you might well prefer to jump straight into one of the following:

4. Peter Smith, *An Introduction to Gödel's Theorems** (2nd edition CUP, 2013: also now downloadable from logicmatters.net/igt).

   Three times the length of *GWT* and ranging more widely, this starts by informally exploring various ideas such as effective computability, and then it proves two correspondingly 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 then 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).

5. Richard Epstein and Walter Carnielli, *Computability: Computable Functions, Logic, and the Foundations of Mathematics* (Wadsworth 2nd edn. 2000: Advanced Reasoning Forum 3rd edn. 2008).

   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 second 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 or PA 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, and a very good historical appendix too.

Those two 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 us half-a-step up in mathematical sophistication, we arrive at a quite delightful book:

6. George Boolos and Richard Jeffrey, *Computability and Logic* (CUP 3rd edn. 1990).

   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. The guiding idea is easy enough, but there is tedious work to be done in implementing it. If you *really* want more details, see e.g. the book by Boolos mentioned in §10.4, or eventually look at the final chapter of the book by Rautenberg mentioned in §12.3.

6.6 Some parallel/additional reading

I should start by mentioning a more elementary book which might well appeal to some for its debunking of myths about the wider significance of Gödelian incompleteness:

7. Torkel Franzén, *Gödel's Theorem: An Incomplete Guide to its Use and Abuse* (A. K. Peters, 2005).

   John Dawson (whom we'll meet again in §6.7) 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!

And next, here's a group of three books at about the same level as those mentioned in the previous section. First, from the Open Logic Project:

8. Jeremy Avigad and Richard Zach, *Incompleteness and Computability: An Open Introduction to Gödel's Theorems**, tinyurl.com/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.

But really, you should take 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, *Computability Theory: An Introduction to Recursion 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'. 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 is 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* (recommended in §3.5) 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**, tinyurl.com/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:

    i. 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.

    ii. 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 at a more advanced level in §12.3. For now, I will mention just four 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.

And next I should mention a very nice book about Gödelian matters:

12. Torkel Franzén, *Inexaustibility: A Non-exhaustive Treatment* (Association for Symbolic Logic/A. K. Peters, 2004). The first two-thirds of the book gives another very readable take on logic, arithmetic, computability and incompleteness. It also interweaves some discussion of ordinals for proof-theoretic applications (a topic that will concern us later). The final chapters tackle a more advanced theme and we'll return to them in §12.3.

We now come to an absolutely stand-out book that you should certainly tackle at some point. But though this starts from scratch, 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, which is why I only mention it now:

13. Raymond Smullyan, *Gödel's Incompleteness Theorems*, Oxford Logic Guides 19 (Clarendon Press, 1992).
    This is delightfully 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 extremely rewarding.

To introduce the fourth 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

14. 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 highly 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."

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

15. Douglas Hofstadter, *Gödel, Escher, Bach** (Penguin, 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.

6.7 A little history

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

16. 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 §6.5.

This will really give you the headline news you initially need. It is then well worth reading

17. Robin Gandy, 'The confluence of ideas in 1936' in R. Herken, ed., *The Universal Turing Machine: A Half-century Survey* (OUP 1988). This seeks to explain why so many of the absolutely key notions all got formed in the mid-thirties.

And then you might enjoy

18. Charles Petzold, *The Annotated Turing* (Wiley, 2008) And intriguing mix of historical context and an extensively annotated exposition of Turing's great 1936 paper 'On Computable Numbers . . . '.
19. 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 §5.4 – but then Gödel was such a very different man. Fascinating, though!

(As far as getting any logical insights goes, you can simply ignore Stephen Budiansky *Journey to the Edge of Reason: The Life of Kurt Gödel*, OUP 2021.)

7 Set theory, less naively

In Chapter 2, 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 §12.4 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 here in this chapter we concentrate on some core basics.

7.1 Set theory and number systems

You won't need to have done very much mathematics at all for there to be no real news for you in this section: feel free to skim and skip.

(a) If you have 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 2.

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

It only takes a finite amount of data to fully specify a particular natural number. Similarly for integers and rational numbers. But not so, in general, for real numbers. As is very familiar, a real can be approached by a sequence of ever-closer rational approximations; but the sequence need never terminate. We need a framework for reasoning about such non-finite data. Set theory provides this. How?

Assume, for the moment, that we already have the rational numbers to hand. Let's now 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. In other words, 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 any sequence of rationals like 1.4, 1.41, 1.414, 1.4142, 1.41421, ..., i.e. rationals whose squares approach 2. And what's a sequence? We can treat an ordered sequence $s_1, s_2, s_3, \ldots$ as a set of pairs $\{\langle 1, s_1\rangle, \langle 2, s_2\rangle, \langle 3, s_3\rangle, \ldots\}$.

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 non-negative 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.

On either approach, real numbers are identified with sets (or sets of sets of sets) of rationals. 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 – these real numbers do indeed have the properties assumed in our informal working theory of real analysis. And given that our set theory is consistent, the resulting theory will be consistent too. Excellent!

We can now go on define functions between real numbers in terms of sets of ordered tuples of reals, so we can develop a theory of analysis. I am not going to spell this out further here. However, you do want to 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.

(c) Now, as far as the 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 universe of rational numbers governed by a prior suite of numerical laws. And that would be entirely fine.

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. Or at least, we will have constructed such a sequence so long as we are allowed to iterate an infinite number of times the operation of forming new sets by applying the 'set of' operation to sets that we have already constructed; and that is an important new idea. But if we *do* allow that, then elementary further reasoning about sets will show that the familiar arithmetic laws about natural numbers will apply to numbers as just constructed (including e.g. the principle of arithmetical induction).

Once we have a natural number sequence in play 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 $\langle m, n \rangle$ of natural numbers. But since for given m and n, $m - n = m' - n'$ 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 – at least once in your mathematical life! – you will want to see them worked through in enough detail to confirm that these constructions can indeed all be done. The point 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 noted in §2.1(vi), a simple argument shows that the set of infinite binary strings is *not* countably infinite. Now, such a 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 all subsets of the naturals, is also not countably infinite.

Note too that a real number between 0 and 1 can be represented in binary by an infinite string. And, by the same argument as before, for any countable list of reals-in-binary between 0 and 1, there will be another such real not on the list. Hence the set of real numbers between 0 and 1 is again not countably infinite. Hence neither is the set of *all* the reals.

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 is too big to be put into one-to-one correspondence with just the natural numbers and is too small to 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. And efforts to confirm or refute the Continuum Hypothesis were a major driver in early developments of set theory. We now know the problem is indeed a profound one – the standard axioms of set theory don'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.

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.

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

Starting perhaps from some given urelements – i.e. 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. Think in terms of a hierarchy of levels – *cumulative* levels, in the sense that a given level still contains all the urelements and all the sets that occur at earlier levels. Then at the next level we add all the new sets which have as members urelements and/or sets which can already be found at the current level. And we keep on going, adding more and more levels.

Now, for purely mathematical purposes such as reconstructing analysis, it seems that we only need a single non-membered base-level entity, and it is tidy to think of this as the empty set. So for internal mathematical purposes, we can take the whole universe of sets to contain only 'pure' sets (when we dig down and look at the members of members of . . . members of sets, we find nothing other than more sets).

But what if we want to be able to apply our 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. However, we can always *code* for widgets, wombats or space-time points using some kind of numbers, and we can treat those numbers as sets. So 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.

7.2 Ordinals, cardinals, and more

(a) Lots of questions arise from the rough-and-ready discussion so far. Here are two of the most pressing ones:

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 once 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 indeed, 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 it will be neat if we can identify these cardinal numbers 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 a great deal about this here, but some initial pointers might still be useful.

(b) We need to start from the notion of a *well-ordered set*. That's a set X together with an order-relation $\prec$ such that (i) $\prec$ is a linear order, and (ii) any subset $S \subseteq X$ has a $\prec$-least member.

For familiar examples, the rational numbers in their natural order are linearly ordered but not well-ordered (e.g. the set of rationals greater than zero has no least member). By contrast, the natural numbers in *their* natural order are well-ordered: the Least Number Principle tells us that, in any set of natural numbers, there is a least one.

Now, an absolutely key fact here is that – just as we can argue by induction over the natural numbers – we can argue by induction over other well-ordered sets. I need to explain.

In your reading on arithmetic, you should have met the so-called Strong Induction Principle (a.k.a. Course-of-Values Induction):[1] this says that, if a number has the property P whenever *all* smaller numbers have that property, then every number has P. This is quite easily seen to be equivalent to the ordinary induction principle we've encountered before in this Guide (§§4.2, 6.3): but this version is the one to focus in the present context.

We can now show that an exactly analogous induction principle holds whenever we are dealing with a set which, like the natural numbers, is also well-ordered. Assume X is well-ordered by the order relation $\prec$. Then the following induction principle holds for any property P:

(W-Ind) Suppose an object x in X has property P if all its $\prec$ predecessors already have property P: then *all* objects in X have property P.

Or putting that semi-formally,

Suppose for any $x \in X$, $(\forall z \prec x)Pz$ implies Px: then for all $x \in X$, Px

Why so?[2] Suppose (i) for any $x \in X$, $(\forall z \prec x)Pz$ implies Px, but also (ii) it *isn't* the case that for all $x \in X$, Px. Then by (ii) there must be some objects in X which don't have property P, and hence by the assumption that X is well-ordered, there is a $\prec$-least such object m such that *not*-Pm. But since m is the $\prec$-least such object, $(\forall z \prec m)Pz$ is true, and by (i) that implies Pm. Contradiction!

(c) Coming down from that level of abstraction, let's now look at some simple examples of well-ordered sets.

Here, then, 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: it is a linear order and, for any numbers we take, one will be the $\sqsubset$-least.

Now let's ask: 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.

[1] See e.g. my *Introduction to Gödel's Theorems*, §9.2.

[2] We in fact are just going to use the same line of argument that you may have seen being used to show that the Least Number Principle implies Strong Induction.

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 relative 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 $\omega + \omega$. And note, since this sequence is well-ordered, we can (if we want) base induction arguments on it – and an induction which takes us transfinitely far along such a sequence is naturally enough called a transfinite induction.

Here's another example. There are familiar maps for coding ordered pairs of natural numbers by a single natural, such as 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.) Since every number is equal to some unique $[m, n]$, $\prec$ is another well-ordering of 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; so the five blocks are in total $\omega \cdot 5$ 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 – hence these sequences are all equinumerous, and can be mapped one-to-one to each other (ignoring order). So they have the same infinite *cardinal* size, but the well-orders are tallied by different infinite *ordinal* numbers. Or so we want to say.

(d) But hold on! Is this sort of talk of transfinite ordinals really legitimate?

Well, it was one of Cantor's great and lasting achievements to make a start at showing that we *can* start to 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 infinite ordinals and their arithmetic (adding, multiplication and exponentiation).

But 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 various orderings. And there is in fact 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 at all entangle us with the kind of rather wildly infinitary and non-constructive ideas characteristic of modern set theory.

(e) However, here in this chapter 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. Recall the idea that we can implement the *finite* natural numbers by starting from the empty set, and taking the set of that, then the set of what we now have, and so on, forming at each stage the set of what we have already constructed:

$$0; \{0\}; \{0, \{0\}\}; \{0, \{0\}, \{0, \{0\}\}\}; \{0, \{0\}, \{0, \{0\}\}, \{0, \{0\}, \{0, \{0\}\}\}\}; \ldots$$

Now the idea is that we iterate this construction into the transfinite. The resulting well-ordered sequence of sets – call them the ordinals$_{\text{vN}}$ – provides a universal measuring scale against which to tally the length of any well-ordering. In other words, any well-ordered collection of objects, however long the ordering, will have the same type of ordering as an initial segment of these ordinals$_{\text{vN}}$. And note, we just said that these ordinals themselves are well-ordered by size: so we will be able to do induction arguments along them – 'transfinite' induction.

And at this point, I'll have to leave it to you to explore the details of the construction of the ordinals$_{\mathrm{vN}}$, and to learn e.g. about transfinite induction along the ordinals in the recommended readings. But once we have the ordinals available, we can say more about the way that the universe of sets is structured; we can take the levels to be well-ordered as we build up the universe in stages, and so the levels will be indexed by some ordinals$_{\mathrm{vN}}$. But there seems to be no natural stopping place; so we arrive at the idea that for every ordinal there is a corresponding level of the universe of sets.

(f) We can now also define a scale of cardinal size. We have already seen that well-orderings of different ordinal length can be equinumerous; hence 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 to leave for the moment for you to explore in the readings. 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 when beginning set theory: where do we start to need to appeal to some Choice principle?)

(g) Ah, I've been getting rather carried away! We are perhaps already rather past the point where scene-setting remarks at this level of generality can be very helpful. Time for you to dive into the details.

One final important observation, however, 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 §2.2 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 about sets that mathematicians actually *needed*, 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, both Zermelo himself and also 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 textbooks 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.

7.3 Main recommendations on set theory

This present chapter is, as advertised, just about the basics of set theory. Even here, however, there 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 let me mention again an introduction to set-theoretic ideas which I noted in §2.3, which you may have skipped past then.

2. Cambridge lecture notes by Tim Button have become incorporated into *Set Theory: An Open Introduction** (2019) tinyurl.com/opensettheory.

   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.

Since Button can't really get into enough detail into his brisk notes, 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 material in rather more depth but still very accessibly:

3. 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 §3.4 – is exceptionally clear, and it is 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 two overview sections. 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.

4. 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 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:

5. Winfried Just and Martin Weese, *Discovering Modern Set Theory I: The Basics* (American Mathematical Society, 1996).

   Covers similar 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, though, 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.)

6. 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 still have a lot to teach us even after we have read the more recent books, and this is very illuminating:

7. 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. make 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

8. 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.

7.4 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.

Let me begin by mentioning a bare-bones, introductory book, a level or so down in coverage and detail from what we really want here, but which some might find a helpful preliminary read:

9. 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". He proceeds pretty naively in the second sense we identified in §2.2. 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. Some might warm to this classic (though perhaps you should ignore the remarks in the Preface about set theory for applications being 'pretty trivial stuff'!).

Next, here are four introductory books at the right sort of level, listed in order of publication; each has many things to recommend it to beginners. Browse through to see which might suit your interests:

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 ZFC 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 of 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).

    Eventually this book 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 in §12.4, and Hrbacek once won a AMA prize for maths writing. So, unsurprisingly, this is a very nicely put together book, which could very well have featured as a main recommendation.

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 first 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 §12.4.)

13. Judith Roitman, *Introduction to Modern Set Theory** (Wiley, 1990: a 2011 version is available at tinyurl.com/roitmanset).

    Relatively short, and very engagingly written, this book covers 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. 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).

    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)

    This is 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 here:

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 and want to see set theory being applied; you could tackle Chs 6 to 8 on the basis of other introductions.

7.5 Further conceptual reflection on set theories

(a) A preliminary point. Go back to our starting point when we introduced set theory as giving us a 'foundation' for real analysis. But what does that really mean? As Penelope Maddy notes, "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 account of what might be meant by talk of foundations in this context. It is *very* well worth reading for orientation:

17. Penelope Maddy, 'Set-theoretic foundations', in A. Caicedo et al., eds., *Foundations of Mathematics* (AMS, 2017), tinyurl.com/maddy-found. See §1 in particular.

(b) Michael Potter's *Set Theory and Its Philosophy* must be the starting point for further 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

18. 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. Knowledgeable and readable.

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

19. 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).

7.6 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 tinyurl.com/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 will probably need to be on top of rather more set theory before getting to grips with *that*.

7.7 Postscript: Other treatments?

What else is there? A classic introduction is given by Patrick Suppes, *Axiomatic Set Theory** (vast Nostrand 1960, republished by Dover 1972). Clear and straightforward as far as it goes: but there are better alternatives now. There is also another classic book by Azriel Levy with the inviting title *Basic Set Theory** (Springer 1979, republished by Dover 2002). However, while this is still '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 §12.4.

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. You can probably already skip over the initial chapter on FOL, 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 considerably 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, some simple constructions and basic results take a *very* long time to arrive. For example, we don't actually get to Cantor's theorem on the uncountability of $\mathcal{P}(\omega)$ until p. 455, long after we have met more sophisticated results. So while this book might be worth dipping into for some of the motivational explanations, I can't myself recommend it overall.

Finally, I also can't recommend Daniel W. Cunningham's *Set Theory: A First Course* (CUP, 2016). Its old-school Definition/Lemma/Theorem/Proof style just doesn't make for an inviting introduction for self-study.

8 Intuitionistic logic

In the briefest headline terms, intuitionistic logic is what you get if you drop the classical principle that $\neg\neg A$ implies A (or equivalently drop the law of excluded middle which says that $A \vee \neg A$ always holds). But why would we want to do *that*? And what further consequences for our logic does that have?

8.1 A formal system

(a) To fix ideas, it will help to have in front of us a particular natural deduction system in Gentzen style, initially for propositional logic.

We assume that at least the three binary connectives $\wedge, \vee, \rightarrow$ are built in, together with the absurdity constant $\bot$.

The connectives are then governed by pairs of introduction and elimination rules. For the record (and for future reference), here are the usual introduction rules, presented in the short-hand way you should now be familiar with from work on standard FOL:

$$(\wedge\text{I})\quad \frac{A \qquad B}{A \wedge B} \qquad\qquad (\vee\text{I})\quad \frac{A}{A \vee B} \quad \frac{B}{A \vee B} \qquad\qquad (\rightarrow\text{I})\quad \begin{array}{c}[A]\\ \vdots \\ B \\ \hline A \rightarrow B\end{array}$$

Each elimination rule then in effect just undoes an application of the corresponding introduction rule (putting it roughly, for each binary connective $\diamond$, its elimination rule allows us to argue onwards from $A \diamond B$ to a conclusion that we could *already* have derived from what was required to derive $A \diamond B$ by its introduction rule):

$$(\wedge\text{E})\quad \frac{A \wedge B}{A} \quad \frac{A \wedge B}{B} \qquad (\vee\text{E})\quad \begin{array}{ccc} & [A] & [B] \\ & \vdots & \vdots \\ A \vee B & C & C \\ \hline & C & \end{array} \qquad (\rightarrow\text{E})\quad \frac{A \qquad A \rightarrow B}{B}$$

We next take the absurdity constant to be governed by the rule that given $\bot$ we can derive anything – *ex falso, quodlibet*.

Finally, what about negation? One option is to treat $\neg A$ as simply an abbreviation for $A \rightarrow \bot$. The introduction and elimination rules given for the conditional then immediately yield the following as special cases:

$$(\neg\text{I}) \quad \dfrac{\begin{matrix}[A]\\ \vdots \\ \bot\end{matrix}}{\neg A} \qquad (\neg\text{E}) \quad \dfrac{A \qquad \neg A}{\bot}$$

Alternatively, we can take these to be the introduction and elimination rules governing a primitive built-in negation connective. Nothing hangs on this choice.

We then define IPL, intuitionistic propositional logic (in its natural deduction version), to be the logic governed by these rules.

The described rules are of course all rules of classical logic too. However, the intuitionistic system is strictly weaker in the sense that the following classically acceptable principles are *not* derived rules of our intuitionistic logic:

$$(\text{DN}) \quad \dfrac{\neg\neg A}{A} \qquad (\text{LEM}) \quad \dfrac{}{A \vee \neg A} \qquad (\text{CR}) \quad \dfrac{\begin{matrix}[\neg A]\\ \vdots \\ \bot\end{matrix}}{A}$$

DN allows us to drop double negations. LEM is the Law of Excluded Middle, which permits us to infer $A \vee \neg A$ whenever we want, from no assumptions. CR is the classical reductio rule. And these three rules are equivalent in the sense that adding any one of them to intuitionistic propositional logic enables us to prove all the same conclusions; each way, we get back full classical propositional logic.

(b) If only for brevity's sake, we will largely be concentrating on propositional logic in the two introductory overviews which follow. But we should briefly note what it takes to get intuitionistic predicate logic in natural deduction form.

Technically, it's very straightforward. Just as the rules for $\wedge$ and $\vee$ are the same in classical and intuitionist logic, the rules for generalized conjunctions and generalized disjunctions remain the same too. In other words, to get intuitionistic predicate logic we simply add to IPL the same two pairs of introduction and elimination rules for $\forall$ and $\exists$ as for classical logic.

But note, because of the different background propositional logic – in particular, because of the different rules concerning negation – these familiar quantifier rules no longer have all the same implications in the intuitionistic setting. For example $\exists x A(x)$ is no longer equivalent to $\neg\forall x \neg A(x)$. More about this below.

8.2 Why intuitionistic logic?

(a) A little experimentation quickly suggests that we indeed cannot derive an instance of excluded middle like $\mathsf{P} \vee \neg\mathsf{P}$ in IPL. But how can we *prove* that this is underivable?

There's a proof-theoretic argument. We examine the structure of proofs in IPL, and thereby show that we can only prove $A \vee B$ as a theorem (i.e. from no premisses) if there is a proof of A or a proof of B. Since neither P nor $\neg\mathsf{P}$ is a theorem of intuitionistic logic (with P atomic), it follows that $\mathsf{P} \vee \neg\mathsf{P}$ isn't a theorem either.

Alternatively, there's a semantic argument. We find some new, non-classical, way of interpreting IPL as a formal system, an interpretation on which the intuitionistic rules of inference are still acceptable, but on which the double negation rule and its equivalents are clearly *not* acceptable. It will then follow that buying into IPL can't by itself commit us to those classical rules. How might this new interpretation go?

It is natural to think of a correct assertion as one that corresponds to some realm of facts (whatever that means exactly). But suppose just for a moment that we instead think of correctness as a matter of being *warranted*, where we understand this in the following strong sense: A is warranted if and only if there is an informal proof which provides a direct certification for A's correctness. Then here is a reasonably natural story about how to characterize the connectives in this new framework (it's a rough version of what's called the BHK – Brouwer-Heyting-Kolmorgorov – interpretation):

(i) $(A \land B)$ is warranted iff (if and only if) A and B are both warranted.

(ii) While there may be other ways of arriving at a disjunction, the direct and ideally informative way of certifying a disjunction's correctness is by establishing one or other disjunct. So we will count $(A \lor B)$ as warranted iff at least one disjunct is certified to be correct, i.e. iff there is a warrant for A or a warrant for B.

(iii) A warranted conditional $(A \to B)$ must be one that, together with the warranted assertion A, will enable us to derive another warranted assertion B by using modus ponens. Hence $(A \to B)$ is directly warranted iff there is a way of converting any warrant for A into a warrant for B.

(iv) $\neg A$ is warranted iff we have a warrant for ruling out A because it leads to something absurd (given what else is warranted).

(v) $\bot$ is never warranted.

Then, in keeping with this approach, we will think of a reliable inference as one that takes us from warranted premisses to a warranted conclusion.

Now, in this framework, the familiar *introduction* rules for the connectives will *still* be acceptable, for they will evidently be warrant-preserving (given our interpretation of the connectives). But as we said, the various elimination rules in effect just 'undo' the effects of the introduction rules: so they should come for free along with the introduction rules. Finally, we can still endorse EFQ, *ex falso quodlibet* – the plausible thought is that if, *per impossible*, the absurd is warrantedly assertible, then all hell breaks loose, and anything goes.

Hence, regarded now as warrant-preserving rules, all our IPL rules can remain in place. However:

1. DN will *not* be acceptable in this framework. We might have a warrant for ruling out being able actually to rule out A, so we can warrantedly assert $\neg\neg A$. But that doesn't put us in a position to warrantedly assert A. We might just have to remain neutral about A

2. Likewise LEM will *not* be acceptable. On the present understanding of the connectives, $(A \lor \neg A)$ would be correct, i.e. directly warranted, just if there is a warrant for A or a warrant for ruling out A. But must there always be a way of justifiably deciding a conjecture A in the relevant area of inquiry one way or the other? Some things may be beyond our ken.

Again, for similar reasons, CR is *not* acceptable either in this framework: but I won't keep mentioning this third rule.

In sum, then, if we want a propositional logic suitable as a framework for regimenting arguments which preserve warranted assertability, we should stick with the core rules of IPL – and shouldn't endorse those further distinctively classical laws.

But be very careful here! It is one thing to stand back from endorsing the law of excluded middle. It would be something else entirely actually to *deny* some instance of the law. In fact, it is an easy exercise to show that, even in IPL, any outright negation of an instance – i.e. any sentence of the form $\neg(A \lor \neg A)$ – entails absurdity.

(b) The double negation rule DN of classical logic is an outlier, not belonging to one of the matched pairs introduction/elimination rules. Now we see the significance of this. Its special status leaves room for an interpretation on which the remaining rules – the rules of IPL – hold good, but DN doesn't. Hence, as we wanted to show, DN is not derivable as a rule of intuitionistic propositional logic. Nor is LEM.

True, our version of the semantic argument as presented so far might seem all a bit too arm-waving for comfort; after all, the notion of warrant as we characterized it can hardly be said to be ideally clear! But let's not fuss about details now. We'll soon meet a rigorous story partially inspired by this notion which gives us an entirely uncontroversial, technically kosher, proof that DN and its equivalents are, as claimed, independent of the rules of IPL.

Things *do* get controversial, though, when it is claimed that DN and LEM really don't apply in some particular domain of inquiry, because in this domain there can be no more to correctness than having a warrant in the form of a direct informal proof. Now, so-called *intuitionists* do hold that mathematics is a case in point. Mathematical truth, they say, doesn't consist in correspondence with facts about abstract objects laid out in some Platonic heaven (after all, there are familiar worries: what kind of objects could these ideal mathematical entities be? how could we possibly know about them?). Rather, the story goes, the mathematical world is in some sense our construction, and being mathematically correct can be no more than a matter of being assertible on the basis of a proof elaborating our constructions – meaning not a proof in this or that formal system but a chain of reasoning satisfying informal mathematical standards for being a direct proof.

Consider, for example, the following argument, intended to show that (C), there is a pair of irrational numbers a and b such that a^b is rational:

> Either (i) $\sqrt{2}^{\sqrt{2}}$ is rational, or (ii) it isn't. In case (i) we are done: we can simply put $a = b = \sqrt{2}$, and hence (C) then holds. In case (ii) put $a = \sqrt{2}^{\sqrt{2}}$, $b = \sqrt{2}$. Then a is irrational by assumption, b is irrational, while $a^b = (\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = \sqrt{2}^2 = 2$ and hence is rational, so (C) again holds. Either way, (C).

It will be agreed on all sides that this argument isn't ideally satisfying. But the intuitionist goes further, and claims that this argument actually fails to establish (C), because we haven't yet constructed a specific a and b to warrant (C). The cited argument assumes that either (i) or (ii) holds, and – the intuitionist complains – we are not entitled to assume this when we are given no reason to suppose that one or other disjunct can be warranted by a construction.

(c) For an intuitionist, then, the appropriate logic is not full *classical* two-valued logic but rather our cut-down *intuitionistic* logic (hence the name!), because *this* is the right logic for correctness-as-informal-direct-provability.

Or so, roughly, goes the story. Plainly, we can't even begin to discuss here the highly contentious issues about the nature of truth and provability in mathematics which first led to the advocacy of intuitionistic logic (if you want to know a bit more, there are some initial references in the recommended reading). But no matter: there are plenty of other reasons too for being interested in intuitionistic logic, which keeps recurring in various contexts (e.g. in computer science and in category theory). And as we will see in the next chapter, the fact that its rules come in matched introduction/elimination pairs makes intuitionistic logic proof-theoretically particularly neat.

For now, let's just say a bit more about what can and can't be proved in IPL and its extension by the quantifier rules, and also introduce one of the more formal ways of semantically modelling it.

8.3 More proof theory, more semantics

(a) We use '$\vdash_{\text{C}}$' to symbolize classical derivability, and '$\vdash_{\text{I}}$' to symbolize derivability in intuitionistic logic. Then:

(i) The familiar classical laws governing just conjunctions and disjunctions stay the same: so, for example, we still have $A \wedge (B \wedge C) \vdash_{\text{I}} (A \wedge B) \wedge C$ and $A \vee (B \wedge C) \vdash_{\text{I}} (A \vee B) \wedge (A \vee C)$. However, although the conditional rules of inference are the same in classical and intuitionist logic, the laws governing the conditional are *not* the same. Classically, we have Peirce's Law, $(A \to B) \to A \vdash_{\text{C}} A$; but we do *not* have $(A \to B) \to A \vdash_{\text{I}} A$.

(ii) Classically, the binary connectives are interdefinable using negation. Not so in IPL. We do have for example $(A \vee B) \vdash_{\text{I}} \neg(\neg A \wedge \neg B)$. But the converse doesn't hold – a good rule of thumb is that IPL makes disjunctions harder to prove. However, $\neg(\neg A \wedge \neg B) \vdash_{\text{I}} \neg\neg(A \vee B)$.

Likewise, we do have $(\neg A \lor B) \vdash_{\text{I}} (A \to B)$. But the converse doesn't hold – though $(A \to B) \vdash_{\text{I}} \neg\neg(\neg A \lor B)$.

(iii) The connectives in IPL are not truth-functional. But their behaviour in a sense still tracks the classical truth-tables.

Take, for example, the classical table for the material conditional. We can read that as telling us that when A and B holds so does $A \to B$; when A holds and B doesn't (so $\neg B$ does), then $A \to B$ doesn't hold (so $\neg(A \to B)$ does); while when $\neg A$ holds, so does $(A \to B)$ (whether we also have B or $\neg B$).

Correspondingly, in intuitionistic logic, we still have $A, B \vdash_{\text{I}} (A \to B)$; $A, \neg B \vdash_{\text{I}} \neg(A \to B)$; and $\neg A \vdash_{\text{I}} (A \to B)$. The intuitionistic conditional therefore shares some of the same unwelcome(?) features as the classical material conditional.

(iv) Glivenko's theorem: if A is a propositional formula, $\vdash_{\text{C}} A$ just when $\vdash_{\text{I}} \neg\neg A$. Note, though, that this doesn't apply in general to quantified formulas.

(v) The so-called disjunction property applies in IPL, i.e. if $\Gamma \vdash_{\text{I}} (A \lor B)$ then either $\Gamma \vdash_{\text{I}} A$ or $\Gamma \vdash_{\text{I}} B$. And, moving to quantified intuitionistic logic, we have the following analogue: we only have $\Gamma \vdash_{\text{I}} \exists x Ax$ if we can provide a witness for the existentially quantified sentence, i.e. for some term t, $\Gamma \vdash_{\text{I}} At$.

(vi) Just as conjunction and disjunction are not intuitionistically interdefinable using negation, so too for the universal and existential quantifiers. Thus while $\exists x A \vdash_{\text{I}} \neg\forall x \neg A$, the converse doesn't hold – though, inserting a double negation, we do have $\neg\forall x \neg A \vdash_{\text{I}} \neg\neg\exists x A$. Likewise, $\forall x A \vdash_{\text{I}} \neg\exists x \neg A$. But again the converse doesn't hold – though $\neg\exists x \neg A \vdash_{\text{I}} \forall x \neg\neg A$.

(vii) A theme is emerging! While some classical results fail in intuitionistic logic, inserting some double negations *will* give corresponding intuitionistic results. This theme can be made more precise, in various ways. Consider, for example, the following translation scheme T for mapping classical to intuitionistic sentences – a *double-negation* translation:

a) $A^T := \neg\neg A$, for atomic wffs A; $\bot^T := \bot$
b) $(A \land B)^T := A^T \land B^T$
c) $(A \lor B)^T := \neg\neg(A^T \lor B^T)$
d) $(A \to B)^T := A^T \to B^T$
e) $(\neg A)^T := \neg A^T$
f) $(\forall x A)^T := \forall x A^T$
g) $(\exists x A)^T := \neg\neg\exists x A^T$

Suppose Γ^T comprises the double-negation translations of the sentences in the set Γ. Then we have the following key theorem due (independently) to Gödel and Gentzen:

$\Gamma \vdash_{\text{C}} A$ if and only if $\Gamma^T \vdash_{\text{I}} A^T$.

(b) Two comments on the Gödel/Gentzen theorem. First, it shows that for every classical result, there is already a corresponding intuitionistic one which has additional double negation signs in the right places. So we can think of classical logic not so much as what you get by *adding* to intuitionist logic but rather as what you get by *ignoring a distinction* that the intuitionist thinks is of central importance, namely the distinction between A and $\neg\neg A$.

Second, note this particular consequence of the theorem: $\Gamma \vdash_{\mathrm{C}} \bot$ if and only if $\Gamma^T \vdash_{\mathrm{I}} \bot$. So if the classical theory Γ is inconsistent by classical standards, then its translated version Γ^T is already inconsistent by intuitionistic standards. Roughly speaking, then, if we have worries about the consistency of a classical theory, retreating to an intuitionistic version isn't going to help. As you'll see from the readings, this observation had significant historical impact in debates in the foundations of mathematics.

(c) Let's now return to those earlier arm-waving semantic remarks in §8.2(a). They can be sharpened up in various ways, but here I'll just briefly consider (a version of) Saul Kripke's semantics for IPL. I'll leave it to you to find out how the story can be extended to cover quantified intuitionistic logic.

Take things in stages. First, imagine an enquirer, starting from a ground state of knowledge g; she then proceeds to expand her knowledge, through a sequence of possible further states K. Different routes forward can be possible, so we can think of these states as situated on a branching array of possibilities rooted at g (not strictly a 'tree' though, as we can allow branches to later rejoin, reflecting the fact that our enquirer can arrive at the same later knowledge state by different routes). If she can get from state $k \in K$ to the state $k' \in K$ by zero or more steps, then we'll write $k \leq k'$. So, to model the situation a bit more abstractly, let's say that

> An *intuitionistic model structure* is a triple $(g, K, \leq)$, where K is a set, $\leq$ is a partial order defined over K, and g is its minimum (so $g \leq k$ for all $k \in K$).

As our enquirer investigates the truth of the various sentences of her propositional language, at any stage k a sentence A is either *established to be true* or *not [yet] established*. We can symbolize those alternatives by $k \Vdash A$ and $k \nVdash A$; it is quite common, for reasons that needn't now detain us, to read '$\Vdash$' as *forces*. And, as far as *atomic* sentences are concerned, the only constraint on a forcing relation is this: once P is established in the knowledge state k, it stays established in any expansion on that state of knowledge, i.e. at any k' such that $k \leq k'$. Knowledge persists. Hence, again to put the point more abstractly, we require the forcing relation $\Vdash$ to satisfy this persistence condition:

> For any atomic sentence P and $k \in K$, if $k \Vdash P$, then $k' \Vdash P$, for all $k' \in K$ such that $k \leq k'$.

And now, next stage, let's expand a forcing relation defined for a suite of atoms so that it now covers *all* wffs built up from those atoms by the connectives. So, for all $k, k' \in K$, and all relevant sentences A, B, we will require

(i) $k \nVdash \bot$.
(ii) $k \Vdash A \wedge B$ iff $k \Vdash A$ and $k \Vdash B$.
(iii) $k \Vdash A \vee B$ iff $k \Vdash A$ or $k \Vdash B$.
(iv) $k \Vdash A \to B$ iff, for any k' such that $k \leq k'$, if $k' \Vdash A$ then $k' \Vdash B$.
(v) $k \Vdash \neg A$ iff, for any k' such that $k \leq k'$, $k' \nVdash A$.

It's a simple consequence of these conditions on a forcing relation that for any A, whether atomic or molecular,

($*$) If $k \Vdash A$, then $k' \Vdash A$, for all k' such that $k \leq k'$.

This formally reflects the idea that once A is established it stays established, whether or not it is an atom.

But what motivates those clauses (i) to (v) in our characterization of $\Vdash$? (i) The absurd is never established as true, in any state of knowledge. And (ii) establishing a conjunction is equivalent to establishing each conjunct, on any sensible story. So we needn't pause over these first two.

But (iii) reveals our enquirer's intuitionist/constructivist commitments! – as per the BHK interpretation, she is taking establishing a disjunction in an acceptably direct way to require establishing one of the disjuncts. For (iv) the thought is that establishing $A \to B$ is tantamount to giving you an inference-ticket: with the conditional established, if you (eventually) get to also establish A, then you will then be entitled to B too. Finally, (v) falls out from the definition of $\neg A$ as $A \to \bot$ and the evaluation rules for $\to$ and $\bot$. Or more directly, the idea is that to establish $\neg A$ is to rule out, once and for all, A turning out to be correct as we later expand our knowledge.

With these pieces in place, we can – next stage! – define a formula of a propositional language to be intuitionistically valid in a natural way. Classically, a propositional formula is valid (is a tautology) if it is true however things turn out with respect to the values of the relevant atoms. Now we say that a propositional formula A is intuitionistically valid if it can be established in the ground state of knowledge, however things later turn out with respect to the truth of relevant atoms as our knowledge expands. Putting that more formally,

> A is intuitionistically valid iff $g \Vdash A$, whatever the model structure $(g, K, \leq)$ and whatever forcing relation $\Vdash$ is defined over the relevant atoms.[1]

And now for the big reveal! Kripke proved in 1965 the following soundness and completeness result:

[1] Fine print, just to link up with other presentations you will meet. First, given ($*$), $g \Vdash A$ holds iff $k \Vdash A$ for all k. So we can redefine validity by saying A is valid just when $k \Vdash A$ for all k. But then, second, we can in fact let g drop right out of the picture. For it is quite easy to see that it will make no difference whether we require the partial order $\leq$ to have a minimum or not: the same sentences will come out valid either way. Third, we don't even require the relation we symbolized $\leq$ to be a true partial order: again, if we allow any reflexive, transitive relation over K in its place, it will make no different to what comes out as valid.

> A formula is a theorem of IPL (can be derived from no premisses) if and only if it is intuitionistically valid.

Neither direction of the biconditional is particularly hard.

Expanding the idea of valuations over an intuitionistic model structure to accommodate quantified formulas and then proving soundness and completeness for quantified intuitionistic logic is, however, rather more involved.

(d) Let's finish by briefly showing that – given Kripke's soundness result that every IPL theorem is intuitionistically valid on his semantic story – it is immediate that the law of excluded middle fails for IPL.

It couldn't be easier. Consider a propositional language with just a single atom P; and take the model structure which has just two states g, k such that $g \leq k$. And now suppose that P is not yet established at g but *is* established at k, hence $g \nVdash \mathsf{P}$ while $k \Vdash \mathsf{P}$. By the rule for negation, $g \nVdash \neg\mathsf{P}$. So $g \nVdash (\mathsf{P} \vee \neg\mathsf{P})$. Hence $\mathsf{P} \vee \neg\mathsf{P}$ is not valid. Hence, by the soundness result, $\mathsf{P} \vee \neg\mathsf{P}$ can't be an IPL theorem.

8.4 Basic recommendations on intuitionistic logic

So much for some introductory remarks – enough, I hope, to spark interest in the topic! There is room, then, for a short introductory book which would develop these and related themes at the kind of accessible level we currently want. And Grigori Mints's *A Short Introduction to Intuitionistic Logic* (Springer, 2000) is brief enough; however, it soon becomes entangled with more advanced topics in a way that will too quickly mystify beginners. So we will have to patch together readings from a few different sources.

We will cherry-pick from the following:

1. Joan Moschovakis, 'Intuitionistic logic', in *The Stanford Encyclopaedia of Philosophy*, §§1–3, §4.1, §5.1. Available at tinyurl.com/sep-intuit.
2. Dirk van Dalen, *Logic and Structure* (Springer, 1980; 5th edition 2012), Chapter 5, 'Intuitionistic logic'.
3. A.S. Troelstra and Dirk van Dalen, *Constructivism in Mathematics, An Introduction: Vol. I* (North-Holland, 1988), Chapter 2, 'Logic', §1, §3 (up to Prop 3.8), §4?, §5, §6?.

You could read these in the order given, initially skimming/skipping over passages that aren't immediately clear.

Or perhaps better, start with (1)'s §1, 'Rejection of Tertium Non Datur', and then (2)'s §5.1, 'Constructive reasoning' which introduces the BHK interpretation of the logical operators.

Then look at a presentation of a *natural deduction* system for intuitionistic logic (as sketched in our overview): this is briskly covered in (2) in the first half of §5.2. But in fact the discussion in (3) – though this is not an introductory textbook – is notably more relaxed and clearer: see §1 of the chapter.

Next, read up on the *double-negation translation* between classical and intuitionistic logic. This is described in (1) §4.1, and explored a bit more in the second half of (2) §5.2. But again, a more relaxed presentation can be found in (3), §3 (up to Prop. 3.8).

Now you want to find out more about *Kripke semantics*, which is also covered in all three resources. (1) §5.1 gives the brisk headline news. (2) gives a compressed account in the first half of §5.3. But again (3) is best: Troelstra and Van Dalen give a much more expansive and helpful account in their Ch. 2 §5 – which sensibly treats propositional logic first before expanding the story to cover full quantified intuitionistic logic.

I would suggest, though, leaving detailed soundness and completeness proofs for Kripke semantics – covered in (2) §5.3 or (3) §6 – for later (if they are tackled at all, at this stage.)

For a few more facts about intuitionistic logic, such as the *disjunction property*, see also the first couple of pages of (2) §5.4 (the rest of that section is interesting but not really needed at this stage).

Return to (1) to look at §2.1 (an *axiomatic* version of intuitionistic logic), and the first half of §3 (on Heyting's intuitionistic arithmetic). Then finally, for more on Heyting Arithmetic and a spelt-out proof that it is consistent if and only if classical Peano Arithmetic is consistent, you could dip into

4. Paolo Mancosu, Sergio Galvan, and Richard Zach, *An Introduction to Proof Theory*, (OUP, 2021). Their §2.15 on 'Intuitionistic and classical arithmetic' can be read as an approachable stand-alone treatment.

8.5 Some parallel/additional reading

Kripke semantics for intuitionistic logic involves evaluating formulas not once and for all but at different points in a relational structure. We informally talked about these points as various 'states of knowledge'; in a different idiom we could have talked about various 'possible worlds'. Now, the use of this kind of relational semantics is characteristic of modal logics – the simplest modal logics being logics of necessity and possibility, with their semantics modelling the idea that being necessarily true is being true at all suitably related possible worlds. So another way of approaching intuitionistic logic is by *first* discussing modal logics more generally, *before* looking at intuitionistic logic in particular. If you want to explore this route, you can jump to this Guide's Chapter 10. In particular, you could perhaps look at Graham Priest's terrific *An Introduction to Non-Classical Logic* mentioned there, which gives tableaux systems first for modal logic and then for intuitionistic logic.

There is also a different way using tableaux for intuitionistic logic (which doesn't rely on first treating modal logic), which is quite nicely explored by

5. Harrie de Swart, *Philosophical and Mathematical Logic* (Springer, 2018), Chapter 18.

However, I prefer the treatment of the same tableau approach in an earlier excellent book:

6. Melvin Fitting, *Intuitionistic Logic, Model Theory, and Forcing* (North Holland, 1969), Part I.

   Ignore the scary title: it is only the beautifully clear but sophisticated first part of the book which concerns us now! It should particularly appeal to those who appreciate mathematical elegance.

For a bit more on natural deduction, the sequent calculus and semantics for intuitionistic logic, you should look at two chapters from a modern classic:

7. Michael Dummett, *Elements of Intuitionism* (OUP, 2nd ed. 2000), Chapters 4 and 5.

In fact, you could well want to read the opening two chapters and the final one as well! There are then many more pointers to technical discussions in Moschovakis's section of 'Recommended reading'.

8.6 A little more history, a little more philosophy

A number of the readings mentioned so far include brief remarks about the history of intuitionism (and constructivism more generally). For something more substantial, look at

8. A.S. Troelstra and Dirk van Dalen, *Constructivism in Mathematics, An Introduction: Vol. I* (North-Holland, 1988), Chapter 1,

which gives a brief characterization of various forms of constructivism (not all of them motivate the adoption of a non-classical logic like intuitionistic logic).

The early days of intuitionism were wild! To get a sense of *how* wild Brouwer's ideas were, you could take a look at

9. Mark van Atten, *On Brouwer* (Wadsworth, 2004), Chapters 1 and 2.

The same author has a *The Stanford Encyclopedia* article on 'The Development of Intuitionistic Logic' at tinyurl.com/dev-intuit; but that's much more detailed than you are likely to want.

Turning to more philosophical discussions – and it is a bit difficult to separate thinking about intuitionism as a philosophy of mathematics from thinking about intuitionistic logic more specifically – one key article that you will want to read (which was hugely influential in reviving interest in a 'tamer' intuitionism among philosophers) is

10. Michael Dummett, 'The philosophical basis of intuitionistic logic' (originally 1973, reprinted in Dummett's *Truth and Other Engimas*).

Then, for more recent discussions, here's a trio of articles:

11. Carl Posy, 'Intuitionism and philosophy'; D. C. McCarty, 'Intuitionism in mathematics'; and Roy Cook, 'Intuitionism reconsidered', all in S. Shapiro, ed., *The Oxford Handbook of the Philosophy of Mathematics and Logic* (OUP, 2005).

9 Elementary proof theory

The story of proof theory starts with David Hilbert and what has come to be known as 'Hilbert's Programme', which inspired the profoundly original work of Gerhard Gentzen in the 1930s.

Two themes from Gentzen are within easy reach for beginners in mathematical logic: (A) the idea of normalization for natural deduction proofs, (B) the move from natural deduction to sequent calculi, and cut-elimination results for these calculi. But the most interesting later developments in proof theory – in particular, in so-called ordinal proof theory – quickly become mathematically rather more sophisticated. Still, at this stage it is at least worth making a first pass at (C) Gentzen's proof of the consistency of arithmetic using a cut-elimination proof which invokes induction over some small countable ordinals. So these three themes from elementary proof theory will be the focus of this chapter.

9.1 Preamble: a very little about Hilbert's Programme

Set theory, for example, is about – or at least, is *supposed* to be about – an extraordinarily rich domain of (mostly) infinite objects. How can we know that such a theory really does make good sense? How can we know that it even gets to the starting line of being internally consistent?

David Hilbert had a wonderful insight. While the *topic* of a mathematical theory T such as set theory might be wildly infinitary, the *theory T itself* is built from thoroughly finite objects – namely sentences, and the finite arrays of sentences that are proofs. So perhaps we can use some very tame assumptions (assumptions that don't tangle with the infinite) to reason about T when it is thought of as a suite of finite objects. And in particular, perhaps we can use tame assumptions to prove T's internal consistency, without needing to worry about T's purported infinitary subject matter.

To make any progress with this idea, we'll need to fully pin down T's basic assumptions and to regiment the principles of reasoning that T can deploy – we'll need, in other words, to have a nice axiomatic formalization of T on the table. This formalization of the theory T (whether it's about sets, widgets, or whatnots) then gives us some definite, mathematically precise, *new* objects to reason about (beyond the sets, widgets, or whatnots), namely the T-wffs and T-proofs that make up the theory. And now, as Hilbert saw, we can set off to mathematically investigate *these*, developing a *Beweistheorie* (a theory about proofs).

We'll return in §9.3 to say something more about the resulting Programme of aiming to use entirely 'safe', merely finitary, reasoning about a theory T in order to prove its consistency (though you should already know that Gödel's Second Incompleteness Theorem is going to cause some trouble). But, for the moment, the point we want is simply this: the Programme presupposes that we can indeed regiment the theory that concerns us into a tidily disciplined formal shape – and in particular, we can regiment its required principles of reasoning into a formal deductive logic. Hence the central importance for Hilbert and his associates of constructing suitable formal systems for logic.

9.2 Deductive systems, normal forms, and cuts

(a) The logical systems developed by Hilbert and Bernays[1] were axiomatic in style, and at some remove from the forms of deduction used in practice in mathematical proofs. It was Bernays' student Gerhard Gentzen who first introduced a style of deductive system which explicitly aimed to come, as he put it, "as close as possible to actual reasoning." The result was Gentzen's natural deduction calculi for intuitionistic and classical predicate logic.

Now, these calculi – which I'll take to be familiar from work on earlier topics in this Guide – have some lovely features: and as advertised, they do allow us to formally track natural lines of reasoning. But they also still allow us to construct some perversely unnatural proofs! For example, consider the following two derivations to show that from $\mathsf{P} \wedge \mathsf{Q}$ we can infer $\mathsf{P} \vee \mathsf{Q}$:

$$\text{(i)}\quad \dfrac{\dfrac{\mathsf{P} \wedge \mathsf{Q}}{\mathsf{P}}}{\mathsf{P} \vee \mathsf{Q}} \qquad\qquad \text{(ii)}\quad \dfrac{\dfrac{\dfrac{\mathsf{P} \wedge \mathsf{Q}}{\mathsf{P}}}{\mathsf{P} \vee (\mathsf{R} \wedge \mathsf{Q})} \qquad \dfrac{[\mathsf{P}]^{(1)}}{\mathsf{P} \vee \mathsf{Q}} \qquad \dfrac{\dfrac{[\mathsf{R} \wedge \mathsf{Q}]^{(1)}}{\mathsf{Q}}}{\mathsf{P} \vee \mathsf{Q}}}{\mathsf{P} \vee \mathsf{Q}}\,(1)$$

(i) is an entirely natural mini-proof. But (ii) takes us on a pointless detour: on the leftmost branch, the 'wrong' disjunction is introduced which involves the quite irrelevant R, before we finally use a disjunction-elimination inference at (1) to finally get the proof back on track.

The detour in (ii) is not just inelegant; there is also a sense in which it makes the proof non-explanatory. After all, if a premiss A logically entails a conclusion C, this – we suppose – results from the conceptual content of A and C. So we want a proof to explain how the contents of A and C generate the entailment. A derivation like (ii), which introduces irrelevant content that is quite unrelated to either the premiss or conclusion, can't do that.

So, generalizing on the example of (ii), let's now define a detour as consisting in the use of the introduction rule for a logical operator (a connective or a

[1] Paul Bernays was nominally Hilbert's assistant, but in fact was an absolutely key figure in his own right, shaping Hilbert's writings on logic.

quantifier) followed by the application of the corresponding elimination rule to this introduced operator. Then, as just noted, it is not merely to avoid inelegancies that we will want detour-free proofs.

Now, simple detours in a Gentzen-style natural deduction proof can easily be removed. For example, a detour which involves introducing a conditional (by conditional proof) and then eliminating it (by modus ponens), as on the left, can be simply smoothed away or *reduced*, as on the right:

$$\dfrac{\begin{array}{c}\vdots\\ A\end{array} \qquad \dfrac{\begin{array}{c}[A]^{(1)}\\ \wr\\ B\end{array}}{A \to B}\,(1)}{B} \qquad \rightsquigarrow \qquad \begin{array}{c}\vdots\\ A\\ \wr\\ B\end{array}$$

For another example, going back to the case of introducing and then eliminating a disjunction, a proof of the shape on the left can be reduced to a proof with the shape on the right:

$$\dfrac{\dfrac{\begin{array}{c}\vdots\\ A\end{array}}{A \vee B} \qquad \begin{array}{c}[A]^{(1)}\\ \wr\\ C\end{array} \qquad \begin{array}{c}[B]^{(1)}\\ \wr\wr\\ C\end{array}}{C}\,(1) \qquad \rightsquigarrow \qquad \begin{array}{c}\vdots\\ A\\ \wr\\ C\end{array}$$

And similarly for other simple detours involving other connectives and the quantifiers. However, what about the case where a detour gets entangled with the application of other rules in more complicated ways? Can detours *always* be removed?

Gentzen was able to show that – at least for his system of intuitionistic logic – if a conclusion can be derived from premisses at all, then there will in fact be a *normal*, i.e. detour-free, proof of the conclusion from the premisses. And he did this by giving a *normalization* procedure – i.e. instructions for systematically removing detours until we are left with a normal proof. The resulting detour-free proofs will then have particularly nice features such as the so-called subformula property: every formula that occurs in a proof will either be a subformula of one of the premisses or a subformula of the conclusion (as usual, counting instances of quantified wffs as subformulas of them). There won't be irrelevancies as in our silly proof (ii) above.

And now note that, as a corollary, we can immediately conclude that intuitionistic logic is consistent: we can't have a proof with the subformula property from no premisses to $\bot$. Which raises a hopeful prospect: can other normalization proofs be used to establish the sort of consistency results that Hilbert wanted?

(b) But now the story gets complicated. For a start, Gentzen himself couldn't find a normalization proof for his natural deduction system of classical logic (you can see why there might be a problem – a classical proof might need e.g. to go via an instance of excluded middle which isn't a subformula of either the premisses or the conclusion). In order to get a classical system for which he *could* prove an appropriate normalization theorem, Gentzen therefore introduced his sequent calculi, about which more in moment. And his normalization proof for intuitionistic logic then remained unpublished for seventy years. In the meantime, the proof was independently rediscovered by Dag Prawitz in his thesis, published as *Natural Deduction* (1965), which also presents a normalization proof for Gentzen's classical natural deduction system without $\lor$ and $\exists$ (which is of course equivalent to the complete system).

Since Prawitz's work brought Gentzen-style natural deduction back to centre stage, there has been a whole cottage industry of tinkering with the inference rules, and tinkering with the definition of a normal proof, in order to produce classical natural deduction systems with nice proof-theoretic features. But I rather think that the typical beginner in mathematical logic won't find the details of *these* further developments particularly exciting. However, it *is* well worth looking at the opening four chapters of Prawitz's wonderful short book, and perhaps noting a few more ideas. This will be enough on our theme (A), natural deduction and normalization.

(c) How do we tell what depends on what in a natural deduction proof? By looking at the geometry of the proof, and its annotations.

For example, consider this derivation of $\mathsf{P} \to (\mathsf{Q} \to \mathsf{R})$ from $(\mathsf{P} \land \mathsf{Q}) \to \mathsf{R}$:

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

Reading upwards from e.g. R, we see that this wff depends on all three of $(\mathsf{P} \land \mathsf{Q}) \to \mathsf{R}$, P, and Q as assumptions (for neither of the last two have yet been discharged). While $\mathsf{Q} \to \mathsf{R}$ on the next line depends only on $(\mathsf{P} \land \mathsf{Q}) \to \mathsf{R}$ and P.

That's clear enough. But we could alternatively record dependencies quite explicitly, line by line. To do this, we will make use of so-called *sequents*. We'll write a sequent in the form $\Gamma \Rightarrow A$, and read this as saying that A is deducible from the finitely many (perhaps zero) wffs Γ.[2] Since an (undischarged) assumption depends just on itself, we can then explicitly record the deducibilities revealed in our last natural deduction proof as follows (check this claim!):

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

[2]For present purposes, we can alternatively think of Γ as given as a set – though in the end we might prefer to treat Γ as a multi-set where repetitions matter: Gentzen himself treated Γ as an ordered sequence.

And now, following Gentzen, instead of thinking of this tree of sequents as in effect just a running commentary on an underlying natural deduction proof, we can treat it as *itself* a new sort of proof in its own right – a proof relating whole sequents rather than individual wffs.

At the tips of branches of this sequent proof about deducibilities we have 'axioms' of the form $A \Rightarrow A$ (since, trivially, A is deducible given A!). And then the proof is extended downwards by the application of two sorts of rules, rules governing specific logical operators, and general structural rules.

For the logical rules, we could replace the familiar natural deduction rules for wffs with corresponding rules for deriving sequents, as in these examples:[3]

$$\frac{A \qquad B}{A \wedge B} \quad \rightsquigarrow \quad \frac{\Gamma \Rightarrow A \qquad \Delta \Rightarrow B}{\Gamma, \Delta \Rightarrow A \wedge B}$$

$$\frac{\begin{array}{c}[A]\\ \vdots \\ B\end{array}}{A \to B} \quad \rightsquigarrow \quad \frac{\Gamma, A \Rightarrow B}{\Gamma \Rightarrow A \to B}$$

There should be nothing mysterious here. After all, the terse schematic presentation of the natural-deduction introduction rule for $\wedge$ is to be read as saying that if we have A (deduced perhaps from some other assumptions) and have B (again perhaps deduced from some other assumptions), we can infer $A \wedge B$ (with those earlier assumptions all remaining in play). And that's what the suggested sequent calculus rule now explicitly says too. Likewise, the natural-deduction introduction rule for $\to$ is to be read as saying that if we derive B from the assumption A (and perhaps from some other assumptions), then we can drop that assumption A and infer $A \to B$ (with those other assumptions kept in play); and that's what the sequent calculus rule says too. There will be similar rules for other connectives and for quantifiers.

As for structural rules, we will mention here two candidates. The first is traditionally called *thinning* or *weakening* (neither of which is perhaps a very helpful label). The simple idea is that, if a wff is deducible from some assumptions, it remains deducible if we add in a further unnecessary assumption. So

$$\frac{\Gamma \Rightarrow C}{\Gamma, A \Rightarrow C}$$

Our second structural rule for sequent proofs corresponds to the structural fact that we can chain natural deduction proofs together into longer proofs. Thus in natural deduction,

$$\text{We can splice a proof } \begin{array}{c}\Gamma\\ \vdots\\ A\end{array} \text{ with a proof } \begin{array}{c}\underbrace{A \qquad \Delta}\\ \vdots\\ B\end{array} \text{ to get } \begin{array}{c}\Gamma\\ \vdots\\ \underbrace{A \qquad \Delta}\\ \vdots\\ B\end{array}.$$

[3] Obvious notation: If we are treating Γ as a set, then Γ, A is the set comprising the members of Γ plus A, while Γ, Δ is the union of the sets Γ and Δ.

In sequent calculus terms this corresponds to the following *cut* rule:

$$\frac{\Gamma \Rightarrow A \qquad \Delta, A \Rightarrow B}{\Gamma, \Delta \Rightarrow B}$$

This intuitively sound rule allows us to cut out the middle man A.

So far, then, so good – though of course, we've left lots of detail to be filled out. And as yet there is nothing really novel involved in reworking natural deduction into sequent style like this. But now, however, Gentzen introduces two *very* striking new ideas.

(d) To introduce the first idea, let's think again about the *elimination* rules for conjunction. As a first shot, we might expect to transform the pair of natural-deduction rules into a corresponding pair of sequent-calculus rules like this:

$$\frac{A \wedge B}{A} \qquad \frac{A \wedge B}{B} \quad \rightsquigarrow \quad \frac{\Gamma \Rightarrow A \wedge B}{\Gamma \Rightarrow A} \qquad \frac{\Gamma \Rightarrow A \wedge B}{\Gamma \Rightarrow B}$$

What could be more obvious? But in fact we could alternatively adopt the following sequent-calculus rule:

$$\frac{\Gamma, A, B \Rightarrow C}{\Gamma, A \wedge B \Rightarrow C}$$

This is obviously valid – if C can be derived from some assumptions Γ plus A and B, it can obviously be derived from Γ plus the conjunction of A and B. And we can use *this* rule introducing $\wedge$ on the *left* of the sequent sign instead of the expected pair of rules eliminating $\wedge$ to the *right* of the sequent sign. For note, given the new rule, we can restore the first of the elimination rules as a derived rule, because we can always give a derivation of this shape:

$$\dfrac{\Gamma \Rightarrow A \wedge B \qquad \dfrac{\dfrac{A \Rightarrow A}{A, B \Rightarrow A}\text{(Weakening)}}{A \wedge B \Rightarrow A}\text{(New rule for }\wedge)}{\Gamma \Rightarrow A}\text{(Cut)}$$

Similarly, of course, for the companion elimination rule.

And the point generalizes. As Gentzen saw, in a sequent calculus for intuitionistic logic, we can get *all* the rules for handling connectives and quantifiers to *introduce* a logical operator – either on the right of the sequent sign (corresponding to a natural-deduction introduction rule) or on the left of the sequent sign (corresponding to a natural-deduction elimination rule).

(e) We can go further. Still working with a sequent calculus for $\Rightarrow$ read as intuitionistic deducibility, we can in fact *eliminate* the cut rule. Anything provable using cut can be proved without it.

This might initially seem pretty surprising. After all, didn't we just have to appeal to the cut rule to show that – using our new introduction-on-the-left rule for $\wedge$ – we can still argue from (1) $\Gamma \Rightarrow A \wedge B$ to (2) $\Gamma \Rightarrow A$? How can we possibly do without cut in this case?

Well, consider how we might actually have arrived at (1) $\Gamma \Rightarrow A \wedge B$. Perhaps it was by the rule for introducing occurrences of $\wedge$ on the right of a sequent. So perhaps, to expose more of the proof from (1) to (2), it has the shape of the left-hand proof below (supposing Γ to result from putting together Γ' and Γ''):

$$\dfrac{\dfrac{\Gamma' \;\Rightarrow\; A \qquad \Gamma'' \;\Rightarrow\; B}{\Gamma \;\Rightarrow\; A \wedge B} \qquad \dfrac{\dfrac{A \;\Rightarrow\; A}{A, B \;\Rightarrow\; A}}{A \wedge B \;\Rightarrow\; A}}{\Gamma \;\Rightarrow\; A}\,{\scriptstyle (\mathrm{Cut})} \qquad \rightsquigarrow \qquad \dfrac{\Gamma' \;\Rightarrow\; A}{\Gamma \;\Rightarrow\; A}\,{\scriptstyle (\mathrm{Weakenings})}$$

But if we *already* have $\Gamma' \Rightarrow A$, as in the proof on the left, then we don't need to go round the houses on that detour, introducing an occurrence of $\wedge$ to get the formula $A \wedge B$, and then cutting out that same formula. We can just get from $\Gamma' \Rightarrow A$ to $\Gamma \Rightarrow A$ by some weakenings (by adding in the wffs from Γ''), as in the proof on the right. Here, then, eliminating the cut is just like normalizing (part of) a natural deduction proof.

OK: that only shows that in just *one* rather special sort of case, we can eliminate a cut. Still, it's a hopeful start! And in fact, we can *always* eventually eliminate cuts from an intuitionistic sequent calculus proof.

But the process can be intricate. For example, take a slight variant of our previous example and suppose we want to eliminate the following cut (remember, combining Γ and Γ gives us Γ!):

$$\dfrac{\dfrac{\Gamma \;\Rightarrow\; A \qquad \Gamma \;\Rightarrow\; B}{\Gamma \;\Rightarrow\; A \wedge B} \qquad \dfrac{\Delta, A, B \;\Rightarrow\; C}{\Delta, A \wedge B \;\Rightarrow\; C}}{\Gamma, \Delta \;\Rightarrow\; C}\,{\scriptstyle (\mathrm{Cut})}$$

Then we can replace this proof-segment with the following:

$$\dfrac{\Gamma \;\Rightarrow\; A \qquad \dfrac{\Gamma \;\Rightarrow\; B \qquad \Delta, A, B \;\Rightarrow\; C}{\Gamma, \Delta, A \;\Rightarrow\; C}\,{\scriptstyle (\mathrm{Cut})}}{\Gamma, \Delta \;\Rightarrow\; C}\,{\scriptstyle (\mathrm{Cut})}$$

Again, as in normalizing a natural deduction proof, we have removed a detour – this time a detour through introducing-$\wedge$-on-the-right and introducing-$\wedge$-on-the-left. So we have now lost the cut on the more complex formula $A \wedge B$, albeit replacing it with two new cuts. But still, these new cuts are on the simpler formulas A and B respectively, and we have also pushed one of the cuts higher up the proof. And that's typical: looking at the range of possible situations where we can apply the cut rule – a decidedly tedious hack though all the cases – we find we can keep reducing the complexity of formulas in cuts and/or pushing cuts up the proof until all the cuts are completely eliminated.

(f) So we arrive at this result. In a sequent-calculus setting, we can use a *cut-free* deductive system for intuitionistic logic where all the rules for the connectives and quantifiers *introduce* logical operators, either to the left or to the right of the sequent sign. Analogously to a normalized natural-deduction proof, there are no detours. As we go down a branch of the proof, the sequents at each stage are steadily more complex (we can make the relevant notion of complexity precise in pretty obvious ways).

This proof-analysis immediately delivers some very nice results.

(i) The subformula property: every formula occurring in the derivation of a sequent $\Gamma \Rightarrow C$ is a subformula of either one of formulas Γ or of C. (By inspection of the rules!)

(ii) There evidently can be no cut-free, ever-more-complex, derivation that ends with $\Rightarrow \bot$; in other words, absurdity isn't intuitionistically deducible from no premisses. Hence intuitionistic logic is internally consistent.

(iii) Equally evidently, the penultimate line of a cut-free, ever-more-complex, derivation of $\Rightarrow A \vee B$ has to be either $\Rightarrow A$ or $\Rightarrow B$, which establishes the disjunction property for intuitionistic logic – see §8.3(a).

Note too that, at least for propositional logic, we can take any sequent and systematically try to work upwards from it to construct a cut-free proof with ever-simpler-sequents: the resulting success or failure then mechanically decides whether the sequent is intuitionistically valid.

(g) I said that Gentzen had two very striking new ideas in developing his sequent calculi beyond a mere re-write of a natural deduction system in which dependencies are made explicit. The first idea was to recast all the rules for logical operators as rules for *introducing* logical operators, now allowing introduction to the left as well as introduction to the right of the sequent sign, and to then show that we can get a cut-free proof (hence, a proof that always goes from less complex to more complex sequents) for any intuitionistically correct sequent.

But this first idea doesn't by itself resolve the problem which Gentzen initially faced. Recall, he ran into trouble trying to find a normalization proof for *classical* natural deduction. And plainly, if we stick with a cut-free all-introduction-rules sequent calculus of the current style, we can't get a classical logical system at all. The point is trivial: one key additional classical principle we need to add to intuitionistic logic is the double negation rule. We need to be able to show, in other words, that from $\Gamma \Rightarrow \neg\neg A$ we can derive $\Gamma \Rightarrow A$. But obviously we can't do *that* in a system where we can only move from logically simpler to logically more complex sequents!

What to do? Well, at this point Gentzen's second (and quite original) idea comes into play. *We now liberalize the notion of a sequent.* Previously, we took a sequent $\Gamma \Rightarrow A$ to relate zero or more wffs on the left to a single wff on the right. Now we pluralize on *both* sides of the sequent sign, writing $\Gamma \Rightarrow \Delta$; and we read that as saying that at least one of Δ is deducible from the wffs Γ. If you like, you can regard Δ as delimiting the field within which the truth must lie if the premisses Γ are granted. (We'll continue, for our purposes, to treat Γ and Δ officially as sets, rather than multisets or lists: note that we will allow either or both to be empty.)

Keeping the idea that we want all our rules for the logical operators to be rules for *introducing* operators to the left or right of the sequent sign, how might these rules now go? There are various options, but the following can work nicely for conjunction and disjunction:

$$(\wedge\text{L})\ \frac{\Gamma, A, B \Rightarrow \Delta}{\Gamma, A \wedge B \Rightarrow \Delta} \qquad (\wedge\text{R})\ \frac{\Gamma \Rightarrow \Delta, A \qquad \Gamma \Rightarrow \Delta, B}{\Gamma \Rightarrow \Delta, A \wedge B}$$

$$(\vee\text{L})\ \frac{\Gamma, A \Rightarrow \Delta \qquad \Gamma, B \Rightarrow \Delta}{\Gamma, A \vee B \Rightarrow \Delta} \qquad (\vee\text{R})\ \frac{\Gamma \Rightarrow \Delta, A, B}{\Gamma \Rightarrow \Delta, A \vee B}$$

I won't give the rules for the conditional and the absurdity constant here. However, let's pause to note the left and right rules for negation (these can either be built-in rules, if negation is treated as a primitive built-in connective, or derived rules, if negation is defined in terms of the conditional and absurdity):

$$(\neg\text{L})\ \frac{\Gamma \Rightarrow \Delta, A}{\Gamma, \neg A \Rightarrow \Delta} \qquad (\neg\text{R})\ \frac{\Gamma, A \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg A}$$

These rules are evidently correct on the classical understanding of the connectives. For the first rule, suppose that given the assumptions Γ, then (at least) one from among Δ and A follows: then given the same assumptions Γ but now also ruling out A, we can conclude that (at least) one of Δ is true. We can argue similarly for the second rule. But with these negation and disjunction rules in place we immediately have the following derivation:

$$\cfrac{\cfrac{A \Rightarrow A}{\Rightarrow A, \neg A}\ (\neg\text{R})}{\Rightarrow A \vee \neg A}\ (\vee\text{R})$$

Out pops the law of excluded middle! – so we know we are dealing with classical calculus.

(h) What about the structural rules for our classical sequent calculus which allows multiple alternative conclusions as well as multiple premisses? We can now allow weakening on both sides of a sequent. And we can generalize the cut rule to take this form:

$$\frac{\Gamma \Rightarrow \Delta, A \qquad \Gamma', A \Rightarrow \Delta'}{\Gamma, \Gamma' \Rightarrow \Delta, \Delta'}$$

(Think why this is a sound rule, given our interpretation of the sequents!) But then, just as with our sequent calculus for intuitionistic logic, we can proceed to prove that we can eliminate cuts. If a sequent is derivable in our classical sequent calculus, it is derivable without using the cut rule.

And as with intuitionist logic, this immediately gives us some nice results. Of course, we won't have the disjunction property (think excluded middle!). But we still have the subformula property in the form that if $\Gamma \Rightarrow \Delta$ is derivable, the every formula in the sequent proof is a subformula of one of Γ, Δ. And again, simply but crucially, $\Rightarrow \bot$ won't be derivable in the cut-free classical system, so it is consistent.

And that's perhaps enough by way of introduction to our theme (B), in which we begin to explore various elegant sequent calculi, prove cut-elimination theorems, and draw out their implications.

9.3 Proof theory and the consistency of arithmetic

Now for our third theme (C), Gentzen's famed proof of the consistency of arithmetic (more precisely, the consistency of first-order Peano Arithmetic). Recall, Hilbert's Programme is the project of using tame proof-theoretic reasoning to prove the consistency of mathematical theories: PA gives us a first test case.

(a) You might very well wonder whether there can be *any* illuminating and informative ways of proving PA to be consistent. After all, proving consistency by appealing to a *stronger* theory like ZFC set theory which in effect contains PA won't be a very helpful (for doubts about the consistency of PA will presumably just carry over to become doubts about the stronger theory). And you already know that Gödel's Second Incompleteness Theorem shows that it is impossible to prove PA's consistency by appealing to a *weaker* theory tame enough to be modelled inside PA (not even full PA can prove PA's consistency).

However, another possibility does remain open. It isn't ruled out that we can prove PA's consistency by appeal to an attractive theory which is weaker than PA in some respects but stronger in others. And *this* is what Gentzen aims to give us in his consistency proof for arithmetic.[4]

(b) Here then is an outline sketch of the key proof idea, in Gentzen's own words.

We start with a formulation of PA using for its logic a classical sequent calculus including the cut rule. (We will initially want the cut rule in making use of PA's axioms, and we can't assume straight off the bat that we can still eliminate cuts once we have more complex proofs appealing to non-logical axioms). Then,

> The 'correctness' of a proof depends on the correctness of certain other simpler proofs contained in it as special cases or constituent parts. This fact motivates the arrangement of proofs in linear *order* in such a way that those proofs on whose correctness the correctness of another proof depends precede the latter proof in the sequence. This arrangement of the proofs is brought about by correlating with each proof a certain transfinite ordinal number.

The idea, then, is that the various sequent proof-trees in this version of PA can be put into an ordering by a kind of dependency relation, with more complex proof trees (on a suitable measure of complexity) coming after simpler proofs. And this can be a well-ordering, so that the position along the ordering can indeed be tallied by an ordinal number: see §7.2(b).

But why is the relevant linear ordering of proofs said to be *transfinite* (in other words, why must it allow an item in the ordering to have an infinite number of predecessors)? Because

---

[4] Gentzen in fact gives four different proofs, developed along somewhat different lines. But the master idea underlying the best known of the proofs is given in a wonderfully clear way in his wide-ranging lecture on 'The concept of infinity in mathematics' reprinted in his *Collected Papers*, from which the following quotations come.

> [it] may happen that the correctness of a proof depends on the correctness of infinitely many simpler proofs. An example: Suppose that in the proof a proposition is proved for *all* natural numbers by complete induction. In that case the correctness of the proof obviously depends on the correctness of every single one of the infinitely many individual proofs obtained by specializing to a particular natural number. Here a natural number is insufficient as an ordinal number for the proof, since each natural number is preceded by only finitely many other numbers in the natural ordering. We therefore need the transfinite ordinal numbers in order to represent the natural ordering of the proofs according to their complexity.

Think of it this way: a proof by induction of the quantified $\forall \mathsf{x} A(\mathsf{x})$ leaps beyond all the proofs of $A(\mathsf{0})$, $A(\mathsf{1})$, $A(\mathsf{2})$, ... And the result $\forall \mathsf{x} A(\mathsf{x})$ depends for its correctness on the correctness of the simpler results. So, in the sort of ordering of proofs which Gentzen has in mind, the proof by induction of $\forall \mathsf{x} A(\mathsf{x})$ must come infinitely far down the list, after all the proofs of the various $A(\mathsf{n})$.

And now Gentzen's key step is to argue by an induction along this transfinite ordering of proofs. The very simplest proofs right at the beginning of the ordering transparently can't lead to contradiction. Then

> once the correctness [and specifically, freedom from contradiction] of all proofs preceding a particular proof in the sequence has been established, the proof in question is also correct precisely because the ordering was chosen in such a way that the correctness of a proof depends on the correctness of certain earlier proofs. From this we can now obviously infer the correctness of all proofs by means of a transfinite induction, and we have thus proved, in particular, the desired consistency.

Transfinite induction, recall, is just the principle that, if we can show that a proof has a property P if all its predecessors in the relevant transfinite ordering have P, then *all* proofs in the ordering have property P.

(c) We can implement this same proof idea the other way around. We show that if any proof *does* lead to contradiction, then there must be an *earlier* proof in the linear ordering of proofs which also leads to contradiction – so we get an infinite sequences of proofs of contradiction, ever earlier in the ordering. But then the ordinals which tally these proofs of contradiction would have to form an infinite descending sequence. And there can't be such a sequence of ordinals, since the ordinals are well-ordered. Hence no proof leads to contradiction and PA is consistent.

(d) Two questions arising. First, *how* do we show that if a proof leads to a contradiction, then there must be another proof *earlier* in the linear ordering which also leads to contradiction? By eliminating cuts using reduction procedures like those involved in the proof of cut-elimination for a pure logical sequent calculus – so here's the key point of contact with ideas we meet in tackling theme (B).

And second, what kind of transfinite ordering is involved here? Gentzen's ordering of possible proof-trees in his sequent calculus for PA turns out to have the order type of the ordinals less than ε_0 (what does that mean? – the references will explain, but these are all the ordinals which are sums of powers of ω). So, what Gentzen's proof needs is the assumption that a relatively modest amount of transfinite induction – induction up to ε_0 – is legitimate.

Now, the PA proof-trees which we are ordering are themselves all finite objects; we can code them up using Gödel numbers in the familiar sort of way. So in ordering the proofs, we are in effect thinking about a whacky ordering of (ordinary, finite) code numbers. And whether one number precedes another in the whacky ordering is nothing mysterious; a computation without open-ended searches can settle the matter.

So what resources does a Gentzen-style argument use, if we want to code it up and formalize it? The assignment of a place in the ordering to a proof can be handled by primitive recursive functions, and facts about the dependency relations between proofs at different points in the ordering can be handled by primitive recursive functions too. A theory in which we can run a formalized version of Gentzen's proof will therefore be one in which we can (a) handle primitive recursive functions *and* (b) handle transfinite induction up to ε_0, maybe via coding tricks. It turns out to be enough to have all p.r. functions available, together with a formal version of transfinite induction just for simple *quantifier-free* wffs containing expressions for these p.r. functions. Such a theory is neither contained in PA (since it can prove PA's consistency by formalizing Gentzen's method, which PA can't), nor does it contain PA (since it needn't be able to prove instances of the ordinary Induction Schema for *arbitrarily complex* wffs).

So, in this sense, we can indeed prove the consistency of PA by using a theory which is weaker than PA in some respects while stronger in others.

(e) Of course, it is a very moot point whether – if you were *really* worried about the consistency of PA – a Gentzen-style proof when fully spelt out would help resolve your doubts. Are the resources the proof invokes 'tame' enough to satisfy you?

Well, if you are globally worried about the use of induction in general, then appealing to an argument which deploys an induction principle won't help! But global worries about induction are difficult to motivate, and perhaps your worry is more specifically that induction over arbitrarily complex wffs might engender trouble. You note that PA's induction principle applies, inter alia, to wffs that themselves quantify over all numbers. And you might worry that if (like Frege) you understand the natural numbers to be what induction applies to, then there's a looming circularity here – numbers are understood as what induction applies to, but understanding some cases of induction involves understanding quantifying over numbers. If *that* is your worry, the fact that we can show that PA is consistent using an induction principle which is only applied to quantifier-free wffs (even though the induction runs over a novel ordering on the numbers) could soothe your worries.

Be that as it may: we can't pursue that kind of philosophical discussion any further here. The point remains that the Gentzen proof is a fascinating achievement, containing the seeds of wonderful modern work in proof theory. Perhaps we haven't quite executed an instance Hilbert's Programme, proving PA's consistency by appeal to entirely tame proof-theoretic reasoning. But in the attempt, we have found how far along the ordinals we need to run our transfinite induction in order to prove the consistency of PA.[5] And we can now set out to discover how much transfinite induction is required to prove the consistency of other theories. But the achievements of that kind of ordinal proof theory will have to be left for you (eventually) to explore ...

9.4 Main recommendations on elementary proof theory

Let's start with a couple of very useful encyclopaedia entries by some notable proof theorists.

First, the following exemplary historical outline is particularly helpful for orientation:

1. Jan von Plato, 'The development of proof theory', *The Stanford Encyclopedia of Philosophy*. Available at tinyurl.com/sep-devproof.

And then look at the first half of the main entry on proof theory:

2. Michael Rathjen and Wilfrid Sieg, 'Proof theory', §§1–3, *The Stanford Encyclopedia of Philosophy*. Available at tinyurl.com/sep-prooftheory.

Skip over any passages that are initially unclear, and return to them when you've worked through some of the readings below.

In keeping with our overviews in the previous two sections, I suggest that – in a first encounter with proof theory – you focus on (A) normalization for natural deduction and its implications; (B) the sequent calculus, cut-elimination and its implications; and (C) a Gentzen-style proof of the consistency of arithmetic. Now, there is book which aims to cover just these topics at the level we want:

3. Paolo Mancosu, Sergio Galvan and Richard Zach, *An Introduction to Proof Theory: Normalization, Cut-Elimination and Consistency Proofs* (OUP, 2021) – henceforth *IPL*.

However, as the authors say in their Preface, "in order to make the content accessible to readers without much mathematical background, we carry out the details of proofs in much more detail than is usually done." And the result isn't anywhere near as reader-friendly as they intend: expositions too often become

[5] Technical remark. There are no worries about using transfinite induction up to any ordinal less than ε_0; for *this* can be handled inside PA. So Gentzen's proof calls on the least possible extension to the amount of induction that can be handled inside PA!

wearyingly laborious. Also the authors stick very closely to Gentzen's own original papers, which isn't always the wisest choice. So, at least on topic areas (A) and (B), I will be highlighting some alternatives.

(A) You could find that the following *Handbook of the History of Logic* article gives some more helpful orientation:

4. F. J. Pelletier and Allen Hazen, 'Natural deduction', §3. Available at tinyurl.com/pellhazen.

   It is §3.1 that is most immediately relevant. But do read the rest of §3. (And, for your general logical education, why not read all this informative survey paper sometime?)

You could next tackle Chs 3 and 4 of *IPL*. But there's a lot to be said for just diving into the brisk opening chapters of a modern classic:

5. Dag Prawitz, *Natural Deduction: A Proof-Theoretic Study** (originally published 1965, reprinted by Dover Publications 2006), Chapters I to IV.

   Ch. I presents the now-standard Gentzen-style natural deduction systems for intuitionistic and classical logic. The short Ch. II explains the sense in which elimination rules are inverses to introduction rules. Then it notes some basic "reduction steps" for eliminating the sort of unnecessary detours which result from the application of an introduction rule being immediately followed by the application of the corresponding elimination rule. Ch. III shows that we can normalize proofs in a classical ND system – or at least, a cut down version without $\vee$ and $\exists$ built in as primitive – by systematically eliminating detours. Ch. IV extends the result to a full system of intuitionistic logic.

And that's perhaps about as much as you need on natural deduction. OK, you might be left wondering whether we can improve on Prawitz's Chapter III result and prove a similar normalization result for a full classical logic with the $\vee$ and $\exists$ rules restored. The answer is 'yes'. *IPL* §4.9 shows how it can be done for Gentzen's original natural deduction system. But it is more interesting to look at what happens if you revise Gentzen's original classical rules and use so-called 'general elimination rules'; this makes establishing normalization rather more straightforward. For something on this, see

6. Jan von Plato, *Elements of Logical Reasoning* (CUP, 2013). Chapters 3 to 6.

These very accessible chapters on intuitionistic and classical propositional logic also introduce the theme of proof-search.

Von Plato's book is, in fact, intended as a first introductory logic text, based on natural deduction: but it, very unusually, has a strongly proof-theoretic emphasis. And non-mathematicians, in particular, could find the whole book very helpful.

(B) Next, moving on to sequent calculi, you could start with Chs 5 and 6 of *IPL*. But the following is very accessibly written, ranges more widely, and is likely to prove quite a bit more enjoyable:

7. Sara Negri and Jan von Plato, *Structural Proof Theory* (CUP, 2001).
   The first four chapters gives us the basics. Ch. 1 helpfully bridges our topics, 'From natural deduction to sequent calculus'. Ch. 2 gives a sequent calculus for intuitionistic propositional logic and proves the admissibility of cut. Ch. 3 does the same for classical propositional logic. Ch. 4 adds the quantifiers.
   You might well want to then read on to Ch. 5 which illuminatingly discusses some variant sequent calculi. Then you can jump to Ch. 8 which takes us 'Back to natural deduction'. This relates the sequent calculus to natural deduction with general elimination rules, shows how to translate between the two styles of logic, and then derives a normalization theorem from the cut-elimination theorem: again this is very instructive.

Negri and von Plato note that, as we 'permute cuts upward' in a derivation – in order to eventually arrive at a cut-free proof – the number of cuts remaining in a proof can increase exponentially as we go along (though the process eventually terminates). So a cut-free proof can be much bigger than its original version. Pelletier and Hazen (4) in their §3.8 make some interesting related comments about sizes of proofs. And you will certainly want to read this famous short paper:

8. George Boolos, 'Don't eliminate cut', reprinted in his *Logic, Logic, and Logic* (Harvard UP, 1998).

And *now*, if you really want to know more (in particular about how Gentzen originally arrived at his cut-elimination proof) you can make use of the relevant *IPL* chapters, skipping over a lot of the tedious proof-details.

(C) Next, on Gentzen's proof of the consistency of arithmetic. In their *SEP* articles, von Plato and Rathjen/Sieg both provide some context for Gentzen's work. And here's a contemporary mathematician's perspective on why we might be interested in the proofs of the consistency of PA:

9. Timothy Y. Chow, 'The consistency of arithmetic', *The Mathematical Intelligencer* 41 (2019), 22–30. Available at tinyurl.com/chow-cons.

Now we have two options, as Rathjen/Sieg makes clear. We can tackle something like one of Gentzen's own consistency proofs for PA; but we then have to tangle with a *lot* of messy detail as we negotiate the complications caused by having to deal with the induction axioms. Or alternatively we can box more cleverly, and prove consistency for a theory PA_ω which swaps the induction axioms for an infinitary rule. The proof uses the same overall strategy, but this time its implementation is a lot less tangled (yet the proof still does the needed job, since PA_ω's consistency implies PA's consistency).

There are a number of versions of the second line of proof in the literature. There is quite a neat but rather terse version here, from which you should be able to get the general idea (it assumes you know a bit about ordinals):

10. Elliott Mendelson, *Introduction to Mathematical Logic*, 'Appendix: A consistency proof for formal number theory' (1st edn., 1964; later dropped but restored in the 6th edn., 2015).

But let's suppose that you do want something much closer to Gentzen's original proof:

There is a rather austere presentation of a Gentzen-style proof in the classic textbook on proof theory by Takeuti which I will mention in the next section: this might suit the more mathematical reader. But the following is more accessible – though with a distracting amount of detail:

3. Mancosu, Galvan and Zach, *IPL*. Read Chapter 8 on ordinal notations first. Then the main line of proof is in Chapters 7 and 9.

Now, after an initial dozen pages saying something about PA, these Chs 7 and 9 together span another *sixty-five* pages(!), and it is consequently easy to get lost/bogged down in the details. And it is not as if the discussion is padded out by e.g. a philosophical discussion about the warrant for accepting the required amount of ordinal induction; the length comes from hacking through more details than any sensible reader will want or need.

However, if you have already tackled a modest amount of other mathematical logic, you should by now have enough nous to be able to read these chapters pausing over the key ideas and explanations while initially skipping/skimming over much of the detail. You could then quite quickly and painlessly end up with a very good understanding of at least the general structure of Gentzen's proof and of what it is going to take to elaborate it. So I suggest first skimming through to get the headline ideas, and then do a second pass to get more feel for the shape of some of the details. You can then drill down further again to work through as much of the remaining nitty-gritty that you then feel that you really want/need (which probably won't be much!).

9.5 Some parallel/additional reading

Here I will start by mentioning (parts of) three other books. Each of them starts again from scratch, but then their varied modes of presentation are perhaps half a step up in mathematical sophistication from the readings in the last section;

11. Gaisi Takeuti, *Proof Theory** (North-Holland 1975, 2nd edn. 1987: reprinted Dover Publications 2013).

    This is a true classic – if only because for a while it was about the only available book on most of its topics. Later chapters won't really be accessible to beginners. But you can certainly tackle Ch. 1 on logic, §§1–7 (and perhaps the beginnings of §8, pp. 40–45, which is easier than it looks if you compare how you prove the completeness of a tableau system of logic). Then tackle Ch. 2, §9 on Peano Arithmetic. You can skip the next section on the incompleteness theorem, and skim §11 on ordinals (which makes heavy weather of what's really needed, which is the claim that a decreasing series of ordinals less than ε_0 can only be finitely long: see p. 98 on). The core consistency proof is then given in §12; read up to at least p. 114. This isn't exactly plain sailing – but if you skip and skim over some of the more tedious proof-details you should pick up a good sense of what happens in the consistency proof.

12. Jean-Yves Girard, *Proof Theory and Logical Complexity. Vol. I* (Bibliopolis, 1987). With judicious skipping, which I'll signpost, this is readable and insightful.

    So: skip the 'Foreword', but do pause to glance over 'Background and Notations' as Girard's symbolic choices need a little explanation. Then the long Ch. 1 is by way of an introduction, proving Gödel's two incompleteness theorems and explaining 'The Fall of Hilbert's Program': if you've read some of the recommendations on arithmetic, you can probably skim this fairly quickly, though noting Girard's highlighting of the notion of 1-consistency.

    Ch. 2 is on the sequent calculus, proving Gentzen's *Hauptsatz*, i.e. the crucial cut-elimination theorem, and then deriving some first consequences (you can probably initially omit the forty pages of annexes to this chapter). Then also omit Ch. 3 whose content isn't relied on later. But Ch. 4 on 'Applications of the *Hauptsatz*' is crucial (again, however, at a first pass you can skip almost 60 pages of annexes to the chapter). Take the story up again with the first two sections of Ch. 6, and then tackle the opening sections of Ch. 7. A rather bumpy ride but very illuminating.

13. A. S. Troelstra and H. Schwichtenberg, *Basic Proof Theory* (CUP 2nd ed. 2000). You can, with a bit of skipping, at this stage usefully read Chs 1–3, the first halves of Chs 4 and 6, and then Ch. 10 on arithmetic again.

The last is a volume in the series 'Cambridge Tracts in Computer Science'. Now, one theme that runs through the book concerns the computer-science idea of formulas-as-types and invokes the lambda calculus: however, it is in fact quite possible to skip over those episodes if (as is probably the case) you aren't yet familiar with the idea. The book, as the title indicates, is intended as a first foray into proof theory, and it *is* reasonably approachable. However it does spend quite a bit of time looking at slightly different ways of doing natural deduction and slightly different ways of doing the sequent calculus, and the differences may matter more for computer scientists with implementation concerns than for others.

Let me add two more recommendations. First, a book that sits rather askew to the mainstream texts I've mentioned so far:

14. Neil Tennant, *Core Logic* (OUP, 2017). This accessible tour-de-force is very well worth reading for its interesting proof-theoretic insights, even if at the end of the day you don't want to buy the relevantist aspects which we'll say more about in §11.1.

And second, let's go back to the beginning of this chapter and find out more about Hilbert's Programme. There is an excellent *SEP* article by Richard Zach. But there's an expanded version here:

15. Richard Zach, 'Hilbert's Programme Then and Now' in D. Jacquette, ed., *Philosophy of Logic: Handbook of the Philosophy of Science, Vol 5* (North-Holland 2007), available at tinyurl.com/zach-hil. This both reviews the history and has intriguing pointers forward.

We will return to consider more advanced texts on proof theory in the final chapter, §12.6.

10 Modal logics

A deduction, Aristotle tells us, requires a conclusion which 'comes about by necessity' given some premisses. So it is no surprise that, from the very beginning, logicians have been interested in the modal notions of necessity and possibility. Modern modal logics aim, at least in the first place, to regiment reasoning about such notions. But as we will see, they can be applied much more widely.

Here's an attractive thought: *it is necessarily true that* A just if A is not only true here in the actual world but also obtains in all relevant possible worlds. Suppose we add to a logical language a symbol $\Box$, where $\Box A$ is to be read as *it is necessarily true that* A. Then, to formally model our attractive thought, we will take some objects to represent possible worlds, and say that $\Box A$ is true at 'world' w in the model just if A is true at all 'worlds' w' suitably related to w.

Compare: in §8.3(c), we described a semantic model for intuitionistic logic with the following key feature – to determine whether the conditional $A \to B$ holds in a situation k in the model, it isn't enough to know whether A holds in k and whether B holds in k; we also need to know whether A and B obtain in *other* situations k' suitably related to k. So now the idea is to use a similar relational semantics for the necessity operator, with the truth of $\Box A$ in one situation w again depending on what happens in *other* related situations w'.

In §10.1, then, we explore this key idea by taking a look at some basic modal logics. These and similar logics will be of interest to quite a few philosophers and also eventually to some mathematicians and computer scientists who investigate relational structures. There is, however, one rather distinctive modal logic which should be of particular interest to *anyone* beginning mathematical logic, namely so-called provability logic: we will highlight that in §10.2. Provability logic can be tackled without a wider background in modal logic; but it certainly doesn't hurt to know a little about the wider picture we introduce first.

10.1 Some basic modal logics

(a) Notation first. As just proposed, we are going to add a one-place operator $\Box$ to our familiar logical languages (propositional, first-order), governed by the new syntactic rule that if A is a wff, so is $\Box A$.

Now, as we said, $\Box$ is typically going to be interpreted as some sort of necessity operator. We could also build into our languages a matching possibility operator

$\Diamond$ (so we read $\Diamond A$ as *it is possibly true that* A). But, to keep things simple, we won't do that, since $\Diamond A$ can equally well be treated as just a definitional abbreviation for $\neg\Box\neg A$. Reflect: it is possibly true that A iff A is true at some possible world, iff it isn't the case that A is false at all possible worlds, iff it isn't the case that $\neg A$ is necessary. So the parallel between the equivalences $\Diamond/\neg\Box\neg$ and $\exists w/\neg\forall w\neg$ is not an accident!

A third modal symbol you will come across is $⥽$, for what is standardly called 'strict implication'. But again, we can treat $A ⥽ B$ as a definitional abbreviation, this time for $\Box(A \to B)$.

Hence, following quite common practice, we will here take $\Box$ to be the sole built-in modal operator in our languages.

(b) The story of modern modal logic begins with C. I. Lewis's 1918 book *A Survey of Symbolic Logic*. Lewis presents postulates for $⥽$, motivated by claims about the proper understanding of the idea of implication, though unfortunately his claims do seem pretty muddled.[1] Later, in C. I. Lewis and C. H. Langford's 1932 *Symbolic Logic*, there are further developments: the authors distinguish five modal logics of increasing strength, which they label *S1* to *S5*. But why multiple logics?

Let's take four schemas, and ask whether we should accept all their instances when the $\Box$ is interpreted in terms of necessary truth:

K	$\Box(A \to B) \to (\Box A \to \Box B)$
T	$\Box A \to A$
S4	$\Box A \to \Box\Box A$
S5	$\neg\Box A \to \Box\neg\Box A$

Well, on any understanding of the idea of necessity, if $A \to B$ and A both hold necessarily, so does B: so we can accept the principle K. And necessary truth implies plain truth: so we can accept T too. But what about the principles S4 and S5 (which are in fact distinctive of Lewis and Langford's systems *S4* and *S5*)?

It seems that different principles about repeated modalities will be acceptable depending on how exactly we interpret the necessity involved. Take a couple of examples. Suppose we interpret $\Box A$ in a mathematical context as meaning that A necessarily holds in the sense that *it is provable that* A (i.e. is provable by ordinary informal standards of proof): then arguably (i) in this case, S4 but not S5 holds. Alternatively, suppose we interpret $\Box$ as indicating analyticity in the old-fashioned philosopher's sense (where it is analytically true that A if A is true just in virtue of its conceptual content): then arguably (ii) in this case, both the S4 and S5 principles hold. But I'm certainly not going to get into the business of assessing the supposed arguments for (i) and (ii) – the issues are far too murky. And that's exactly the point to make here: the early discussions of systems of modal logic, and the supposed semantic justifications for various

[1] The modern reader might well suspect confusion between ideas that we now demarcate by using the distinguishing notations $\to, \vdash$ and $\vDash$ (cf. §3.2(e)).

suggested principles, were entangled with contentious philosophical arguments. No wonder then that modal logic initially had a somewhat shady reputation!

(c) The picture radically changed some thirty years after Lewis and Langford, when Saul Kripke (in particular) developed a sharply characterized framework for giving semantic models for various modal logics.

Let's begin with the headline news about some modal *propositional* logics. In this subsection we'll describe a family of semantic models. In the next subsection we'll describe a family of deductive modal proof systems. Then the following subsection makes the Kripkean connections between the two.

So let's assume we are working in some suitable language L with the absurdity constant $\bot$ built in alongside the other usual propositional connectives, plus the unary operator $\Box$. And to define a relational semantics for such a language, we obviously need to start by introducing relational structures:

1. The basic ingredients we need are some objects W and a relation R defined over them. For the moment, think of W as a collection of 'possible worlds' and then wRw' will say that the world w' is possible relative to w (or if you like, w' is an accessible possible world, as seen from w).
2. And we will pick out an object w_0 from W to serve as the 'actual world'.

But we need an important further idea:

3. To get different flavours of relational structure (for interpreting different flavours of modal deductive system) we will want to specify different conditions S that the relation R needs to satisfy. For just one example, we might be particularly interested in relational structures where R is specified as being transitive and reflexive.

Let's say, for short, that a relational structure where the relation R satisfies the condition S is an S-structure.

Next we define the idea of a valuation of L-sentences on an S-structure. The story starts unexcitingly!

i. We initially assign a value, either *true* or *false*, to each propositional letter of L with respect to each world w. Then,
ii. The propositional connectives behave in the now entirely familiar classical ways. For example, $A \to B$ is true at w if and only if either A is false at w or B is true at w; and so forth.

The only real novelty, as trailed at the outset, is in the treatment of the modal operator $\Box$. We stipulate

iii. $\Box A$ is true at a world w if and only if A is true at every world that is possible relative to w, i.e. A is true at every world w' such that wRw'.

Evidently, given (ii) and (iii), every valuation ends up assigning a value to each L-wff A at each world.

Let's say that an S-structure together with such a valuation for L-sentences is an S-model for L. Then, continuing our list of definitions, when A is an L-sentence,

iv. A is (simply) true in a given S-model for L if and only if A takes the value *true* at the actual world w_0 in the model.

Finally, and predictably, we say

v. A is S-valid if and only if it is true in every S-model.

So that sets up the general framework for a relational semantics for a propositional modal language. But we are now going to be interested in four different particular versions got by filling out the specification S in different ways, and so giving us four different notions of validity for propositional modal wffs:

(K) K-validity is defined in terms of K-models which allow any relation R (the specification condition S is null).
(T) T-validity is defined in terms of T-models which require the relation R to be reflexive.
(S4) *S4*-validity is defined in terms of *S4*-models which require the relation R to be reflexive and transitive.
(S5) *S5*-validity is defined in terms of *S5*-models which require the relation R to be reflexive, transitive and symmetric (i.e. R has to be an equivalence relation).

As we will soon discover, the labels we have chosen are significant!

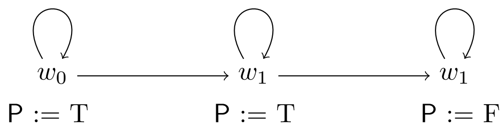
(d) Let's look at a couple of very instructive mini-examples. Take first the following two-world model, with an arrow $w \longrightarrow w'$ depicting that wRw', and with the values of P at each world as indicated:

Now, in this model, $\Box$P is true at w_0, since P is true at every world accessible from w_0, namely w_1. $\Box$P is also true at w_1, since P is again true at every world accessible from w_1, namely w_1 itself. And so $\Box\Box$P is true at w_0, since $\Box$P is true at every world accessible from w_0.

But note $\Box$P $\rightarrow$ P is *false* at w_0. So in a model like this one where the accessibility relation is not reflexive, not every instance of the schema T is true. Conversely, a moment's reflection shows that in T-models, which require that the accessibility relation is reflexive, instances of the schema T must always be true (because if $\Box A$ is true at w_0 then A is true at all accessible worlds, which will include w_0 by the reflexiveness of accessibility).

Moral: if $\Box$ is to be interpreted as necessary truth, where instances of the schema T should always come out true, then we'll want our semantic models to be built using a reflexive relation R.

For our second example, take this three-world model:



Note, this *is* not only a K model but also a T-model, because the diagrammed accessibility relation R is reflexive; but it is not an $S4$ model since R is not transitive (we have w_0Rw_1 and w_1Rw_2 but not w_0Rw_2).

Now, in this model, $\Box\mathsf{P}$ is true at w_0 (because P is true at both the accessible-from-w_0 worlds, i.e. at w_0 and w_1). But $\Box\mathsf{P}$ is false at w_1 (because P is false at the accessible-from-w_1 world w_2). And then since $\Box\mathsf{P}$ is false at w_1 and w_1 is accessible from w_0, it follows that $\Box\Box\mathsf{P}$ is false at w_0. And hence in this model $\Box\mathsf{P} \to \Box\Box\mathsf{P}$ is false (i.e. false at w_0). Moral: the $\mathsf{S4}$ principle can fail in models where the accessibility relation is not transitive.

But we can also show the reverse – in other words, in $S4$ models where the accessibility relation is transitive, the $\mathsf{S4}$ principle holds. That follows because $\mathsf{S4}$ can only fail in a model if the accessibility relation is non-transitive:

> Suppose something of the form $\Box A \to \Box\Box A$ *is* false in a given model, so (i) $\Box A$ is true at w_0 while (ii) $\Box\Box A$ is false at w_0. But for (ii) to hold, there must be a world w_1 such that w_0Rw_1 and (iii) $\Box A$ is false at w_1. And for (iii) to hold there must be a world w_2 such that w_1Rw_2 and (iv) A is false at w_2. But then (iv) w_2 must be 'invisible' from w_0, or else (i) couldn't hold: i.e. we can't have w_0Rw_2. In sum, for $\Box A \to \Box\Box A$ to fail we need three worlds such that w_0Rw_1, w_1Rw_2 but not w_0Rw_2 – which requires R to be non-transitive.

So our two mini-examples very nicely make the connection between a structural condition on models and the obtaining of a general modal principle such as T or $\mathsf{S4}$. More about this very shortly.

(e) Since our main concern here is with the formalities, we won't delve into the arguments about which specification conditions S appropriately reflect which intuitive notions of necessity (though note that even the condition T can fail if e.g. we want to model deontic necessities – i.e. necessities of duty: since what ought to be the case may not in fact be the case!). We can leave it to the philosophers to fight things out. For now, it might be more useful to pause to summarize our semantic story in the style of our earlier account of intuitionistic semantics in §8.3(c).

So, an S-structure is a triple (w_0, W, R) where W is a set, $w_0 \in W$, and R is a relation defined over W which satisfies the conditions S. Then an S-model for a modal propositional language L is an S-structure together with a valuation relation $\Vdash$ ('makes true') between members of W and wffs of L such that

(i) $w \nVdash \bot$.
(ii) $w \Vdash \neg A$ iff $w \nVdash A$.
(iii) $w \Vdash A \wedge B$ iff $w \Vdash A$ and $w \Vdash B$.
(iv) $w \Vdash A \vee B$ iff $w \Vdash A$ or $w \Vdash B$.
(v) $w \Vdash A \to B$ iff $w \nVdash A$ or $w \Vdash B$.
(vi) $w \Vdash \Box A$ iff, for any w' such that wRw', $w' \Vdash A$.

We say that A is true in a given S-model when $w_0 \Vdash A$. As before, A is S-valid when A is true in all S-models. And for the moment the most significant conditions S on the accessibility relation R in a model are K (null), T (reflexivity), *S4* (reflexivity and transitivity), *S5* (equivalence).[2]

(f) Now let's turn to consider some proof systems for propositional modal logics. And, just because it is simplest way to do things, let's give an old-school axiomatic presentation (leaving natural deduction and tableaux versions to be explained in the recommended reading). Here then are four key systems, starting with the simplest:

(K) The modal axiomatic system K is the theory whose axioms are

(Ax i) All instances of tautologies.
(Ax ii) All instances of the schema K.

And whose rules of inference are

(MP) From A and $A \to B$, infer B.
(Nec) If A is deducible as a theorem, infer $\Box A$.

To explain briefly: Read (Ax i) as meaning that, given a schema for a classical tautology, the result of systematically substituting *any* wffs of our modal propositional language for schematic letters – even substituting modalized wffs – will be an axiom of K. So, for example, $(A \wedge B) \to A$ is a schema for a classical tautology. Hence the result of substituting $\Box\mathsf{P}$ for A and $\Box\mathsf{Q}$ for B, giving us $(\Box\mathsf{P} \wedge \Box\mathsf{Q}) \to \Box\mathsf{P}$, is an axiom of K. Such instances of tautologies are still, surely, logical truths.

We've already said that instances of (Ax ii) look good on any suitable reading of the box. And our old friend the modus ponens rule (MP) is uncontentious.

[2]As in §8.3, fn. 1, I need to link up what I've just said with other presentations you'll encounter.

First, note that what I've called S-structures are more standardly called *frames*.

Second, and more importantly, note that – although Kripke's original presentation did involve, as here, picking out a 'world' w_0 from W to play the role of the 'actual' world – it is clear that we can drop that step and can equivalently re-define S-validity as truth at *all* worlds in any S-model.

(Why? Obviously, if A is valid on the revised definition it is valid on our original definition. While if A is not valid on the revised definition, A must be false at some world, and so it will be false on the Kripke model with that world chosen as the 'actual' world w_0.)

Which leaves the necessitation rule (Nec). This is to be very sharply distinguished from what would evidently be the quite unacceptable axiom schema $A \to \Box A$: obviously, A can be true without being necessarily true. However, the idea justifying (Nec) is that if A is actually a logical theorem – i.e. if A is deducible from logical principles alone – then it will indeed be necessary (on almost any sensible understanding of 'necessary'). Here's an example of the rule (Nec) in use in a K-proof:

1.	$((\mathsf{P} \land \mathsf{Q}) \to \mathsf{P})$	Axiom, by (Ax i)
2.	$\Box((\mathsf{P} \land \mathsf{Q}) \to \mathsf{P})$	By (Nec), since 1 is a theorem
3.	$\Box(((\mathsf{P} \land \mathsf{Q}) \to \mathsf{P}) \to (\Box(\mathsf{P} \land \mathsf{Q}) \to \Box\mathsf{P}))$	Axiom, by (Ax ii)
4.	$(\Box(\mathsf{P} \land \mathsf{Q}) \to \Box\mathsf{P})$	From 2 and 3 by (MP)
5.	$\Box(\Box(\mathsf{P} \land \mathsf{Q}) \to \Box\mathsf{P})$	By (Nec), since 4 is a theorem

In sum, then, all the theorems of the weak system K – i.e. all the wffs deducible from axioms alone – should be logical truths on (almost all) readings of $\Box$ read as a kind of necessity.

And now here are three nested ways of strengthening the system K:

(T) T is the axiomatic system K augmented with all instances of the schema T as axioms.
(S4) $S4$ is T augmented with all instances of the schema $\mathsf{S4}$ as axioms.
(S5) $S5$ is $S4$ augmented with all instances of the schema $\mathsf{S5}$ as axioms.

The readings will give lots of examples of these (or equivalent) proof systems in action.

(g) So now at last for the big reveal – except of course I've entirely sploit any element of surprise by the parallel labelling of the flavours of modal semantics and the flavours of axiomatic proof system!

What Kripke famously showed is the following lovely result:

> Whether S is K, T, $S4$, $S5$, a wff A is an S-theorem if and only if it is S-valid.

In short, we have soundness and completeness theorems for our proof systems. And there are some nice immediate implications. Searching for an appropriate countermodel which shows that a wff is not S-valid is a finite business, so it is decidable what's S-valid – and hence it is decidable what's an S-theorem.[3]

These soundness and completeness results are not mathematically very difficult. Perhaps Kripke's real achievement was the prior one in developing the general semantic framework and in finding the required simple proof systems –

[3]Suppose we define in the now obvious ways (i) the idea of a conclusion being an S-valid consequence of some finite number of premisses, and (ii) the idea of that conclusion being deducible in system S from those premisses. Then again we have soundness and weak completeness proofs linking valid consequences with deductions, and we have corresponding decidability results too. We won't worry however about strong completeness (cf. §3.2(e)), which does actually fail for some modal logics, e.g. for GL which we meet in the next section.

some of them different from any of the systems proposed by Lewis and Langford – thereby making his very elegant result possible.

(h) And now, with the apparatus of relational semantics available, the floodgates really open! After all, the objects in a S-model don't have to represent 'possible worlds' (whatever they are conceived to be); they can stand in for any points in a relational structure. So perhaps they could represent states of knowledge, points of a time series, positions in a game, states in the execution of a program, levels in a hierarchy . . . with different classes of accessibility relations appropriate for different cases and so with different deductive systems to match. The resulting applications of propositional modal logics are very many and various, as you will see.

(i) And what about quantified modal logics, where we add the modal operator $\Box$ to a first-order language? Why might we be interested in them?

Well, philosophers make play with questions like this: Does it make sense to suppose the very same objects can appear in the domains of different possible worlds? If it does, do all possible worlds contain the same objects (perhaps some of them actualized, some not)? Does a proper name (formally a constant term) denote the same thing at any possible world at which it denotes at all? Are basic identity statements, if true at all, necessarily true? Questions of this stripe pile up, and they motivate different ways of tweaking quantified modal logic in formally modelling and so clarifying the philosophical ideas: for example, we can consider how things go with model structures where all the worlds have the same domain of objects, and then consider other model structures where domains can vary from world to world. For more on this, see the readings.

However, the resulting logics don't seem to be of particular interest to non-philosophers (apart from quantified intuitionistic logic, if we consider that as belonging to the family); the wider logical community has been *much* more interested in propositional modal logics.

Still, the beginnings of the technical story about first-order modal logics are pretty accessible. And the suggested readings will enable you to get some headline news about different proof systems and their formal semantics, without getting too entangled in unwanted philosophical debates.

10.2 Provability logic

As just noted, propositional modal logics have a very wide range of applications. But there is one that stands out as being of pre-eminent relevance to anyone beginning mathematical logic. And that is provability logic.

(a) Let's start with some reminders of what you should already know from tackling Gödel's incompleteness theorems (see §6.4). So take a theory in which we can do enough arithmetic: to fix on an example, take first-order Peano Arithmetic. Choose a sensible system of Gödel-numbering. Then you can construct a relational predicate in the language of arithmetic – one which we can abbreviate

$\mathsf{Prf(x,y)}$ – that nicely[4] represents the relation which obtains between two numbers x, y, when x is the Gödel number of a PA proof of the sentence with Gödel number y. Now define $\mathsf{Prov(y)}$ to be the expression $\exists \mathsf{x} \mathsf{Prf(x,y)}$. Then $\mathsf{Prov(y)}$ represents the property that a number y has if it numbers a theorem of PA – so Prov is naturally called a *provability predicate*.

If A is a wff of arithmetic, let $\ulcorner A \urcorner$ be shorthand for A's Gödel-number, and let $\overline{\ulcorner A \urcorner}$ be shorthand for the formal numeral for $\ulcorner A \urcorner$. Then, given our definitions, $\mathsf{Prov}(\overline{\ulcorner A \urcorner})$ says that A is provable in PA.

Now we introduce yet another bit of shorthand: let's use $\boxdot A$ as a simple abbreviation for $\mathsf{Prov}(\overline{\ulcorner A \urcorner})$.[5] With some effort, we can then show that PA proves (unpacked versions of) all instances of the following familiar-looking schemas

$$\mathsf{K}\cdot \quad \boxdot(A \to B) \to (\boxdot A \to \boxdot B)$$
$$\mathsf{S4}\cdot \quad \boxdot A \to \boxdot \boxdot A$$

And moreover we have an analogue of the modal Necessitation rule:

(Nec·) If A is deducible as a PA theorem, then so is $\boxdot A$.

That package of facts about PA is standardly reported by saying that the theory satisfies the so-called *HBL derivability conditions* (named in honour of Hilbert and Bernays who first isolated such conditions, and Löb who gave an improved version). And appealing to these facts together with the First Incompleteness Theorem, it is then easy to derive the Second Theorem that PA cannot prove $\neg\boxdot\bot$ (i.e. can't prove that $\bot$ isn't provable, i.e. can't prove that PA is consistent).[6]

(b) The obvious next question might well seem to be: what *other* modal principles/rules should our dotted-box-as-a-provability-predicate obey, in addition to the dotted principles $\mathsf{K}\cdot$ and $\mathsf{S4}\cdot$, and the rule (Nec·)? What is its appropriate modal logic?

But hold on! We are getting ahead of ourselves, because we so far only have the illusion of modal formulas here. The box as just defined simply doesn't have the right grammar to be a modal operator. Look at it this way. In a proper modal language, the operator $\Box$ is applied to a wff A to give a complex wff $\Box A$ in which A appears as a subformula. But in our newly defined usage where the dotted $\boxdot A$ is short for $\mathsf{Prov}(\overline{\ulcorner A \urcorner})$, the formula A doesn't appear as a subformula at all – what fills the appropriate slot(s) in the predicate Prov is a *numeral* (the numeral for the number which happens to code the formula A).

In short, the surface form of our dotted notation $\boxdot A$ is entirely misleading as to its logical form. Which is why the logically pernickety might not be very happy with the notation.

[4]'Nicely' waves a hand at some details which are important but which we won't need to delay over here!

[5]I've dotted the box here – not the usual notation – for clarity's sake. The reason will appear in just a moment.

[6]For more details, if this is new to you, see for example Chapter 33 of my *An Introduction to Gödel's Theorems* (downloadable from logicmatters.net/igt).

However, it remains the case that our abbreviatory notation is highly suggestive. And what it suggests is starting with a kosher modal propositional language of the kind now familiar for §10.1, where the box is genuinely a unary operator applied to wffs. And then we consider *arithmetical interpretations* which map sentences A of our genuinely modal language to corresponding sentences A^* of PA, interpretations which have the following shape:

i. An interpretative map sends each atomic letter A of our modal language to some corresponding arithmetical sentence A^*, any you like.
ii. The map then respects the propositional connectives: for example, it sends conjunctions in the modal language to conjunctions in the arithmetic language, so $(A \wedge B)^*$ is $(A^* \wedge B^*)$; it sends the absurdity constant to the absurdity constant, i.e. $\bot^*$ is $\bot$; and so on.
iii. Then – the crucial bit – the map sends the modal sentence $\Box A$ to $\boxdot A^*$, i.e. to $\mathsf{Prov}(\overline{\ulcorner A^* \urcorner})$.

There is now no notational jiggery pokery; we have a respectable modal language on the one side, and various interpretative mappings from its sentences into a regular arithmetical language on the other side.

And *now* we can ask a cogent version of the misplaced question we wanted to ask before. In particular, we can ask: what are the modal sentences which are such that, on *any* interpretative mapping into PA, their translations are arithmetical theorems? What, for short, is the correct modal logic for the $\Box$ interpreted *this* way as tracking formal provability in PA?

(c) Here's a reminder of another result we can get from the HBL conditions, namely Löb's Theorem.

Using again our now somewhat deprecated dotted-box-as-abbreviation notation, this rather surprising theorem says:

If PA proves $\boxdot A \to A$, then it proves A.[7]

We will presumably want to reflect this theorem in a logic for the genuinely modal $\Box$ operator interpreted as arithmetical provability: a natural move, then, is to build into our modal logic the rule that, if $\Box A \to A$ is deducible as a theorem, then we can infer A.

So putting this thought together with our previous remarks, let's consider the following modal logic – the 'G' in its name is for Gödel who made some prescient remarks, and the 'L' is for Löb:

(GL) The modal axiomatic system GL is the theory whose axioms are

(Ax i) All instances of tautologies
(Ax ii) All instances of the schema K: $\Box(A \to B) \to (\Box A \to \Box B)$
(Ax iii) All instances of the schema $\mathsf{S4}$: $\Box A \to \Box\Box A$

And whose rules of inference are

[7] See Chapter 34 of *An Introduction to Gödel's Theorems.*

(MP) From A and $A \to B$, infer B
(Nec) If A is deducible as a theorem, infer $\Box A$
(Löb) If $\Box A \to A$ is deducible as a theorem, infer A.

You can immediately see, by the way, that we *don't* also want to add all instances of the T-schema $\Box A \to A$ to this modal logic. For a start, doing that would make $\Box\bot \to \bot$ a theorem and hence $\neg\Box\bot$ would be a theorem. But that can't correspond on arithmetic interpretation to a theorem of PA, since we know that PA can't prove $\neg \boxdot \bot$ (that's the Second Incompleteness Theorem).

And there's worse: leaving aside the desired interpretation of this logic, if we add all instances of $\Box A \to A$ as axioms, then in the presence of the rule (Löb), we can derive any A, and the logic is inconsistent.

Now, given our motivational remarks in defining *GL*, it won't be a surprise to learn that it is *sound* on the provability interpretation. Once we have done the (non-trivial!) background work required for showing that the HBL derivability conditions and hence Löb's theorem hold in PA, it is quite easy to go on to establish that, on every interpretation of the modal language into the language of arithmetic, every theorem of *GL* is a theorem of PA.

And (with more decidedly non-trivial work due to Robert Solovay) it can also be shown that *GL* is *complete* on the provability interpretation. In other words, if a modal sentence is such that every arithmetic interpretation of it is a PA theorem, then that sentence is a theorem of the modal logic *GL*.

Which is all very pleasingly neat!

(d) We should pause to note that there is another way of presenting this provability logic.

Suppose we drop the Löb inference rule from *GL*, and replace the instances of the S4 schema as axioms with instances of the Löb-like schema

L $\Box(\Box A \to A) \to \Box A$

It is then quite easy to see that this results in a modal logic with exactly the same theorems (because *GL* in our original formulation implies all instances of L; and conversely we can show that all instances of S4 can be derived in the new formulation, for which the Löb rule is also a derived rule of inference). Hence either formulation gives us the provability logic for PA.

(e) Now, we've so far been working with arithmetic interpretations of our modal wffs. But we can also give a more abstract Kripke-style relational semantics for *GL* (it is a nice question, though, whether this 'semantics' has much to do with meaning!). We start by defining a *GL*-model in the usual sort of way as comprising a valuation with respect to some worlds W with a relation R defined over them, where R satisfies ...

Well, what conditions *do* we in fact need to place on R so that *GL*-theorems match with the *GL*-validities (the truths that hold at every world, for every *GL*-model)? Clearly, we *mustn't* require R to be reflexive – or else all instances of the T-schema would come out *GL*-valid, and we don't want *that*. Equally clearly, we *must* require R to be transitive – or else instances of the S4-schema could fail to be GL-valid. But we need more: what further condition on R is required to make all the instances of the L-schema come out valid?

It turns out that what is needed is that there is no infinite chain of R-related worlds $w_0, w_1, w_2, w_3, \ldots$ such that $w_0 R w_1 R w_2 R w_3 \ldots$ (and that condition ensures that R is irreflexive, for otherwise we would have some infinite chain $wRwRwRw \ldots$). Call that the finite chain condition. Then define a GL-model as one where the accessibility relation R is transitive and satisfies the finite chain condition. Then a modal sentence is GL theorem if and only if it is GL-valid (true in all worlds in all GL-models).

This new soundness and completeness theorem has a lovely upshot. As with the other modal logics we've met, there is a systematic way of testing for GL-validity (by systematically searching for Kripke-style countermodels). So it is decidable what's a GL theorem.

(f) That last result, together with the fact that GL is sound and complete for arithmetical interpretations into theorems of PA, shows something rather remarkable. Although PA as a whole is an undecidable theory, there is a very interesting *part* of that theory – roughly, what it can say by applying propositional logic and its provability predicate to arithmetical wffs – which *is* decidable.

For example, consider this question: for any arithmetical sentence A, does PA know – i.e. can it prove? – that, if A is provably equivalent to the claim it isn't provable, then A is provably equivalent to saying that PA is consistent? In symbols, using the dotted-box-as-abbreviation, can PA prove

$$\boxdot(A \leftrightarrow \neg \boxdot A) \to \boxdot(A \leftrightarrow \neg \boxdot \bot)$$

Well it can so long as the corresponding modal wff

$$\Box(\mathsf{P} \leftrightarrow \neg\Box\mathsf{P}) \to \Box(\mathsf{P} \leftrightarrow \neg\Box\bot)$$

is a GL theorem – and that's decidable (in fact, it *is* a theorem).

This way, we easily find out a lot more about what PA can prove about what it can and can't prove. And this is just one example of the kind of payoff we get from applying modal logic to questions of provability in arithmetics. Hence the interest of provability logic.

10.3 First readings on modal logic

(a) There is, as so often, a good entry in that wondrous resource the Stanford encyclopaedia, one which should provide more very helpful orientation:

1. James W. Garson, 'Modal logic', *The Stanford Encyclopedia of Philosophy*: read §§1–11 and 15. Available at tinyurl.com/sep-modal.

Now, because of its interest, modal logic is often taught to philosophers without much logical background, and so there are a number of introductions written primarily for them. One often recommended example is the very accessible

2. Rod Girle, *Modal Logics and Philosophy* (Acumen 2000; 2nd edn. 2009). Part I of this book provides a clear introduction, which in 136 pages explains the basic syntax and relational semantics, covering both trees (tableaux) and natural deduction for some propositional modal logics, and extends as far as the beginnings of quantified modal logic.

Philosophers may well very want to go on to read Part II of this book, on applications of modal logic.

But there is a clearer and better-organized account in an extraordinarily useful book by Graham Priest. I'll highlight this not only because it is crisper on modal logics, but because we also get an account of intuitionistic logic in the same tableaux framework:

3. Graham Priest, *An Introduction to Non-Classical Logic** (CUP, much expanded 2nd edition 2008). This treats a whole range of logics systematically, concentrating on semantic ideas, and using a tableaux approach. Chs 1 and 12 provide quick revision tutorials on tableaux for classical propositional and predicate logic. Then Chs 2 and 3 give the basics on propositional modal logics. You can then either fill in more about modal logics in Ch 4 or skip to Ch. 6 on propositional intuitionistic logic. Then Chs 14 and 15 introduce the basics on quantified modal logics. You can then fill in more about quantified modal logics in Chs 16–18 or can then skip to Ch. 20 on quantified intuitionistic logic.

   This whole book – which we will revisit in our next chapter – is a terrific achievement and enviably clear and well-organized.

Then, going half-a-step up in sophistication, though still starting from scratch, we find another excellent book (elegantly done in a way which might appeal more to mathematicians):

4. Melvin Fitting and Richard L. Mendelsohn, *First-Order Modal Logic* (Kluwer 1998). This gives both tableaux and axiomatic systems for various modal logics, in an approachable style and with lucid discussions of options at various choice points. Despite its more mathematical flavour, the book still includes some interesting discussions of the conceptual motivations for different modal logics.

   Read the first half of this book to get a compact but sufficient introduction to propositional modal logics, and also the initial headlines about quantified modal logics. Philosophers will then want to read on.

And let me also mention:

5. Johan van Bentham, *Modal Logic for Open Minds* (CSLI Publications, 2010). This ranges widely and is good at highlighting main ideas and

making cross-connections with other areas of logic. Particularly interesting and enjoyable to read in parallel with the main recommendations.

10.4 Suggested readings on provability logic

Provability logic is nicely introduced in:

6. Rineke Verbrugge, 'Provability logic' §§1–4 and perhaps §6, *The Stanford Encyclopedia of Philosophy*. Available at tinyurl.com/prov-logic.

Or you could dive straight into the very first published book on our topic, which I think still makes for the most attractive entry-point:

7. George Boolos, *The Unprovability of Consistency: An Essay in Modal Logic* (CUP, 1979), particularly Chs 1–12. This fairly short book is a famous modern classic, yet very approachable. And you don't need any prior acquaintance with modal logic in order to tackle it. Boolos has an engaging presentational style (and the book can be read surprisingly quickly in order to get the main news if you are happy to initially skip some of the longer proofs).

However, this seems to be one of the very few distinguished mathematical logic books which is *not* readily available online. So I also need to mention

8. George Boolos, *The Logic of Provability* (CUP, 1993). This is a significantly expanded and updated version of his earlier book. And so you could read the first half of this instead, though I do retain a fondness for the somewhat more streamlined presentations in the shorter version. The main occasion for the update is the presentation of proofs of major results about quantified provability logic which were discovered after Boolos wrote his first book: but these results are really more than you need in a first encounter with provability logic.

And here is another classic introductory book:

9. Craig Smoryński, *Self-Reference and Modal Logic* (Springer-Verlag, 1985). This is a lovely alternative or accompaniment to Boolos's 1979 book. Not lovely to look at, as it oddly printed in extremely small type emulating an electric typewriter, which doesn't make for comfortable reading: but the content is extremely lucidly and elegantly presented, with a lot of helpful explanatory/motivating chat alongside the more formal work. Also highly recommended.

Then, for more pointers towards recent work on related topics you could look at §5 of Verbrugge's article and/or at the following interesting overview:

10. Sergei Artemov, 'Modal logic in mathematics' §§1–5, in *The Handbook of Modal Logic*, edited by P. Blackburn et al. (Elsevier, 2005).

10.5 Alternative and further readings on modal logics

(a) *Other introductory readings for philosophers* The first part of Theodore Sider's *Logic for Philosophy** (OUP, 2010) is poor as an introduction to FOL. However, the second part, which is entirely devoted to modal logic and related topics like Kripke semantics for intuitionistic logic, is very much better, and philosophers could find it rather useful. For example, the chapters on quantified modal logic (and some of the conceptual issues they raise) are brief and approachable.

Sider is, however, closely following a particularly clear old classic by G. E. Hughes and M. J. Cresswell *A New Introduction to Modal Logic* (Routledge, 1996, updating their much earlier book). This can still be recommended and may suit some readers, though it does take a rather old-school approach.

If your starting point has been Priest's book or Fitting/Mendelson, then you might want at some point to supplement these by looking at a treatment of natural deduction proof systems for modal logics. One option is to dip into Tony Roy's long article 'Natural derivations for Priest', in which he provides ND logics corresponding to the propositional modal logics presented in tree form in Priest's book, though this gets much more detailed than you really need: available at tinyurl.com/roy-modal. But a smoother introduction to ND modal systems is provided by Chapter 5 of Girle, or by my main alternative recommendation for philosophers, namely

11. James W. Garson, *Modal Logic for Philosophers** (CUP, 2006; 2nd end. 2014). This again is intended as a gentle introductory book: it accessibly deals with both ND and semantic tableaux (trees), and covers quantified modal logic. It is quite a long book (one reason for preferring the snappier Fitting/Mendelsohn as a first recommendation), with a good coverage of quantified modal logics.

(b) *Modal logics for philosophical applications* If you are interested in applications of propositional modal logics to tense logic, epistemic logic, deontic logic, etc. then the relevant chapters of Girle's book give helpful pointers to more readings on these topics. If your interests instead lean to modal metaphysics, then – once upon a time – a discussion of quantified modal logic at the level of Fitting/Mendelsohn or Garson would have probably sufficed. And for a bit more on first-order quantified modal logics, see

12. James W. Garson, 'Quantification in modal logic' in *Handbook of Philosophical Logic, Vol. 3*, edited by Dov M. Gabbay and F. Guenther (Reidel, 2nd edition 2001).

However, Timothy Williamson's notable book *Modal Logic as Metaphysics* (OUP, 2013) calls on rather more, including e.g. second-order modal logics. There doesn't seem to be general guide/survey of higher-order modal logics at the right sort of level, with the right sort of coverage to recommend here. There *is* a text by Nino B. Cocchiarella and Max A. Freund, *Modal Logic: An Introduction* *to its Syntax and Semantics* (OUP, 2008), whose blurb announces that "a variety of modal logics at the sentential, first-order, and second-order levels are developed with clarity, precision and philosophical insight". However, the treatments in this book are relentlessly and rebarbatively formal. In its last two chapters, the book does cover second-order modal logic: but the highly unfriendly mode of presentation will probably put the discussion out of reach of most philosophers who might be interested. You have been warned.

(c) *Four more technical books* In order of publication, here are some more advanced/challenging texts I can suggest to sufficiently interested readers:

13. Sally Popkorn, *First Steps in Modal Logic* (CUP, 1994). The author is, at least in this possible world, identical with the late mathematician Harold Simmons. This book, which entirely on propositional modal logics, is written for computer scientists. The Introduction rather boldly says 'There are few books on this subject and even fewer books worth looking at. None of these give an acceptable mathematically correct account of the subject. This book is a first attempt to fill that gap.' This considerably oversells the case: but the result is illuminating.
14. Alexander Chagrov and Michael Zakharyaschev *Modal Logic* (OUP, 1997). This is a volume in the Oxford Logic Guides series and again concentrates on propositional modal logics. Definitely written for the more mathematically minded reader, it tackles things in an unusual order, starting with an extended discussion of intuitionistic logic, and is good but rather demanding.
15. Patrick Blackburn, Maarten de Ricke and Yde Venema, *Modal Logic* (CUP, 2001). This is one of the Cambridge Tracts in Theoretical Computer Science: but don't let that provenance put you off! This is an accessibly and agreeably written text on propositional modal logic – certainly compared with the previous two books in this group – with a lot of signposting to the reader of possible routes through the book, and with interesting historical notes. I think it works pretty well, and will also give philosophers an idea about how non-philosophers can make use of propositional modal logic.
16. Lloyd Humberstone, *Philosophical Applications of Modal Logic** (College Publications, 2015). This very large volume starts with a book-within-a-book, an advanced 176 page introduction to propositional modal logics. And then there are extended discussions at a high level of a wide range of applications of these logics that have been made by philosophers. A masterly compendium to consult as/when needed.

10.6 Finally, a very little history

Especially for philosophers, it is very well worth getting to know a little about how mainstream modern modal logic emerged from the to-and-fro between philosophical debate and technical developments. So do read e.g. one of

17. Roberta Ballarin, ‘Modern origins of modal logic’, *The Stanford Encyclopedia of Philosophy*. Available at tinyurl.com/mod-orig.
18. Sten Lindström and Krister Segerberg, ‘Modal logic and philosophy’ §1, in *The Handbook of Modal Logic*, edited P. Blackburn et al. (Elsevier, 2005).

11 Other logics?

So far we have looked at just three variants or extensions of standard FOL:

i. One limitation of FOL is that we can only quantify over objects, as opposed to properties, relations and functions. Yet seemingly, we quantify over properties etc. in informal mathematical reasoning. In Chapter 4, we therefore considered adding second-order quantifiers. (This is just a first step: there is a rich mathematical theory of higher-order logic, a.k.a. type theory, which you will eventually want to explore – but I deem that to be a more advanced topic, so we will return to it in the final chapter, §12.7.)

ii. In Chapter 8 we looked at what happens if we drop the classical law of excluded middle. The resulting intuitionistic logic is mathematically elegant and also widely applicable (in constructive reasoning, in theoretical computer science, in category theory).

iii. Then in Chapter 10 we explored the use of the kind of relational semantics we first met in the context of intuitionistic logic, but now in extending FOL with modal operators. Again, the development on the formal side is mathematically quite elegant: and some modal logics – in particular, provability logic – have worthwhile mathematical applications.

And now, what other exhibits from the wild jungle of variants and/or extensions of standard FOL are equally worth knowing about at this stage, as you begin studying mathematical logic? What other logics are intrinsically mathematically interesting, have significant applications to mathematical reasoning, but can be reasonably regarded as entry-level topics?

A good question. In this chapter, I'll be looking at three relatively accessible variant logics that philosophers in particular have discussed, namely relevant logic, free logic and plural logic. And – spoiler alert! – I'm going to be suggesting that mathematical logicians can cheerfully pass by the first, should have a fleeting acquaintance with the second, and might like to pause a bit longer over the third.

11.1 Relevant logic

(a) Let's concentrate here on one theme. The usual definition of logical consequence makes an inference of the shape $A, \neg A \therefore C$ come out valid, for any

A and for any quite unconnected C; and correspondingly, in proof systems for FOL, we can argue from the premisses A and $\neg A$ to the arbitrary conclusion C. But should we *really* count arguments as valid even when, as in this sort of case, the premisses are totally irrelevant to the conclusion? Shouldn't our formal logic respect the intuitive idea – arguably already in Aristotle – that a conclusion in a valid deduction must have something to do with the premisses?

Debates about this issue go back at least to medieval times. So let's ask: what might a suitable *relevance-respecting logic* look like? Is it worth the effort to use such a logic?

(b) When we very first encounter it in Logic 101, the claim that A and $\neg A$ together entail any arbitrary conclusion C indeed initially seems odd. But we soon learn that this result follows immediately from seemingly uncontentious assumptions. Consider, in particular, these two principles:

> *Disjunctive syllogism is valid.* From $A \lor C$ and $\neg A$ we can infer C.
>
> *Entailment is transitive.* In the simplest case, if A entails B and B entails C, then A entails C. More generally, if Γ and Δ stand in for zero or more premisses, then if Γ entail B and Δ, B entail C, then Γ, Δ entail C

These seem irresistible. Disjunctive syllogism is a principle we use all the time in informal arguments (everyday ones and mathematical ones too). If we've established that one of two options must hold, and can then rule out the first, this surely establishes the second. And the transitivity of entailment is what allows us to chain together shorter valid proofs to make longer valid proofs: reject it, and it seems that the whole practice of proof in mathematics collapses.

But now take the following three arguments:

$$\frac{\mathsf{P}}{\mathsf{P} \lor \mathsf{Q}} \qquad \frac{\mathsf{P} \lor \mathsf{Q} \qquad \neg\mathsf{P}}{\mathsf{Q}} \qquad \frac{\dfrac{\mathsf{P}}{\mathsf{P} \lor \mathsf{Q}} \qquad \neg\mathsf{P}}{\mathsf{Q}}$$

The first just reflects our understanding of inclusive disjunction. The second is the simplest of instances of disjunctive syllogism. The third argument chains together the first two and, since they are valid entailments, this too is valid according to the transitivity principle. So we have shown that P and $\neg\mathsf{P}$ entail Q. And of course, we can generalize. In the same way, we can get from any pair of premisses A and $\neg A$ to an arbitrary conclusion C.

We have just three options, then:

1. Reject disjunctive syllogism as a universally valid principle (or at least, reject disjunctive syllogism for the kind of disjunction for which the inference A so $A \lor C$ is uncontentiously valid).
2. Reject the unrestricted transitivity of entailment.
3. Bite the bullet, and accept what is often called 'explosion', the principle that from contradictory premisses we can infer anything at all.

The large majority of logicians take the first two options to be entirely unpalatable. So they conclude that we should, as in standard FOL, learn to live with explosion. And where's the harm in that? After all, the explosive inference can't actually be used to take us from jointly true premisses to a false conclusion!

Still, before resting content with the explosive nature of FOL, perhaps we should pause to see if there *is* any mileage in either option (1) or option (2). What might a *paraconsistent* logic – one with a non-explosive entailment relation – look like?

(c) Logicians are an ingenious bunch. And it isn't difficult to cook-up a formal system for e.g. a propositional language equipped with connectives written $\wedge$, $\vee$ and $\neg$, for which analogues of disjunctive syllogism and explosion don't generally hold.

For example, suppose we adopt a natural deduction system with the usual introduction and elimination rulers for $\wedge$ and $\vee$ (as in §8.1). But the additional rules governing negation are now just De Morgan's Laws and a double negation rule (the double inference lines indicate that you can apply the rules both top to bottom and also the other way up).

$$(\neg\wedge)\quad \frac{\neg(A \wedge B)}{\neg A \vee \neg B} \qquad (\neg\vee)\quad \frac{\neg(A \vee B)}{\neg A \wedge \neg B} \qquad (\neg\neg)\quad \frac{\neg\neg A}{A}$$

The resulting logic is standardly called *FDE* for reasons that needn't delay us. And a little experimentation should convince you that, with only the *FDE* rules in place, we can't warrant either disjunctive syllogism or explosion.

But so what? By itself, the observation that dropping some classical rules stops you proving some classical results has little interest. Contrast the intuitionist case, for example. There we are given a semantic story (the BHK account of the meaning of the connectives) which aims to *justify* dropping the classical double negation law. Can we similarly give a semantic story here which would again justify dropping some classical rules and this time only underpin *FDE*?

(d) Suppose – just suppose! – we think that there are *four* truth-related values a proposition can take. Label these values T, B, N, F. And suppose that, given an assignment of such values to atomic wffs, we compute the values of complex wffs using the following tables:

$A \wedge B$	T	B	N	F
T	T	B	N	F
B	B	B	F	F
N	N	F	N	F
F	F	F	F	F

$A \vee B$	T	B	N	F
T	T	T	T	T
B	T	B	T	B
N	T	T	N	N
F	T	B	N	F

A	$\neg A$
T	F
B	B
N	N
F	T

These tables are to be read in the obvious way. So, for example, if P takes the value B, and Q takes the value N, then $\mathsf{P} \wedge \mathsf{Q}$ takes the value F, $\mathsf{P} \vee \mathsf{Q}$ takes the value T, and $\neg\mathsf{P}$ takes the value B.

Suppose in addition that we define a quasi-entailment relation as follows: some premisses Γ entail* a given conclusion C – in symbols $\Gamma \vDash^* C$ – just if, on any valuation which makes each premiss either T or B, the conclusion is also either T or B.

Then, lo and behold, we can show that *FDE* is sound and complete for this semantics – we can derive C from premisses Γ if and only if $\Gamma \vDash^* C$. And note, as we wanted, the analogue of disjunctive syllogism is not always a correct entailment*: on the same suggested valuations, both $\mathsf{P} \lor \mathsf{Q}$ and $\neg\mathsf{P}$ are either T or B, while Q is N, so $\mathsf{P} \lor \mathsf{Q}, \neg\mathsf{P} \nvDash^* \mathsf{Q}$. And we don't always get explosion either, since both P and $\neg\mathsf{P}$ are B while Q is N, it follows that $\mathsf{P}, \neg\mathsf{P} \nvDash^* \mathsf{Q}$.

Which is all fine and good in the abstract. But what are these imagined four truth-related values? Can we actually give some interpretation so that our tables really do have something to do with truth and falsity, with negation, conjunction and disjunction, and so that entailment* does arguably become a genuine consequence relation?

Well, suppose – just suppose! – that propositions can not only be plain true or plain false but can also be *both* true and false at the same time, or *neither* true nor false. Then there will indeed be four truth-related values a proposition can take – T (true), B (both true and false), N (neither), F (false).

And, interpreting the values like that, the tables we have given arguably respect the intuitive meaning of the connectives. For example, if A is both true and false, the same should go for $\neg A$. While if A is both true and false, and B is neither, then $A \lor B$ is true because its first disjunct is, but it isn't also false as that would require both disjuncts to be false (or so we might argue). Similarly for the other table entries. Moreover, the intuitive idea of entailment as truth-preservation is still reflected in the definition of entailment*, which says that if the premisses are all true (though maybe some are false as well), the conclusion is true (though maybe false as well).

(e) What on earth can we make of this supposition that some propositions are both true and false at the same time? At first sight, this seems simply absurd.

However, a vocal minority of philosophers *do* famously argue that while, to be sure, regular sentences are either true or false but not both, there are certain special cases – e.g. the likes of the paradoxical liar sentence 'This sentence is false' – which *are* simultaneously both true and false.

It is fair to say that rather few are persuaded by this extravagant suggestion. But let's go along with it just for a moment. And now note that it isn't immediately clear that this really helps. For suppose we *do* countenance the possibility that certain special sentences have the deviant status of being both true and false (or being neither). Then we might reasonably propose to add to our formal logical apparatus an operator '!' to signal that a sentence is *not* deviant in that way, an operator governed by the following table:

A	$!A$
T	T
B	F
N	F
F	T

Why not? But then it is immediate that $!\mathsf{P}, \mathsf{P}, \neg\mathsf{P} \vDash^* \mathsf{Q}$. And similarly, if (say) P and Q are the atoms present in A, then $!\mathsf{P}, !\mathsf{Q}, A, \neg A \vDash^* C$ always holds. So, if built out of regular atoms (expressing ordinary non-paradoxical claims), a contradictory pair entails* anything. Yet surely, if we were seriously worried by the original version of explosion, then this modified form will be no more acceptable.

(f) We said that most logicians bite the bullet, and accept explosion because they deem it harmless. But are they right?

It seems fundamental to a conditional connective $\to$ that it obeys the principle of conditional proof. In other words, if the set of premisses Γ plus the temporary assumption A together entail C, that shows that Γ entails $A \to C$. But then suppose we *do* accept the explosive inference from $\neg A$ and A to C. Applying conditional proof, we will have to agree that given $\neg A$, it follows that $A \to C$, for any unrelated consequent C. And this, some will say, is just the unacceptable face of the classical (or intuitionistic) conditional: so we should reject explosion, not just for its prima facie oddity, but also to get a nice conditional.

Now, if you have learnt to live happily with the standard conditional of classical or intuitionistic logic as an acceptable regimentation for serious mathematical purposes, then you won't be much moved by this argument. But what if you *do* want to add a conditional connective where the inference from $\neg A$ to $A \to C$ generally fails?

Within an *FDE*-like framework, we can play with four-valued tables again, now for the connective $\to$. But on the more plausible ways of doing this, we will still have $!\mathsf{P}, \neg\mathsf{P} \vDash^* \mathsf{P} \to \mathsf{Q}$; and more generally, for wffs built out of regular atoms, the conditional is just the material conditional again. So again, if we were worried about the material conditional before, we should surely stay worried about this sort of four-valued replacement.

(g) Let's very briefly take stock.

We can run up proof systems like *FDE* which lack disjunctive syllogism and explosion and where $\neg A$ doesn't imply $A \to C$. Further, we can give these systems what *looks* like a semantics e.g. using four values (or alternatively we could use Kripke-style valuations over some relational structure). But if this exercise isn't just to be an abstract game, then we do need to tell a story about how to interpret the formal 'semantics' in order to link everything up with considerations about truth and falsity and inference. And as we see in the initial case of *FDE*, the supposed linkage *can* embroil us with highly implausible claims (some propositions can be both true *and* false – really?). Moreover, while our resulting logic may not be classical overall, if we are allowed to distinguish regular true-or-false propositions from those that behave deviantly according to the enhanced semantic story, then in its application to the regular propositions, the new logic *can* simply collapse back into classical logic again (with an entailment relation and a conditional that don't respect worries about relevance).

So already the price of avoiding exposition by rejecting disjunctive syllogism in the manner of *FDE* is beginning to look as if could be unattractively high while the real gains remain pretty unclear.

But of course, all this is just an opening skirmish. There is a great deal more than can be said, and which has been said, as you will find (to repeat, logicians are an ingenious bunch). Though by my lights things only get worse when we move on from the relatively simple *FDE* to fancier relevant logics such as the one standardly called simply R. In the case of R, for example, the semantic story is not superficially-clear-but-implausible (as for *FDE*) but downright obscure without any attractive motivation for ordinary logical use. Or so say most of us.

I'll give readings on these sorts of semantically deviant relevant logics which you can follow up if you want: but this is a rabbit hole that most mathematical logicians very sensibly won't want to disappear down. (I didn't say that this Guide would never be opinionated!)

(h) What about avoiding explosion not by rejecting disjunctive syllogism but by rejecting the unrestricted transitivity of entailment? At first sight, this idea might seem to be complete non-starter: as Timothy Smiley once put it, "the whole point of logic as an instrument, and the way in which it brings us new knowledge, lies in the contrast between the transitivity of 'entails' and the non-transitivity of 'obviously entails', and all this is lost if transitivity cannot be relied on."

But perhaps, after all, there is wriggle-room here. Yes, in general, it is essential to maintain the transitivity principle that if Γ entails B and Δ, B entail C, then Γ, Δ entail C. But what about the special case where Γ includes A while Δ includes $\neg A$: shouldn't that give us pause before we put Γ and Δ together as joint premisses? Rather than combining those explicitly inconsistent premisses and arguing onwards regardless, shouldn't we instead – so to speak – raise a red flag, and declare that Γ, Δ together are absurd, and only allow the inference from Γ, Δ to $\bot$? In other words, the suggestion might go, transitivity holds except when it shouldn't, i.e. except when we have explicitly contradictory premisses on the table and we should flag the absurdity. (So we can't cogently put the inference A to $A \vee C$ together with the disjunctive syllogism from $A \vee C$ and $\neg A$ to C to justify the explosive entailment from A and $\neg A$ to C: we should restrain ourselves and stick to the inference from A and $\neg A$ to $\bot$.)

Now, compared with the proposal that we should achieve a relevant logic by adopting a deviant semantics and rejecting disjunctive syllogism, this actually seems a positively attractive suggestion. But can we actually develop the leading idea into a smoothly workable logical system without its own oddities?

Well, Neil Tennant has long been arguing that we can arrange things so that we get very recognizable natural deduction rules but only the described more restricted form of transitivity. In other words, we can get a proof system in which we can paste proofs together when we ought to be able to, or else we must combine the proofs to expose that we now can generate a contradiction. And this, as Tennant emphasizes, looks like an epistemic plus-point, if we are forced to highlight a contradiction when one is there to be exposed.

Tennant advertises his proof system as *core logic* – it comes in two versions, one classical and one intuitionistic. His claim is that core logic captures what we actually need in mathematical and scientific reasoning (classical or constructive), without some of the unwanted extras. However, to avoid explosion reappearing, the operations of Tennant's natural deduction system for his core logic are inevitably subject to additional constraints as compared with the more free-wheeling proof-structures allowed in standard systems for classical or intuitionistic systems. See the reading for more details.

So here's the obvious next question: is the occasional potential epistemic gain from requiring proofs to obey the strictures of 'core logic' actually worth the additional effort of strictly following its rules? A judgement call, of course. But most mathematical logicians are going to return a negative verdict and, despite Tennant's energetic advocacy, feel quite comfortable on cost-benefit grounds of sticking with their familiar ways.

11.2 Readings on relevant logic

A familiar resource once more provides some excellent entry-points:

1. Graham Priest, 'Paraconsistent logic', *The Stanford Encyclopedia of Philosophy*, tinyurl.com/paracons. As Priest notes, any logical system counts as paraconsistent as long as it is not explosive; there are a variety of motivations for a variety of paraconsistent systems. This is a very clear introduction to some of the options.
2. Edwin Mares, 'Relevance logic', *The Stanford Encyclopedia of Philosophy*, tinyurl.com/rel-logic. This, among other things, very usefully summarizes a number of semantic interpretations that have been proposed for relevant logics. Some depend on information-theoretic ideas that might e.g. be of use in computer science: it is much less clear what their significance for mathematical reasoning might be.

Or instead of (2) you could look at

3. Edwin Mares and Robert Meyer, 'Relevant logics', in L. Goble, ed, *The Blackwell Guide to Philosophical Logic* (Blackwell 2001).

And if you just want to know what it takes to get a relevance-respecting logic by the route of semantic revisionism, these initial pieces should suffice. You may well then quickly decide that you don't want to pay the price, and be happy to accept the verdict of e.g.

4. John Burgess, 'No requirement of relevance', in S. Shapiro, ed., *The Oxford Handbook of the Philosophy of Mathematics and Logic* (OUP, 2005). (Initially, you can skip the later pages of §3, on Tennant.)

If, however, you *are* tempted to explore further, the following is a terrific resource, already familiar from the recommended readings on modal logic:

5. Graham Priest, *An Introduction to Non-Classical Logic** (CUP, 2nd edition 2008). As we said before, this treats a whole range of logics systematically, concentrating on semantic ideas, and using a tableaux approach. Chs 7–10 discuss some propositional many-valued logics (including ones with truth-value 'gaps' and 'gluts'), *FDE*, *R*, and much else besides: then Chs 21–24 discuss their quantificational counterparts.

And, taking a step up in level, here is the same author again vigorously making the case for taking paraconsistent logics seriously:

6. Graham Priest, 'Paraconsistent logic', in the *Handbook of Philosophical Logic*, Vol. 6, ed. by D. Gabbay and F. Guenthner, (Kluwer 2nd edition 2001), pp. 287–393.

You could also follow up Mares's *SEP* article by taking a look at his book:

7. Edwin Mares, *Relevant Logic: A Philosophical Interpretation* (CUP 2004). As the title suggests, this book has very extensive conceptual discussion alongside the more formal parts elaborating what might be called the mainstream tradition in relevance logics.

However, I for one am unpersuaded and remain on Burgess's side of the debate, at least as far as relevance-via-semantic-revisionism is concerned.

Going now in a different direction, what about Tennant's idea of instead buying a certain amount of relevance by restricting the transitivity of entailment? For a very lucid introductory account, see

8. Neil Tennant, 'Relevance in reasoning', in S. Shapiro, ed., *The Oxford Handbook of the Philosophy of Mathematics and Logic* (OUP, 2005).

And for a full-blown development of these ideas, see

9. Neil Tennant, *Core Logic* (OUP, 2017). This an ambitious and rich book, though mostly very accessible, and as I noted in §9.5 it is well worth reading for its many more general proof-theoretic insights, even if you are not persuaded by Tennant's version of relevantism.

In his final chapter, by the way, Tennant responds to the technical challenges laid down by Burgess in §3 of his paper.

11.3 Free logic

It is often said that pure logic should be topic-neutral. But FOL arguably isn't entirely topic-neutral. In particular it isn't neutral about existence assumptions:

(i) Domains of quantification are assumed to be non-empty.
(ii) Names are assumed to have denotations in the domain, and definite descriptions (constructions of the kind *the x such that Fx*) are massaged away because they might lack a denotation.

(iii) Functions are assumed to be total, i.e. for each object of the domain as input, the function returns a value.

Do any of these features really matter? Is it worth the effort to construct a suitable logic free of such existence assumptions?

(a) Start with an elementary point: in standard FOL, $\forall xFx$ entails $\exists xFx$. And here's a Gentzen-style natural deduction derivation to prove the point:

$$\cfrac{\cfrac{\forall xFx}{Fa}}{\exists xFx}$$

The first line states our premiss. At the second line, the story goes, we pick an arbitrary member of the domain and dub it with a temporary name and then infer . . .

But not so fast! What if the domain is empty? Then there is nothing to pick out and dub.

So our natural deduction derivation at the second line in effect presupposes that the domain is non-empty. Which ties in, of course, with the usual semantics for an FOL language, where we stipulate that domains of quantification are always indeed non-empty.

Deploying standard FOL to regiment a theory about Xs and using quantifiers which range over Xs, then, makes an ontological assumption – namely, that there *are* some Xs (at least one). For example, when we adopt the usual first-order logical framework for doing formalized set theory, with quantifiers ranging over sets, we are assuming that some sets exist (at least one) for our quantifiers to range over.[1]

Suppose then that we want to drop the existential presumption and allow for the possibility that our domain of quantification is empty. In an empty domain, $\forall xFx$ can be vacuously true (anything in the domain satisfies F!) while $\exists xFx$ is false; so we'll have to revise our logical laws. But should we bother?

Here's a line of argument on one side:

> An inference is logically valid just if it is necessarily truth-preserving in virtue of topic-neutral features of its structure. And formal logic is the study of logical validity in this sense, using regimented languages to enable us to bring out how arguments of certain forms are valid irrespective of their subject-matter.
>
> Now, sometimes we want to argue logically about the properties of things which we already know to exist (electrons, say). Other times we want to argue in an exploratory way, in ignorance of whether what

[1] Oliver and Smiley in their *Plural Logic* – about which more in the next section – have fun chastising some set theorists for getting sloppy about this. For example, they quote J.R. Shoenfield saying "we can use the usual axioms of logic to conclude that there is at least one set". But this is, strictly speaking, to get things exactly upside down: it is only because we have *already* presupposed that there is at least one set that we can deploy the usual axioms of FOL in doing formalized set theory.

> we are talking about exists (superstrings, perhaps). While sometimes we want to argue about things that we believe don't exist, precisely in order to try to show that they don't exist (tachyons, perhaps). And logic should aim to regiment correct forms of inference which we can apply topic-neutrally across these different cases, without our taking any stance about how things are in the world.
>
> Hence *one* way our formal logic should be topic-neutral is by allowing empty domains. But standard FOL rules – being incorrect for empty domains – are not topic-neutral. So they don't reliably capture only logical validities and logical truths. Therefore our standard logic needs revision.

And how might the defender of our standard FOL logic reply?

> There is no One True Logic. Choosing a formal logic always involves weighing up costs and benefits. And the very small benefit of having a logic whose inferential principles also hold in empty domains is just not worth the albeit minor additional cost. After all, when we want to argue about things that do not/might not exist, we already have sufficient resources while still using standard logic.
>
> First, a suitably inclusive wider domain is usually easily found (one will typically be in play when engaged in serious inquiry as opposed to artificial classroom examples). For example, suppose we are arguing about tachyons. Instead of taking the domain to be tachyons and regimenting the proposition that tachyons are really weird as $\forall x\, Wx$, we can more naturally take the domain more inclusively to be, say, physical particles. We can then regiment that proposition that tachyons are weird along the lines of $\forall x(Tx \to Wx)$ and lose the unwanted FOL inference that some really weird particles exists, $\exists x\, Wx$.
>
> But put that first manoeuvre aside. Suppose we want to adopt a domain to work in but we have lingering doubts about its legitimacy. Then, second, we can and do proceed in an exploratory, noncommittal, suppositional mode.
>
> For example, consider some mathematical inquiry which proceeds in the supposedly all-inclusive framework of full-blown set theory. What if we are sceptical about this wildly proliferating world of sets? No problem. We can bracket our set-theoretic investigations with an unspoken 'Ok, let's take it, for the sake of argument, that there *is* this extravagantly infinitary universe that standard set theory supposedly talks about, and see what follows . . . '. And then, within the scope of that bracketing assumption, we plunge in and quantify over sets in the usual way, and continue our explorations *as if* we are dealing with a suitably populated domain, to see where our investigations get to. (Of course, if we start off assuming in a hypothetical spirit

that there are at least some Xs, our enquiries might in fact lead us in the end to backtrack and reject that assumption!)

Now, once we have made the supposition for the sake of further exploration that there *are* sets (or superstrings or whatever Xs we are interested in), we might very reasonably want the same logical laws to apply in each case, topic-neutrally. But there is no need for this logic we use, once we are working within the scope of the supposition that we *are* talking about something, to *continue* to remain neutral about whether there is anything in the domain.

In other words, the topic-neutrality we want can be downstream from the fundamental presumption that we are talking about something rather than nothing.

Now, the debate, all too predictably, will continue. But we have perhaps said enough to give some support to the usual view: particularly for the purposes of regimenting mathematical reasoning, the suggestion goes, it is quite defensible to stick with a standard logic (classical or intuitionist) which relies on the presumption that we aren't talking about nothing at all.[2]

See the suggested readings, however, for accounts of how to give a so-called *inclusive* version of FOL which allows empty domains, if you really *do* want one.

(b) In standard first-order logic we assume not only that the domain of quantification is populated, but also that every name (individual constant) in a FOL language successfully denotes some object in the domain. In other words, we ordinarily ban not only empty domains but also empty names.

In informal argumentation, by contrast, we quite often use empty names. This can be by mistake – as when nineteenth century astronomers used 'Vulcan', the name introduced for a postulated intra-Mercurial planet, or perhaps as when we now use 'Homer', if that's the name for the supposed common creator of the *Iliad* and the *Odyssey*. We can also use empty names more knowingly – as when we use 'Athena' or 'Hogwarts'.

Now, since logic is supposed to be topic neutral, we should be able to regiment argumentation with names independently of whether they successfully refer. So we need a logic free of the assumption that all names denote. Or so the story goes.

But on the other side, it might be responded that we can and should cheerfully set aside concerns about *fictions* like Athena or Hogwarts. It might be quite tricky to give a good general story about straightforwardly fictional discourse, but *that* problem needn't delay the mathematical logician. Further, it might be said, we don't need a new logic to deal with the serious but mistaken use of a name which in fact fails to refer: the mistaken reasoner who uses the usual valid

[2]Looking ahead, model theorists find it useful to allow empty structures with nothing in their domain: see e.g. the books by Hodges and Rothmaler mentioned in §12.2. But as those authors note, this is a matter of convenience on which nothing hangs, and it is quite compatible with that to continue to prefer to define first-order consequence in terms of models which are required to have populated domains.

forms of arguments has simply failed to meet the conditions for their correct application. We don't need to revise their logic but to get them straight about their reference-failure. Nor do we need to be revisionist to deal e.g. with the more tentative use of name in the scope of an assumption for the sake of argument, as in 'OK, assuming now that there *is* such a planet as Vulcan, . . . ': within the scope of the assumption we use standard logic.

As with empty domains, then, it isn't *obvious* that a sensible commitment to the topic neutrality of logic requires us to revise our logic to accommodate empty names. But suppose we *do* want a logic free from existence-assumptions. Then we will have to revise our logical laws.

For example, $\forall x Fx$ now won't always entail Fc, whichever name c we choose, as that name might not denote something in the domain. We'll need some sort of existence predicate available (conventionally written $E!$), so that $E!c$ holds just when c really does denote something in the domain of quantification. And then our $\forall$-elimination rule can be (a more general version of): from $\forall x Fx$ and $E!c$ we can infer Fc. And we'll similarly need to doctor other quantifier rules. For example, our $\exists$-introduction rule will be (a more general version of): from Fc and $E!c$ we can infer $\exists x Fx$.

So the idea is that we allow empty terms, but in effect restrict the application of the quantifier rules to the non-empty ones. But there are complications. Suppose $\neg E!c$, so c is an empty name. Then what is the truth value of a wff like Fc? One line to take is that it is always simply false. Another line to take is that such a wff can sometimes be true (compare 'Athena is Athena', 'Athena is a goddess'). A third line is that simple sentences like Fc with empty terms are simply truth-valueless – if there is a gap where the reference of the name should be, then there is a truth-value gap. Pursuing these lines lead to, respectively, negative, positive, and neutral free logics. For some details, again see the readings: I leave you to judge the relative merits of these three lines.

(c) If our concern is to regiment mathematical reasoning, it is rather unclear what we gain by officially allowing empty domains and/or empty names. Some argue, though, that free logic comes more into its own when we turn to the treatment of definite descriptions of the form *the F*.

If we want to regiment an informal claim of the form *The F is G* into a standard, unaugmented, first-order language, the best we can do is this (or one of its logical equivalents):

$$\exists x(Fx \land \forall y(Fy \to y = x) \land Gx).$$

That's quite uncontroversial. What *is* controversial is Bertrand Russell's claim that this rendition in some sense correctly captures the underlying logical form of the ordinary language claim (that is his famed 'Theory of Descriptions'): in other words, definite descriptions – on his view – are not genuine singular terms, but are to be massaged away via a contextual definition.

But can't we after all treat definite descriptions as genuine terms? It would certainly seem more natural to add to the resources of a first-order language a description-forming operator which takes a predicate F, for example, and forms the expression $\imath x F x$ (read *the x such that Fx*) which is a *term* referring to the one and only thing which satisfies the predicate. And then our formal regimentations can more closely respect the surface form claims involving a definite description. Just as *Kripke is clever* might get formally rendered as Gk, something like *The inventor of modal semantics is clever* might get formally rendered as $\mathsf{G}(\imath x F x)$.

But of course, the snag is that there may be nothing at all that satisfies a given predicate F, or there might be too many things that satisfy the predicate. In either case the term $\imath x F x$ will lack a reference, and will be an empty term. So now the options fork.

1. We can add a definite description operator to our language, but only allow its application to a predicate F if we are entitled to assume that there is exactly one thing that satisfies F. In which case $\imath x F x$ in effect behaves like a newly minted name which, like standard names, has a reference, and we can cheerfully sail on, still using standard FOL.

2. We can allow unrestricted use of the description operator, without prior checks that the terms we form have a reference. In which case we will need to adopt a free logic to cope with the cases of empty definite descriptions. There are various strategies for doing this, depending on whether we want our free logic to be negative, positive or neutral.

In practice, mathematical reasoners tend in *many* cases simply to follow an informal version of option (1). For example, having shown that there is an F less than all the others, they will then cheerfully talk about *the minimum F* – so they first ensure that the use of the description will be backed up with an existence and uniqueness proof. And then the logic for dealing with such an introduced singular term can then remain standard FOL.

However, there is a special class of further cases; and this is – I think – where things get more interesting.

(d) The standard semantic story treats function expressions of a FOL language as denoting total functions – for any object of the domain as input, the function yields a value in the domain as output. Mathematically, however, we often work with partial functions: that's particularly the case in computability theory, where the notion of a partial recursive function is pivotal. Partial recursive functions, recall, are defined by allowing the application of a minimization or least search operator, which is basically a definite description operator which may fail to return a value (see §6.2(c)). So, it might well seem that in order to reason about computable functions we need a logic which is neutral about whether function values always exists, i.e. a free logic which can accommodate partial functions and definite descriptions that fail to refer.

This is a claim often made by proponents of free logic. It is vigorously pressed by Oliver and Smiley (in the chapter mentioned in the next section). Yet they give no examples at all of places where mathematical reasoners doing recursive function theory actually use arguments that need to be regimented by changing our standard logic. And if we turn to mainstream theoretical treatments of partial recursive functions in books on computability – including those by philosophically minded authors like Enderton, Epstein/Carnielli or Boolos/Jeffrey (see §6.5) – we find not a word about needing to revise our standard logic and adopt a free logic. So what's going on here?

I think we have to distinguish two quite different claims:

1. Suppose we want to revise the usual first-order language of arithmetic to allow partial recursive functions, and then construct a formal theory in which we can e.g. do computations of the values of the partial recursive functions (when they have one) in the way we can do simpler formal computations as derivations inside PA (or inside PRA, formal Primitive Recursive Arithmetic). Then this formal theory with its partial functions will need to be equipped with a free logic to allow for reference failures.

2. When, it comes to proving general results *about* partial recursive functions in our usual informal mathematical style, we need to deploy reasoning which presumes a free logic.

Now, (1) may be true. But mathematicians in fact seem to have very little interest in that formalization project (though some computer scientists have written around the topic). What they care about is the general theory of computability.

And there seems no good reason for supposing (2) is true. Work through a mathematical text on the general theory of computability, and you'll see that some care is taken to handle cases where a function has no output. For example, we introduce the notation $f(x)\downarrow$ to indicate that f in fact has an output for input x; and we introduce the notation $f(x) \approx g(x)$ to indicate that either (i) both $f(x)\downarrow$ and $g(x)\downarrow$ and $f(x) = g(x)$ or (ii) neither $f(x)$ nor $g(x)$ is defined. And then our theorems are framed using this sort of notation to ensure that the mathematical propositions which are stated and proved are straightforwardly true (and aren't threatened with e.g. truth-valueness because of possibly empty terms). Reflection on the arguments actually deployed by Enderton etc. suggests that the silence of those authors on the question of revising our logic is entirely appropriate. Theorists of computability, it seems, don't need a free logic.

(e) I have suggested, then, that it is – to say the least – far from clear that mainstream mathematicians going about their ordinary business need an inclusive logic admitting empty domains or a logic admitting empty names or definite descriptions in general (though there has been a dissenting tradition about this). The case for admitting partial functions in our formalized object language is more interesting; but it still seems that in regimenting our mathematical general enquiry about such functions, we still don't need a free logic. So is there any interest in free logic for those beginning mathematical logic?

Well, philosophers occasionally get into a tangle deploying arguments where existence assumptions are smuggled in, and using a free logic to regiment the arguments will expose where the existence assumptions are needed.[3]

[3]See for example this paper where Michael Potter and Timothy Smiley diagnose a failure

Or turn to quantified modal logics where we use a 'possible worlds' semantics. Here we might want to consider relational structures where the domains vary from world to world, and then some things that we have names for at the actual world may not exist in some worlds, and we'll need a free logic in evaluating wffs at different worlds. And, relatedly, it is a nice question how we should treat questions of identity and existence in quantified Intuitionistic logic – there are troublesome issues here which are touched on in the *SEP* article mentioned below. But we have perhaps said enough for now.

11.4 Readings on free logic

Philosophers might appreciate this gentle warm-up introduction:

1. David Bostock, *Intermediate Logic* (OUP 1997), Ch. 8.

Then, for rather more, see one of

2. Karel Lambert, 'Free logics', in L. Goble, ed, *The Blackwell Guide to Philosophical Logic* (Blackwell 2001).
3. John Nolt, 'Free logic', *The Stanford Encyclopedia of Philosophy*, available at tinyurl.com/free-log.

Or even better,

3. John Nolt, 'Free logics', in D. Jacquette, ed., *Philosophy of Logic: Handbook of the Philosophy of Science, Vol 5* (North-Holland 2007), pp. 1023-1060. A more expansive essay covering the ground of the same author's *SEP* article.

   This is a judicious and even-handed survey of many of the main issues and options. Nolt writes "Though unsullied by existential commitment, free logic does not reveal a tidy and compelling realm of logical truth. In fact, the whole business is disappointingly messy." But for all that, he does conclude that "In logic, as elsewhere, freedom, though messy, is often desirable."

And here's a similar survey essay:

4. Ermanno Bencivenga, 'Free Logics', in D. Gabbay and F. Guenthner, eds., *Handbook of Philosophical Logic, vol. III: Alternatives to Classical Logic* (Reidel, 1986). Reprinted in D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, 2nd edition, vol. 5 (Kluwer 2002).

Moving on from general introductions to detailed formal treatments of various kinds, the following are worth looking at:

to allow for empty terms in one kind of argument for a so-called neo-logicist foundation for arithmetic: 'Abstraction by recarving'. *Proc. of the Aristotelian Society*, 101 (2001), 327–38. They recommend using a free logic to expose where the argument goes wrong.

5. Elliott Mendelson, *Introduction to Mathematical Logic* (Chapman and Hall/CRC, 6th edn. 2015). §2.16, 'Quantification theory allowing empty domains', presents an inclusive logic in an axiomatic framework.
6. Neil Tennant, *Natural Logic* (Edinburgh UP 1978, 1990), §7.10. Available at tinyurl.com/nat-logic. An early and original presentation of a free logic in a natural deduction framework.
7. Graham Priest, *An Introduction to Non-Classical Logic** (CUP, 2nd edition 2008), Ch. 13, and also Ch. 21. As you would now expect, neatly and briskly presented tableau systems for various free logics.
8. Alex Oliver and Timothy Smiley, *Plural Logic* (OUP 2013: revised and expanded second edition, 2016). Before giving formal systems for plural logics in later chapters, Ch. 11 gives an original neutral free logic with interesting features.

Finally, let me mention two collection of articles around and about our topic, slightly old now, but likely still to be of some interest to philosophers: Karel Lambert, ed., *Philosophical Applications of Free Logic* (OUP 1991) reprints some classic papers including a famous and influential one by Dana Scott; and for essays by Lambert alone, see his *Free Logic: Selected Essays* (CUP 2003).

11.5 Plural logic

(a) Committed proponents of relevant logic claim that the entailment relation built into standard FOL is badly flawed, because it doesn't respect relevance requirements. Proponents of free logics claim that FOL is badly flawed by not being fully topic-neutral. Proponents of plural logic, by contrast, need have no beef with our standard logic: but they argue that we should extend our logical resources to cope with an important class of arguments which are valid in virtue of their form but which arguably escape being regimented in FOL, namely those which depend on the use of plural locutions.

For a simple non-mathematical example, take the argument 'The Brontë sisters were inseparable; Anne, Charlotte and Emily are the Brontë sisters; so Anne, Charlotte and Emily were inseparable'. Plainly valid, and surely valid in virtue of its form not its specific subject matter. And note, the predicate 'were inseparable' is a so-called collective predicate – meaning that it applies to the sisters, plural, taken collectively together, but not to any one sister taken individually. For another example, take the quantified argument 'Whoever successfully stormed the citadel co-ordinated their attack well. The Greek warriors led by Odysseus successfully stormed the citadel. So the Greek warriors led by Odysseus co-ordinated their attack well.' Surely valid in virtue of its form – and we can note again that 'co-ordinated their attack' is another collective predicate requiring a plural subject.

Next, let's emphasize that we argue with plurals all the time not just in non-mathematical contexts but in informal mathematical English too. For example, we use plural denoting terms like '2, 4, 6, and 8', 'the prime numbers', 'the real numbers between 0 and 1', 'the complex solutions of $z^2 + z + 1 = 0$', 'the points where line L intersects curve C', 'the ordinals', 'the sets that are not members of themselves'. There are mathematical collective predicates which require plural subjects like 'are colinear', 'are countable', 'are isomorphic', 'are well-ordered', 'are co-extensive' and the like. We also often generalize by using plural quantifiers like 'any natural numbers' or 'some reals' together with linked plural pronouns such as 'they' and 'them'. For example, here is a plural version of the Least Number Principle: 'Given any natural numbers, at least one, then one of them must be the least.' A contrasting claim: 'There are some reals – those strictly between 0 and 1 are a case in point – such that no one of *them* is the least.'

If we are in the business of regimenting arguments in mathematical English, then, such examples suggest we should be interested in developing a plural logic. We will want to introduce logical devices going beyond those available in FOL languages – such as plural denoting terms in addition to singular terms, predicates allowing or even requiring plural subjects, and plural quantifiers and matched plural pronouns – and then we will want to explore the rules for arguing with these devices. Why not?

(b) So let's start by introducing some (not-quite-standard) notation.

It is conventional to use early/mid-alphabet lower-case letters as singular constants, and end-of-alphabet lower-case letters as variables which take singular values. We'll now adopt the general policy of using capitalization to indicate the plural counterparts to singular expressions. So:

1. Some capitalized letters such as N, Q, R serve as *plural constants*, typically denoting more than one object (as it might be, the natural numbers, the rationals, the reals).
2. Capitalized end-alphabet letters such as X, Y, Z serve as *plural variables.*
3. $\forall X$, $\exists Y$, etc, are then plural quantifiers – *for any objects* X (from some domain of quantification), *for some objects* Y.

We will then want to be able say of some objects, plural, that they include a particular individual object:

4. $n \varepsilon N$, $x \varepsilon X$ say that n *is one of [the objects]* N, x *is one of* X.

We can now define

5. $X \Subset Y =_{\text{def}} \forall x(x \varepsilon X \; \rightarrow \; x \varepsilon Y)$, which says that X *are among* Y.

Then, for example, the restricted quantifier in $(\forall X \Subset N)\varphi(X)$ unpacks in the obvious way, to give us $\forall X(X \Subset N \rightarrow \varphi(X))$, saying that for any objects we take among the natural numbers, φ holds of them.

What logical laws will govern these initial plural devices? Plural quantifiers will interact with plural terms (constants and free variables) via introduction and elimination rules parallel to the laws governing the interaction of singular quantifiers and singular terms. Then arguably we will want a comprehension principle which tells us that, so long as φ is satisfied by at least one object, then there are some objects which are the φs:

$$\exists x\varphi(x) \rightarrow \exists X(x \mathrel{\varepsilon} X \leftrightarrow \varphi(x))$$

And there will be other candidate laws too. But we needn't go into more details here and now. See the readings for some options. (There is no settled 'best buy': but Linnebo's *SEP* article mentioned below gives contenders for minimal core plural logics which he calls PLO and PLO$^+$.)

(c) Using our shiny new plural notation, we can now state LNP, the Least Number Principle, in plural form as follows, with N denoting the natural numbers:

$$(\forall X \Subset N)[\exists x\, x \mathrel{\varepsilon} X \rightarrow (\exists x \mathrel{\varepsilon} X)(\forall y \mathrel{\varepsilon} X)(x \neq y \rightarrow x < y)].$$

Which is fine. But, of course, it would be more usual – *much* more usual! – to present LNP in set-theoretic guise:

$$(\forall X \subseteq \mathbb{N})[\exists x\, x \in X \rightarrow (\exists x \in X)(\forall y \in X)(x \neq y \rightarrow x < y)].$$

In *this* version $\mathbb{N}$ is a singular term denoting the set of natural numbers, and X plays the more familiar role of a typed *singular* variable running over sets.

Our plural version of LNP therefore has a direct correlate that mentions sets.[4] But this raises an immediate question: what's to choose between the ε-version and the $\in$-version?[5] If the set version is already so available, requiring no change to our logical apparatus, why not just settle for that?

Generalizing, can't we simply use ordinary logic and a modicum of set theory to regiment propositions and arguments involving plurals, without needing a special plural logic? For example, on second thoughts why can't we treat 'the Brontë sisters' and 'Anne, Charlotte and Emily' as just two different ways of picking out the same *set* (a set of three people)? – and then our inseparability inference is just a boring instance of Leibniz's Law.

(d) Or is this getting things back to front? Should we draw a different moral from the close connection between the plural and set versions of LNP and similar cases?

Recall our remarks about virtual classes, right back in §2.4. There we suggested that it seems that a good deal of elementary set talk in mathematics can be treated as just a handy façon de parler. Yes, it is a useful and familiar idiom for talking about many things at once; but in elementary contexts apparent talk about a *set of Fs* can very often be paraphrased away into direct talk about those Fs, plural, without any serious loss of relevant content.

And here we have a case in point. The useful content of the Least Number Principle is already there in the plural version; and this just goes to show that

[4] I've chosen my slightly deviant plural notation exactly to bring out the parallel.

[5] And of course we've met a closely related issue before at the end of §4.2, when we similarly wondered about the relation between a second-order version of the Induction Principle and a version explicitly written in set-theoretic terms. We'll connect the issues shortly.

the set version is overkill, importing an unnecessary commitment to additional objects, sets, over and above the numbers that are the Principle's topic. Or so the argument might go.

So which way should we jump? Do we take plurals seriously, and then perhaps use plural talk to gloss at least some low-level set talk? Or should we go the other way around, and logically tame plural claims by regimenting them into set theoretic versions? Or – an inviting option which isn't always tabled – should we be pragmatic, and let our policy vary from context to context?

(e) It's worth saying that, when we get down to details, a general strategy of systematically replacing plural referring terms with terms referring to sets is not as straightforward to implement as it might sound.

The plan, we said, is to regiment the superficially plural term in e.g.

1. The Brontë sisters were inseparable

by a singular term referring to a set. Presumably the same will go for the same plural term in

2. The Brontë sisters lived in Howarth.

In which case, the plural predicate 'lived in Howarth' will have to be rendered using a matching predicate applying not to people but to sets (with a content along the lines of 'is such that every member lived in Howarth'). But in that case, what about

3. The Brontë sisters lived in Howarth, and so did Bramwell.

It would seem *very* artificial to radically split the renditions of the two occurrences of 'lived in Howarth', rendering the plural version by a predicate which can only be sensibly applied to a set, and the singular version by a quite different predicate applying to a person.

So shall we backtrack and say that when 'The Brontë sisters' refers to a set in (1) but when it takes a non-collective predicate that can also take a singular subject, as in example (2), it needs a different treatment? Then how is this proposal supposed to work? And now what would we say about

4. The Brontë sisters lived in Howarth and were inseparable,

where the same subject term would have to get regimented in two different ways to deal with the two conjuncts? It all quickly gets a bit of mess.

Now, some proponents of plural logic make a lot of fuss about these sorts of considerations, taking them as already providing very strong grounds against a sweeping plan of trading in plural terms for terms referring to sets. But it is rather unclear quite how much weight such considerations should carry for the typical mathematical logician, who is – after all – usually not too worried about adopting somewhat procrustean formal regimentations, case by case, so long as they work for the local purposes at hand.

(f) So let's not rush to making general claims about plurals and sets, one way or the other, but rather let's consider a couple of different contexts where we encounter plurals, contexts which on reflection invite diametrically opposite treatments:

1. For a simple logical example, take the informal entailment relation in propositional logic which holds between one and more premisses (plural!) and a conclusion. When we semi-formally regiment our metalanguage, it is standard practice to officially use a two-place predicate $\vDash$ which relates a *set* of premisses on the left to a conclusion on the right. And this entirely familiar manoeuvre works fine in practice.

Now, as we mentioned in §2.4, this is just the sort of case where the talk of sets seems – strictly speaking – an *unnecessary* step, and where we could have stuck to plural talk instead (except that we don't have to hand a ready-made plural logic to handle it, if we want to semi-formalize our metalanguage). But equally, the talk of sets here seems a *harmless* step. After all, when defining formal languages even for baby propositional logic, we will have already taken on quite a lot of abstract baggage. For example, wffs are *arbitrarily long sequences of symbols* constructible according to certain rules, with instances longer than could be ever written down. It is not easy to see a sensible position which cheerfully allows us such abstract entities as these wffs but balks at very modest talk about sets of them. In short, then, the standard policy of treating $\vDash$ as relating a *set* of premisses to a conclusion allows us to draw on the benefits of a well-understood framework, without taking on extra commitments which need actually worry us in context. So why not just fall in with the standard policy here, and apply the sensible maxim "Where it doesn't itch, don't scratch"?

2. For an extreme contrasting case, take set theory, where we want to make claims such as these: the ordinals are themselves well-ordered by membership; the sets which are not self-membered are all the sets. Now we know that we can't hope to regiment *these* plural terms, the Fs, via singular terms referring to the set of Fs: for according to standard set theories there is no set of ordinals, and there is, even more famously, no set of the sets which are not self-membered. The plural terms *here* can't be regimented away as singular terms for sets.

That does leave open the possibility of construing a plural term like 'the ordinals' as a disguised singular term referring to some new sort of thing which isn't a set – perhaps a 'proper class', whatever that may be if it isn't a virtual class. But do we really want to take on a mysterious new commitment here? This time, it looks distinctly more inviting to insist that we have to take the plural term 'the ordinals' at face value, rather than trying to regiment the plural away.

This gives us, then, a first clue about who might be most interested in plural logic. It won't be the mathematical logician going about their humdrum daily business, who can and will cheerfully use a little bit of set theory when they want to talk formally about many things at once. Rather it will be theorists interested in more sweeping conceptual questions, where calibrating our required commitments can matter, e.g. as when we want to talk about things that are too many to form a set.

For a less exotic general conceptual issue, consider second-order logic again (see §§4.2, 4.4). Second-order logic and second-order theories are claimed by some to have an important foundational role. And as we saw, it's a nice question how much second-order logic can be treated non-set-theoretically, as in effect plural logic in disguise (remember Boolos!). Again, for those interested in the general project of so-called reverse mathematics (where we investigate just how strong the axioms really have to be if we are to derive e.g. standard theorems of elementary classical analysis), it will be important to see how much can be achieved using no more than the amount of set theory that can in effect be regarded as equivalent to some plural logic. And so it goes. To pursue such general conceptual questions, we will need to know more about plural logic. And we will need to convince ourselves that we aren't just temporarily putting off the set-theoretic day but can – for example – treat the semantics of plural logic in its own plural terms.

So there *is* real interest in questions about the nature and scope of plural logic here, particularly relevant to those with foundational interests.

11.6 Readings on plural logic

For a gentle and discursive introduction, see

1. Salvatore Florio and Øystein Linnebo, *The Many and the One* (OUP 2021), Chapter 2, 'Taking plurals at face value'. Available open access at tinyurl.com/flmany.

Then we have the excellent

2. Øystein Linnebo, 'Plural Quantification', *The Stanford Encyclopedia of Philosophy*, tinyurl.com/pluralq

This is particularly lucid and helpful. And from the many papers which Linnebo lists, I'd perhaps pick these classics (I mentioned the Boolos papers before in §4.4: read them now if you haven't read them before):

3. 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).

4. Alex Oliver and Timothy Smiley, 'Strategies for a logic of plurals', *Philosophical Quarterly* (2001) pp. 289–306.

Boolos's papers are influential early defences of the idea that taking plurals seriously is logically important. Oliver and Smiley forcefully argue the point that there is indeed a real topic here: you can't readily eliminate all plural talk and plural reasoning across the board in favour e.g. of singular talk and reasoning about sets (they won't approve of the pragmatic line I take at the end of the last section!).

But now where? The book on *Plural Predication* by Thomas McKay (OUP 2006) is worth reading by philosophers for its discussion of non-distributive predicates, plural descriptions etc. But for logicians, the key text has to be the philosophically argumentative, more than occasionally tendentious, but formally rich tour de force

5. Alex Oliver and Timothy Smiley, *Plural Logic* (OUP 2013: revised and expanded second edition, 2016).

However, Oliver and Smiley's eventual logical system in their Chapter 13, 'Full plural logic', will strike many as having (so to speak) unnecessarily many moving parts, as they aim – all at once – to accommodate empty domains, empty names, a plural description operator, partial functions, multivalued functions, even 'copartial functions' (which supposedly map nothing to something).

Oliver and Smiley, among others, make quite bold claims for plural logic and its relation to set theory. For a critical look at many claims of defenders of plural logic, see

5. Salvatore Florio and Øystein Linnebo, *The Many and the One* (OUP 2021), Chapter 3 onwards.

   According to the blurb, this book "provides a systematic analysis of the relation between this logic and other theoretical frameworks such as set theory, mereology, higher-order logic, and modal logic. The applications of plural logic rely on two assumptions, namely that this logic is ontologically innocent and has great expressive power. These assumptions are shown to be problematic." Open access tinyurl.com/flmany.

In particular, the authors argue that the sort of comprehension principle which is standardly built into plural logics is problematic. Florio and Linnebo propose circumscribing comprehension.

Their book is approachable and argumentative. I in fact think some of Florio and Linnebo's arguments are resistible: see my comments on the first two parts of the book, tinyurl.com/many-one. But well worth reading for an entrée to a number of current debates.

12 Going further

This has been a Guide to *beginning* mathematical logic. So far, then, the suggested readings on different areas have been at entry level, or only a step or so up from that. In this final chapter, by contrast, we take a look at some of the more somewhat advanced literature on a selection of topics, taking us another step or two further.

If you have been tackling enough of the introductory readings, you should in fact be able to now follow your interests wherever they lead, without really needing help from this chapter. For a start, you can explore the many mathematical logic entries in *The Stanford Encyclopedia of Philosophy*, which are mostly excellent and have large bibliographies. The substantial essays in the eighteen(!) volumes of *The Handbook of Philosophical Logic* are of varying quality, but there are some good ones on straight mathematical logic topics, again with large bibliographies. Internet sites like math.stackexchange.com and the upper-level mathoverflow.net can be searched for useful lists of recommended books. And then there is always Google!

However, those resources do cumulatively point to a rather overwhelming range of literature to pursue. So perhaps some readers will still appreciate a few more limited menus of suggestions (even if they are considerably less systematic and more shaped by my personal interests than in the core Guide).

Of course, the 'vertical' divisions between entry-level coverage and the further explorations in this chapter are pretty arbitrary; and the 'horizontal' divisions into different subfields can in places also be quite blurred. But we do need to impose *some* organization! So this chapter is divided up as follows. First, we make a very brief foray into logic-relevant algebra:

12.1 A very little light algebra for logic?

There follows a series of sections taking up the core topics of Chapters 5–7 and 9 in the same order as before:

12.2 More model theory
12.3 More on formal arithmetic and computability
12.4 More on mainstream set theory
12.5 Choice, and the choice of set theory
12.6 More proof theory.

Then there is a final section which introduces a further topic area which is the focus of considerable recent interest:

12.7 Higher-order logic, the lambda calculus, and type theory.

We could continue; but this is more than enough to be going on with . . . !

12.1 A very little light algebra for logic?

Depending on what you have read on classical propositional logic, you may well have touched on the notion of a Boolean algebra. And depending on what you have read on intuitionistic logic, you may have also also encountered Heyting algebras (a.k.a. pseudo-Boolean algebras). It is worth getting to know a bit more about these algebras, both because of their relevance to classical and intuitionistic logic, but also because Boolean algebra features in independence arguments in set theory.

For a gentle and clear first introduction (aimed at those with little mathematical background), see

1. Barbara Hall Partee, Alice G. B. ter Meulen, and Robert Eugene Wall, *Mathematical Methods in Linguistics* (1990, Springer). The (short!) Chs 9 and 10 introduce some basic concepts of algebra (you can omit §10.3); Ch. 11 is on lattices; Ch. 12 is then on Boolean and Heyting algebras, and briefly connects Kripke's relational semantics for intuitionistic logic to Heyting algebras.

Also very accessible, adding a little more on Heyting algebras:

2. Morten Heine Sørensen and Pawel Urzyczyn, *Lectures on the Curry-Howard Isomorphism* (Elsevier, 2006), Ch. II, 'Intuitionistic logic'.

Then, for rather more about Boolean algebras, you need *very* little background to start tackling the opening chapters of

3. Steven Givant and Paul Halmos, *Introduction to Boolean Algebras* (Springer, 2009). This is an update of a classic book by Halmos, and is very accessible; any logician will want eventually to know the elementary material in the first third of the book.

If you already know a smidgin of algebra and topology, however, then there is a faster-track introduction to Boolean algebras in

4. René Cori and Daniel Lascar, *Mathematical Logic, A Course with Exercises: Part I* (OUP, 2000), Chapter 2.

And for a higher-level treatment of intuitionistic logic and Heyting algebras, you could read Chapter 5 of the book by Dummett mentioned in §8.5, or work up to Chapter 7 on algebraic semantics in the book on modal logic by Chagrov and Zakharyaschev mentioned in §10.5.

Then, if you want to pursue more generally e.g. questions about when propositional logics do have nice algebraic counterparts (in the sort of way that classical and intuitionistic logic relate respectively to Boolean and Heyting Algebras), then you *might* get something out of Ramon Jansana's 'Algebraic propositional logic' in *The Stanford Enclyclopedia of Philosophy*, tinyurl.com/alg-logic. But this does strike me as too rushed to be particularly useful. So instead, you could make a start reading

5. Josep Maria Font, *Abstract Algebraic Logic: An Introductory Textbook* (College Publications, 2016). This is attractively written in an expansive and accessible style, and is well worth diving into.

12.2 More model theory

(a) If you want to explore beyond the entry-level material of Chapter 5 on model theory, why not start with a quick warm-up, with some reminders of headlines and some very useful pointers to the road ahead:

1. Wilfrid Hodges and Thomas Scanlon, 'First-order model theory', *The Stanford Encyclopedia of Philosophy*, tinyurl.com/sep-fo-model.

Now, we noted in §3.7(c) and §5.3 that the wide-ranging mathematical logic texts by Hedman and Hinman cover a substantial amount of model theory. But why not look at two classic stand-alone treatments of the area which really choose themselves? In order of both first publication and eventual difficulty:

2. C. Chang and H. J. Keisler, *Model Theory** (originally North Holland 1973: the third edition has been inexpensively republished by Dover Books in 2012). This is the Old Testament, the first systematic text on model theory. Over 550 pages long, it proceeds at an engagingly leisurely pace. It is particularly lucid and is extremely nicely constructed with different chapters on different methods of model-building. A really fine achievement that I think *still* remains a good route in to the serious study of model theory.
3. Wilfrid Hodges, *A Shorter Model Theory* (CUP, 1997). The New Testament is Hodges's encyclopedic *Model Theory* (CUP 1993). This shorter version is half the size but still really full of good things. It does get tougher as the book progresses, but the earlier chapters of this modern classic, written with this author's characteristic lucidity, should certainly be readily manageable.

More specifically, my suggestion would be to read the first three long chapters of Chang and Keisler, and then perhaps pause to make a start on

4. J. L. Bell and A. B. Slomson, *Models and Ultraproducts** (North-Holland 1969; Dover reprint 2006). Very elegantly put together: as the title suggests, the book focuses particularly on the ultra-product construction.

At this point read the first five chapters for a particularly clear introduction.

You could then return to Ch. 4 of C&K to look at (some of) their treatment of the ultra-product construction, before perhaps putting the rest of their book on hold and turning to Hodges.

(b) A level up again, here are two further books that should definitely be mentioned. The first has been around long enough to have become regarded as a modern standard text. The second is a bit more recent but also comes widely recommended. Their coverage is significantly different – so I suppose that those wanting to get really seriously into model theory should take a look at both:

5. David Marker, *Model Theory: An Introduction* (Springer 2002). Despite its title, this book would surely be hard going if you haven't already tackled some model theory (at least read Manzano or Kirby first). But despite being sometimes a rather bumpy ride, this highly regarded text will teach you a great deal. Later chapters, however, probably go far over the horizon for all except those most enthusiastic readers of this Guide who are beginning to think about specializing in model theory – it isn't published in the series 'Graduate Texts in Mathematics' for nothing!
6. Katrin Tent and Martin Ziegler, *A Course in Model Theory* (CUP, 2012). From the blurb: "This concise introduction to model theory begins with standard notions and takes the reader through to more advanced topics such as stability .... The authors introduce the classic results, as well as more recent developments in this vibrant area of mathematical logic. Concrete mathematical examples are included throughout to make the concepts easier to follow." Again, although it starts from the beginning, it could be a challenge to readers without some mathematical sophistication and some prior exposure to the elements of model theory – though I, for one, find it more approachable than Marker's book.

(c) So much for my principal suggestions. Now for an assortment of additional/alternative texts. Here are two more books which aim to give general introductions:

7. Philipp Rothmaler's *Introduction to Model Theory* (Taylor and Francis 2000) is, overall, comparable in level of difficulty with, say, the first half of Hodges. As the blurb puts it: "This text introduces the model theory of first-order logic, avoiding syntactical issues not too relevant to model theory. In this spirit, the compactness theorem is proved via the algebraically useful ultraproduct technique (rather than via the completeness theorem of first-order logic). This leads fairly quickly to algebraic applications, ..." Now, the opening chapters are very clear: but oddly the introduction of the crucial ultraproduct construction in Ch. 4 is done very briskly (compared, say, with Bell and Slomson). And thereafter it seems to me that there is some unevenness in the accessibility of the

book. But others have recommended this text more warmly, so I mention it as a possibility worth checking out.

8. Bruno Poizat's *A Course in Model Theory* (English edition, Springer 2000) starts from scratch and the early chapters give an interesting and helpful account of the model-theoretic basics, and the later chapters form a rather comprehensive introduction to stability theory. This often-recommended book is written in a rather distinctive style, with rather more expansive class-room commentary than usual: so an unusually engaging read at this sort of level.

Another book which is often mentioned in the same breath as Poizat, Marker, and now Tent and Ziegler is *A Guide to Classical and Modern Model Theory*, by Annalisa Marcja and Carlo Toffalori (Kluwer, 2003) which also covers a lot: but I prefer the previously listed books.

The next two suggestions are of books which are helpful on particular aspects of model theory:

9. Kees Doets's short *Basic Model Theory** (CSLI 1996) highlights so-called Ehrenfeucht games. This is enjoyable and very instructive.
10. Chs 2 and 3 of Alexander Prestel and Charles N. Delzell's *Mathematical Logic and Model Theory: A Brief Introduction* (Springer 1986, 2011) are brisk but clear, and can be recommended if you want a speedy review of model theoretic basics. The key feature of the book, however, is the sophisticated final chapter on serious applications to algebra, which might appeal to mathematicians with interests in that area.

Indeed, as we explore model theory, we quickly get entangled with algebraic questions. And as well as going (so to speak) in the direction from logic to algebra, we can make connections the other way about, starting from algebra. For something on this approach, see the following short, relatively accessible, and illuminating book:

11. Donald W. Barnes and John M. Mack, *An Algebraic Introduction to Mathematical Logic* (Springer, 1975).

(d) As an aside, let me also mention the sub-area of Finite Model Theory which arises particularly from consideration of problems in the theory of computation (where, of course, we are interested in *finite* structures – e.g. finite databases and finite computations over them). What happens, then, to model theory if we restrict our attention to finite models? Trakhtenbrot's theorem, for example, tells that the class of sentences true in any finite model is not recursively enumerable. So there is no deductive theory for capturing such finitely valid sentences (that's a surprise, given that there's a complete deductive system for the sentences which are valid in the usual broader sense). It turns out, then, that the study of finite models is surprisingly rich and interesting. So why not dip into one or other of

12. Leonid Libkin, *Elements of Finite Model Theory* (Springer 2004).

13. Heinz-Dieter Ebbinghaus and Jörg Flum, *Finite Model Theory* (Springer 2nd edn. 1999).

Both are good, though I prefer Libkin.

(e) In §5.3 I warmly recommended that you read at least early chapters of *Philosophy and Model Theory* by Button and Walsh. Now you know more model theory, do revisit that book and read on.

Finally, I suppose I should mention John T. Baldwin's *Model Theory and the Philosophy of Mathematical Practice* (CUP, 2018). This presupposes a lot more background than Button and Walsh. Maybe some philosophers might be able to excavate more out of Baldwin's book than I did: but I find this book badly written and unnecessarily hard work.

12.3 More on formal arithmetic and computability

(a) The readings in §6.5 have introduced you to the canonical first-order theory of arithmetic, first-order Peano Arithmetic, as well as to some subsystems of PA (in particular, Robinson Arithmetic) and second-order extensions. So what to read next on formal arithmetics?

You will know by now that first-order PA has non-standard models: in fact, it even has uncountably many non-isomorphic models which can be built just out of natural numbers. It is worth pursuing this theme. For a taster, you could look at lecture notes by Jaap van Oosten, on 'Introduction to Peano Arithmetic: Gödel Incompleteness and Nonstandard Models', tinyurl.com/oosten-peano. But better to dive into

1. Richard Kaye's *Models of Peano Arithmetic* (Oxford Logic Guides, OUP, 1991), which tells us a great deal about non-standard models of PA. This reveals more about what PA can and can't prove, and will also introduce you to some non-Gödelian examples of incompleteness. This is a terrific book, and deservedly a modern classic.

As a sort of sequel, there is also another volume in the Oxford Logic Guides series for enthusiasts with more background in model theory, namely Roman Kossak and James Schmerl, *The Structure of Models of Peano Arithmetic*, OUP, 2006. But this is much tougher going. For a more accessible set of excellent lecture notes, see

2. Tin Lok Wong, 'Model theory of arithmetic', downloadable lecture by lecture from tinyurl.com/wong-model.

Next, going in a rather different direction, and explaining a lot about arithmetics weaker than full PA, here's another modern classic:

3. Petr Hájek and Pavel Pudlák, *Metamathematics of First-Order Arithmetic* (Springer 1993). This is pretty encyclopaedic, but at least the first three chapters do remain surprisingly accessible for such a work. This

is, eventually, a must-read if you have a serious interest in theories of arithmetic and incompleteness.

And what about going beyond first-order PA? We know that full second-order PA (where the second-order quantifiers are constrained to run over *all* possible sets of numbers) is unaxiomatizable, because the underlying second-order logic is unaxiomatizable. But there are axiomatizable subsystems of second-order arithmetic. These are wonderfully investigated in another encyclopaedic modern classic:

4. Stephen Simpson, *Subsystems of Second-Order Arithmetic* (Springer 1999; 2nd edn CUP 2009). The focus of this book is the project of 'reverse mathematics' (as it has become known): that is to say, the project of identifying the weakest theories of numbers-and-sets-of-numbers that are required for proving various characteristic theorems of classical mathematics.

   We know that we can reconstruct classical analysis in pure set theory, and rather more neatly in set theory with natural numbers as unanalysed urelements. But just *how much* set theory is needed to do the job, once we have the natural numbers? The answer is: stunningly little. The project of exploring what's needed is introduced very clearly and accessibly in the first chapter, which is a must-read for anyone interested in the foundations of mathematics. This introduction is freely available at the book's website tinyurl.com/2arith.

(b) Next, Gödelian incompleteness again. You could start with a short old *Handbook* article which is still well worth reading:

5. Craig Smoryński, 'The incompleteness theorems', in J. Barwise, editor, *Handbook of Mathematical Logic*, pp. 821–865 (North-Holland, 1977), which covers a lot very compactly. Available at tinyurl.com/smory.

Now, the further readings on incompleteness suggested in §6.6 finished by mentioning two wonderful books which could arguably have appeared on our main list of introductory readings. However – a judgement call – I suggested that the more abstract stories they tell can probably only be fully appreciated if you've first met the basics of computability theory and the incompleteness theorems in a more conventional treatment. But certainly, now is the time to read them, if you didn't tackle them before:

6. Raymond Smullyan, *Gödel's Incompleteness Theorems*, Oxford Logic Guides 19 (Clarendon Press, 1992). Proves beautiful, slightly abstract, versions of the incompleteness theorems. A modern classic.

7. Equally short and equally elegant is Melvin Fitting's, *Incompleteness in the Land of Sets** (College Publications, 2007). There is a simple correspondence between natural numbers and 'hereditarily finite sets' (i.e. sets which have a finite number of members which in turn have a finite

number of members which in turn ... where all downward membership chains bottom out with the empty set). Relying on this fact gives us another route in to proofs of Gödelian incompleteness, and other results of Church, Rosser and Tarski. Beautifully done.

After these, where should you go if you want to know more about matters more or less directly to do with the incompleteness theorems? Here are some resources, in order of publication date:

8. Craig Smoryński, *Logical Number Theory I, An Introduction* (Springer, 1991). There are three long chapters. Ch. I discusses pairing functions and numerical codings, primitive recursion, the Ackermann function, computability, and more. Ch. II concentrates on 'Hilbert's tenth problem' – showing that we can't mechanically decide the solubility of certain equations. Ch. III considers Hilbert's Programme and contains proofs of more decidability and undecidability results, leading up to a version of Gödel's First Incompleteness Theorem. (The promised Vol. II which would have discussed the Second Incompleteness Theorem has never appeared.)

   The level of difficulty is rather varied, and there are a lot of historical disgressions and illuminating asides. So this is an idiosyncratic book; but is still an enjoyable and very instructive read.

9. Raymond Smullyan's *Diagonalization and Self-Reference*, Oxford Logic Guides 27 (Clarendon Press 1994) is an investigation-in-depth around and about the idea of diagonalization that figures so prominently in proofs of limitative results like the unsolvability of the halting problem, the arithmetical undefinability of arithmetical truth, and the incompleteness of arithmetic. Read at least Part I.

10. Per Lindström, *Aspects of Incompleteness* (Association for Symbolic Logic/ A. K. Peters, 2nd edn., 2003). This rather terse book is probably for enthusiasts. It is not always reader-friendly in its choices of notation and the brevity of its arguments. However, the more mathematical reader will find that it again repays the effort.

11. Torkel Franzén, *Inexaustibility: A Non-exhaustive Treatment* (Association for Symbolic Logic/A. K. Peters, 2004). I recommended most of this book in §6.6. The final chapters interestingly discuss what happens if we extend PA by adding $\mathsf{Con}_{\mathrm{PA}}$ – the arithmetic sentence expressing the consistency of PA – as a new axiom, and then add the consistency sentence for this expanded theory, and then add the consistency sentence for *that* theory, and keep on going ....

12. Wolfgang Rautenberg, *A Concise Introduction to Mathematical Logic* (Springer, 2nd edn. 2006). Chapters 6 and 7 are a compressed but rather elegant discussion of incompleteness, undecidability, and self-reference. Rautenberg does the detailed work in deriving the HBL derivability con-

ditions and so fully proving the second incompleteness theorem. There is also a discussion of provability logic. Excellent.

And if you want the bumpier ride of a lecture course with problems assigned as you go along, this is notable:

13. Tin Lok Wong, 'The consistency of arithmetic', downloadable lecture by lecture from tinyurl.com/wong-consis.

(c) Now let's turn to books on computability. Among the Big Books on mathematical logic, the one with the most useful treatment is probably

14. Peter G. Hinman, *Fundamentals of Mathematical Logic* (A. K. Peters, 2005). Chs 4 and 5 on recursive functions, incompleteness etc. strike me as the best written, most accessible (and hence most successful) chapters in this very substantial book. The chapters could well be read after my *IGT* as somewhat terse revision for mathematicians, and then as sharpening the story in various ways. Ch. 8 then takes up the story of recursion theory (the author's home territory).

However, good those these chapters are, I'd still recommend starting your more advanced work on computability with

15. Nigel Cutland, *Computability: An Introduction to Recursive Function Theory* (CUP 1980). This is a rightly much-reprinted classic and is beautifully lucid and well-organized. This *does* have the look-and-feel of a traditional maths textbook of its time (so perhaps with fewer of the classroom asides we find in some modern, more discursive books). However, if you got through most of e.g. Boolos and Jeffrey without too much difficulty, you ought certainly to be able to tackle this as the next step. Very warmly recommended.

And of more recent books covering computability at this level, I particularly like

16. S. Barry Cooper, *Computability Theory* (Chapman & Hall/CRC 2003). A very nicely done modern textbook. Read at least Part I of the book (about the same level of sophistication as Cutland, but with some extra topics), and then you can press on as far as your curiosity takes you, and get to excitements like the Friedberg-Muchnik theorem.

By contrast, I found Robert I. Soare's densely written *Turing Computability: Theory and Applications* (Springer 2016) a very much less attractive proposition.

Of course, the inherited literature on computability is huge. But, being *very* selective, let me mention three classics from different generations:

17. Rósza Péter, *Recursive Functions* (originally published 1950: English translation Academic Press 1967). This is by one of those logicians who was 'there at the beginning'. It has that old-school slow-and-steady unflashy lucidity that makes it still a considerable pleasure to read. It remains very worth looking at.

18. Hartley Rogers, Jr., *Theory of Recursive Functions and Effective Computability* (McGraw-Hill 1967) is a heavy-weight state-of-the-art-then classic, written at the end of the glory days of the initial development of the logical theory of computation. It quite speedily gets advanced. But the action-packed opening chapters are excellent. At least take it out of the (e)library, read a few chapters, and admire!
19. Piergiorgio Odifreddi, *Classical Recursion Theory*, Vol. 1 (North Holland, 1989) is well-written and discursive, with numerous interesting asides. It's over 650 pages long, so it goes further and deeper than other books on the main list above (and then there is Vol. 2). But it certainly starts off quite gently paced and very accessible and can be warmly recommended for consolidating and then extending your knowledge.

(d) Classical computability theory abstracts away from considerations of practicality, efficiency, etc. Computer scientists are – surprise, surprise! – interested in the theory of feasible computation, and any logician should be interested in finding out at least a little about the topic of computational complexity. Here are three introductions to the topic, in order of increasing detail:

20. Herbert E. Enderton, *Computability Theory: An Introduction to Recusion Theory* (Associated Press, 2011). Chapter 7.
21. Shawn Hedman *A First Course in Logic* (OUP 2004): Ch. 7 on 'Computability and complexity' has a nice review of basic computability theory before some lucid sections discussing computational complexity.
22. Michael Sipser, *Introduction to the Theory of Computation* (Thomson, 2nd edn. 2006) is a standard and very well regarded text on computation aimed at computer scientists. It aims to be very accessible and to take its time giving clear explanations of key concepts and proof ideas. I think this is very successful as a general introduction and I could well have mentioned it before. But I'm highlighting this book now because its last third is on computational complexity.

And for rather more expansive, stand-alone treatments, here are three more suggestions:

23. I don't mention many sets of lecture notes in this Guide, as they tend to be rather too terse for self-study. But Ashley Montanaro has an excellent and extensive lecture notes on *Computational Complexity*, lucid and detailed. Available at tinyurl.com/cocomp.
24. Oded Goldreich, *P, NP, and NP-Completeness* (CUP, 2010). Short, clear, and introductory stand-alone treatment.
25. You could also look at the opening chapters of the pretty encyclopaedic Sanjeev Arora and Boaz Barak *Computational Complexity: A Modern Approach* (CUP, 2009). The authors say that '[r]equiring essentially no background apart from mathematical maturity, the book can be used

as a reference for self-study for anyone interested in complexity, including physicists, mathematicians, and other scientists, as well as a textbook for a variety of courses and seminars.' And at least it starts very readably! A late draft of the book can be freely downloaded from tinyurl.com/arora.

12.4 More on mainstream set theory

(a) Some of the readings on set theory suggested in Chapter 7 were beginning to get quite sophisticated: but still, we weren't tangling with more advanced topics like 'large cardinals' and 'forcing'. Now we move on.

And one option is immediately to go for broke and dive in to the modern bible, which is highly impressive not just for its size:

1. Thomas Jech, *Set Theory*, The Third Millennium Edition (Springer, 2003). The book is in three parts: the first, Jech says, every student should know; the second part every budding set-theorist should master; and the third consists of various results reflecting 'the state of the art of set theory at the turn of the new millennium'. Start at page 1 and keep going to page 705 – or until you feel glutted with set theory, whichever comes first!

This book is a masterly achievement by a great expositor. And if you've happily read e.g. the introductory books by Enderton and then Moschovakis mentioned earlier in the Guide, then you should be able to cope pretty well with Part I of the book while it pushes on the story a little with some material on 'small large cardinals' and other topics. Part II of the book starts by telling you about independence proofs. In particular, the Axiom of Choice is consistent with ZF and the Continuum Hypothesis is consistent with ZFC, as proved by Gödel using the idea of 'constructible' sets. While the Axiom of Choice is independent of ZF, and the Continuum Hypothesis is independent with ZFC, as proved by Cohen using the much more tricky but extraordinarily prolific technique of 'forcing'. The rest of Part II tells you more about large cardinals, and about descriptive set theory. Part III is for enthusiasts.

(b) Now, Jech's book is wonderful, but let's face it, the sheer size makes it a trifle daunting. It goes quite a bit further than many will need, and to get there it occasionally speeds along a bit faster than some will feel comfortable with. So what other options are there for if you want to take things more slowly?

Let's start with a book which I mentioned in passing in §7.7:

2. Azriel Levy, *Basic Set Theory** (Springer 1979, republished by Dover 2002). This is 'basic' in the sense of not dealing with topics like forcing. However it *is* a quite advanced-level treatment of the set-theoretic fundamentals at least in its mathematical style, and even the earlier parts are I think best tackled once you know some set theory (they could be very useful, though, as a rigorous treatment consolidating the basics –

a reader comments that Levy's is his "go to" book when he needs to check set theoretical facts that don't involve forcing or large cardinals.). The last part of the book starts on some more advanced topics.

Levy's book ends with a discussion of some 'large cardinals'. However another much admired older book remains the recommended first treatment of this topic:

3. Frank R. Drake, *Set Theory: An Introduction to Large Cardinals* (North-Holland, 1974). This overlaps with Part I of Jech's bible, though at perhaps a gentler pace. But it also will tell you about Gödel's Constructible Universe and then some more about large cardinals. Very lucid.

For some other topics you could also look at the second volume of a book whose first instalment was a main recommendation in §7.3:

4. Winfried Just and Martin Weese, *Discovering Modern Set Theory II: Set-Theoretic Tools for Every Mathematician* (American Mathematical Society, 1997).

   This contains, as the authors put it, "short but rigorous introductions to various set-theoretic techniques that have found applications outside of set theory". Some interesting topics, and can be read independently of Vol. I.

(c) But now the crucial next step – that perhaps marks the point where set theory gets challenging – is to get your head around Cohen's idea of forcing used in independence proofs. However, there is not getting away from it, this is tough. In the admirable

5. Timothy Y. Chow, 'A beginner's guide to forcing', tinyurl.com/chowf

Chow writes:

> All mathematicians are familiar with the concept of *an open research problem*. I propose the less familiar concept of *an open exposition problem*. Solving an open exposition problem means explaining a mathematical subject in a way that renders it totally perspicuous. Every step should be motivated and clear; ideally, students should feel that they could have arrived at the results themselves. The proofs should be 'natural' . . . [i.e., lack] any ad hoc constructions or brilliancies. I believe that it is an open exposition problem to explain forcing.

In short: if you find that expositions of forcing – including Chow's – tend to be hard going, then join the club.

Here though is a very widely used and much reprinted textbook, which nicely complements Drake's book and which has (inter alia) a relatively approachable introduction to forcing arguments:

6. Kenneth Kunen, *Set Theory: An Introduction to Independence Proofs* (North Holland, 1980). If you have read (some of) the introductory set theory books mentioned in the Guide, you should actually find much of

this text now pretty accessible, and can probably speed through some of the earlier chapters, slowing down later, until you get to the penultimate chapter on forcing which you'll need to take slowly and carefully. This is a rightly admired classic text.

Kunen has since published another, totally rewritten, version of this book as *Set Theory** (College Publications, 2011). This later book is quite significantly longer, covering an amount of more difficult material that has come to prominence since 1980. Not just because of the additional material, my current sense is that the earlier book may remain the somewhat gentler read.

Now, Kunen's classic text takes a 'straight down the middle' approach, starting with what is basically Cohen's original treatment of forcing, though he does relate this to some other approaches. Here are two of them:

7. Raymond Smullyan and Melvin Fitting, *Set Theory and the Continuum Problem* (OUP 1996, Dover Publications 2010). This medium-sized book is divided into three parts. Part I is a nice introduction to axiomatic set theory (in fact, officially in its NBG version – see §12.5). The shorter Part II concerns matters round and about Gödel's consistency proofs via the idea of constructible sets. Part III gives a different take on forcing. This is beautifully done, as you might expect from two writers with a quite enviable knack for wonderfully clear explanations and an eye for elegance.

8. Keith Devlin, *The Joy of Sets* (Springer 1979, 2nd edn. 1993) Ch. 6 introduces the idea of Boolean-Valued Models and their use in independence proofs. The basic idea is fairly easily grasped, but the details perhaps trickier.

   For more on this theme, see John L. Bell's classic *Set Theory: Boolean-Valued Models and Independence Proofs* (Oxford Logic Guides, OUP, 3rd edn. 2005). The relation between this approach and other approaches to forcing is discussed e.g. in Chow's paper and the last chapter of Smullyan and Fitting.

(d) Here is a selection of another three books with various virtues, in order of publication:

9. Akihiro Kanamori, *The Higher Infinite: Large Cardinals in Set Theory from Their Beginnings* (Springer, 1997, 2nd edn. 2003). This blockbuster is subtitled 'Large Cardinals in Set Theory from Their Beginnings', and is very clearly put together with a lot of helpful and illuminating historical asides. A classic.

10. Lorenz J. Halbeisen, *Combinatorial Set Theory, With a Gentle Introduction to Forcing* (Springer 2011). From the blurb "This book provides a self-contained introduction to modern set theory and also opens up some more advanced areas of current research in this field. The first part offers an overview of classical set theory wherein the focus lies on

the axiom of choice and Ramsey theory. In the second part, the sophisticated technique of forcing, originally developed by Paul Cohen, is explained in great detail. With this technique, one can show that certain statements, like the continuum hypothesis, are neither provable nor disprovable from the axioms of set theory. In the last part, some topics of classical set theory are revisited and further developed in the light of forcing."

True, this book gets quite hairy towards the end: but the earlier parts of the book should be much more accessible. This book has been strongly recommended for its expositional merits by more reliable judges than me; but I confess I didn't find it notably more successful than other accounts of forcing. A late draft is available: tinyurl.com/halb-set.

11. Nik Weaver, *Forcing for Mathematicians* (World Scientific, 2014) is less than 150 pages (and the first applications of the forcing idea appear after just 40 pages: you don't have to read the whole book to get the basics). From the blurb: "Ever since Paul Cohen's spectacular use of the forcing concept to prove the independence of the continuum hypothesis from the standard axioms of set theory, forcing has been seen by the general mathematical community as a subject of great intrinsic interest but one that is technically so forbidding that it is only accessible to specialists ... This is the first book aimed at explaining forcing to general mathematicians. It simultaneously makes the subject broadly accessible by explaining it in a clear, simple manner, and surveys advanced applications of set theory to mainstream topics." This does strike me as a helpful attempt to solve Chow's basic exposition problem, to explain the Big Ideas very directly.

I did have hopes for Mirna Džamonja's *Fast Track to Forcing* (LMS Student Texts, CUP 2021), which certainly aims to be accessible to a likely reader of this Guide: but I'd say the book fails to fulfil its brief.

12.5 Choice, and the choice of set theory

But now let's leave the Higher Infinite and other excitements and get back down to earth, or at least to less exotic topics! And, to return to the beginning, we might wonder: is ZFC the 'right' set theory? How do we choose which set theory to adopt?

(a) Let's start by thinking about the Axiom of Choice in particular. It is comforting to know from Gödel that AC is consistent with ZF (so adding it doesn't lead to contradiction). But we also know from Cohen's forcing argument that AC is independent with ZF (so accepting ZF doesn't commit you to accepting AC too). So why buy AC? Is it an optional extra?

Quite a few of the readings already mentioned will have touched on the question of AC's status and role. But for a useful overview/revision of some basics, see

1. John L. Bell, 'The axiom of choice', *The Stanford Encyclopedia of Philosophy*, tinyurl.com/sep-axch.

And for a short book also explaining some of the consequences of AC (and some of the results that you need AC to prove), see

2. Horst Herrlich, *Axiom of Choice* (Springer 2006), which has chapters really rather tantalizingly entitled 'Disasters without Choice', 'Disasters with Choice' and 'Disasters either way'.

Herrlich perhaps already tells you more than enough about the impact of AC: but there's also a famous book by H. Rubin and J.E. Rubin, *Equivalents of the Axiom of Choice* (North-Holland 1963; 2nd edn. 1985) worth browsing through: it gives over two hundred equivalents of AC!

Then next there is the nice short classic

3. Thomas Jech, *The Axiom of Choice** (North-Holland 1973, Dover Publications 2008). This proves the Gödel and Cohen consistency and independence results about AC (without bringing into play everything needed to prove the parallel results about the Continuum Hypothesis). In particular, there is a nice presentation of the so-called Fraenkel-Mostowski method of using 'permutation models'. Then later parts of the book tell us something about mathematics without choice, and about alternative axioms that are inconsistent with choice.

And for a more recent short book, taking you into new territories (e.g. making links with category theory), enthusiasts might enjoy

4. John L. Bell, *The Axiom of Choice** (College Publications, 2009).

(b) From earlier reading you should certainly have picked up the idea that, although ZFC is the canonical modern set theory, there are other theories on the market. I mention just a selection here (I'm certainly not suggesting you need to follow up all these pointers – but it is worth stressing again that set theory is not quite the monolithic edifice that some presentations might suggest).

For a brisk overview, putting many of the various set theories we'll consider below into some sort of order, and mentioning yet further alternatives, see

5. M. Randall Holmes, 'Alternative axiomatic set theories', *The Stanford Encyclopedia of Philosophy*, tinyurl.com/alt-set.

At this stage, you might well find this a bit *too* brisk and allusive, but it is useful to give you a preliminary sense of the range of possibilities here. And I should mention that there is a longer version of this essay which you can return to later:

6. M. Randall Holmes, Thomas Forster and Thierry Libert. 'Alternative set theories'. In Dov Gabbay, Akihiro Kanamori, and John Woods, eds. *Handbook of the History of Logic, vol. 6, Sets and Extensions in the Twentieth Century*, pp. 559-632. (Elsevier/North-Holland 2012).

(c) It quickly becomes clear that some alternative set theories are more alternative than others! So let's start with the one which is the closest sibling to standard ZFC, namely NBG. You will have very probably come across mention of this already (e.g. even in the early pages of Enderton's set theory book).

We know that the universe of sets in ZFC is not itself a set. But we might think that this universe is a *sort* of big collection. Should we explicitly recognize, then, two sorts of collection, sets and (as they are called in the trade) proper classes which are too big to be sets? Some standard presentations of ZFC, such as Kunen's, do in fact introduce symbolism for classes, but then make it clear that class-talk is just a useful short-hand that can be translated away. NBG (named for von Neumann, Bernays, Gödel: a few say VBG) takes classes a bit more seriously. But things are a little delicate: it is a nice question just what NBG commits us to. An important technical feature is that its principle of class comprehension is 'predicative'; i.e. quantified variables in the defining formula for a class can't range over proper classes but range only over sets. Because of this we get a conservative extension of ZFC (nothing in the language of sets can be proved in NBG which can't already be proved in ZFC). For more, see:

7. Abraham Fraenkel, Yehoshua Bar-Hillel and Azriel Levy, *Foundations of Set-Theory* (North-Holland, 2nd edition 1973). Their Ch. II §7 remains a classic general discussion of the role of classes in set theory.

And also worth quickly consulting is

8. Michael Potter, *Set Theory and Its Philosophy* (OUP 2004) Appendix C is a brisker account of NBG and of other theories with classes as well as sets, such as MK, Morse-Kelley set theory.

Then, if you want detailed presentations of set-theory via NBG, you can see either or both of

9. Elliott Mendelson, *Introduction to Mathematical Logic* (CRC, 4th edition 1997), Ch.4. is a classic and influential textbook presentation.
10. Raymond Smullyan and Melvin Fitting, *Set Theory and the Continuum Problem* (OUP 1996, Dover Publications 2010), Part I is another development of set theory in its NBG version.

(d) Recall, earlier in the Guide, we very warmly recommended Michael Potter's book which we just mentioned again. This presents a version of an axiomatization of set theory due to Dana Scott (hence 'Scott-Potter set theory', SP). This axiomatization is consciously guided by the conception of the set theoretic universe as built up in levels (the conception that, supposedly, also warrants the axioms of ZF). What Potter's book aims to reveal is that we can get a rich hierarchy of sets, more than enough for mathematical purposes, without committing ourselves to *all* of ZFC (whose extreme richness comes from the full Axiom of Replacement). If you haven't read Potter's book before, now is the time to look at it. Alternatively for a slightly simplified presentation of SP, see

11. Tim Button, ‘Level Theory, Part I’, *Bulletin of Symbolic Logic*, preprint available at tinyurl.com/level-th.

(e) We now turn to a somewhat more radical departure from standard ZF(C), namely ZFA (which is, in a sense to be explained, ZF − AF + AFA).

Here again is the now-familiar hierarchical conception of the set universe: We start with some non-sets (maybe zero of them in the case of pure set theory). We collect them into sets (as many different ways as we can). Now we collect what we've already formed into sets (as many as we can). Keep on going, as far as we can. On this ‘bottom-up’ picture AF, the Axiom of Foundation, is compelling (that's the axiom that any downward chain linked by set-membership will bottom out, and won't go round in a circle).

But here's another alternative conception of the set universe. Think of a set as a gadget that points you at some things, its members. And those members, if sets, point to *their* members. And so on and so forth. On this ‘top-down’ picture, the Axiom of Foundation is not so compelling. As we follow the pointers, can't we for example come back to where we started? It is well known that in much of the usual development of ZFC the Axiom of Foundation AF does little work. So what about considering a theory of sets ZFA which drops AF and instead has an Anti-Foundation Axiom, AFA, which allows self-membered sets? To explore this idea, see

12. Start with Lawrence S. Moss, ‘Non-wellfounded set theory’, *The Stanford Encyclopedia of Philosophy*, tinyurl.com/sep-zfa.
13. Keith Devlin, *The Joy of Sets* (Springer, 2nd edn. 1993), Ch. 7. The last chapter of Devlin's book, added in the second edition of his book, starts with a very lucid introduction, and develops some of the theory.
14. Peter Aczel, *Non-well-founded Sets* (CSLI Lecture Notes 1988). This is a very readable short classic book, available at tinyurl.com/aczel.
15. Luca Incurvati, ‘The graph conception of set’ *Journal of Philosophical Logic* (2014) pp. 181-208, or his *Conceptions of Set and the Foundations of Mathematics* (CUP, 2020), Ch. 7, very illuminatingly explores the motivation for such set theories.

(f) Now for a much more radical departure from ZF.

Standard set theory lacks a universal set because, together with other standard assumptions, the idea that there is a set of all sets leads to contradiction. But by tinkering with those other assumptions, there are coherent theories with universal sets, of which Quine's ‘New Foundations’ is the probably the best known. For the headline news, see

16. T. F. Forster, ‘Quine's New Foundations’, *The Stanford Encyclopedia of Philosophy*, tinyurl.com/quine-nf.

For a very readable presentation concentrating on NFU (‘New Foundations’ with urelements), and explaining motivations as well as technical details, see

17. M. Randall Holmes, *Elementary Set Theory with a Universal Set* (Cahiers du Centre de Logique No. 10, Louvain, 1998). Now freely available at tinyurl.com/holmesnf.

The following is rather tougher going, though with many interesting ideas:

18. T. F. Forster, *Set Theory with a Universal Set* Oxford Logic Guides 31 (Clarendon Press, 2nd edn. 1995).

(g) Famously, Zermelo constructed his theory of sets by gathering together some principles of set-theoretic reasoning that seemed actually to be used by working mathematicians (engaged in e.g. the rigorization of analysis or the development of point set topology), hoping to get a theory strong enough for mathematical use while weak enough to avoid paradox. The later Axiom of Replacement was added in much the same spirit. But does the result overshoot? We've already noted that SP is a weaker theory which may suffice. For a more radical approach, see this very engaging short piece:

19. Tom Leinster, 'Rethinking set theory'. Gives an advertising pitch for the merits of Lawvere's Elementary Theory of the Category of Sets (ETCS). tinyurl.com/leinst.

And for more on that, you could see e.g.

20. F. William Lawvere and Robert Rosebrugh, *Sets for Mathematicians* (CUP 2003) gives a presentation which in principle doesn't require that you have already done any category theory. But I suspect that it won't be an easy ride if you know no category theory (and philosophers will find it conceptually puzzling too – what *are* these 'abstract sets' that we are supposedly theorizing about?). In my judgement, to really appreciate what's going on, you will have to start engaging with more category theory. Which is a whole new ball game . . .

(h) I'll finish by briefly mentioning two other directions you could go in!

First, ZF/ZFC has a classical logic: what if we change the logic to intuitionistic logic? what if we have more general constructivist scruples? The place to start exploring is

21. Laura Crosilla, 'Set Theory: Constructive and Intuitionistic ZF', *The Stanford Encyclopedia of Philosophy*, tinyurl.com/crosilla.

Second, you'll recall from elementary model theory that Abraham Robinson developed a rigorous formal treatment that takes infinitesimals seriously. Later, a simpler and arguably more natural approach, based on so-called Internal Set Theory, was invented by Edward Nelson. He advertises it here:

22. Edward Nelson, 'Internal Set Theory: a new approach to nonstandard analysis', *Bulletin of The American Mathematical Society* 83 (1977), pp. 1165–1198. tinyurl.com/nelson-ist.

You can follow that up by looking at the approachable early chapters of Nader Vakil's *Real Analysis through Modern Infinitesimals* (CUP, 2011), a monograph developing Nelson's ideas.

12.6 More proof theory

(a) In §9.5, I mentioned three excellent books which are introductory in intent but which take us a step up from the basic steps in proof theory, namely Takeuti's *Proof Theory*, Girard's *Proof Theory and Logical Complexity*, and Troelstra and Schwichtenberg's *Basic Proof Theory*. If you didn't take a look at them before, now might be the time to do so!.

Also worth reading is the editor's own first contribution to

1. Samuel R. Buss, ed., *Handbook of Proof Theory* (North-Holland, 1998). Later chapters of this very substantial handbook do get pretty hard-core, though you might want to look at some of them later. But the 78 pp. opening chapter by Buss himself, a 'Introduction to Proof Theory', is readable, and freely downloadable from tinyurl.com/buss-intro.

(b) And now the paths through proof theory fork. One path investigates what happens when we tinker with the structural rules shared by classical and intuitionistic logic.

Note for example the inference which takes us from the trivial $P \vdash P$ by weakening to $P, Q \vdash P$ and on, via conditional proof, to $P \vdash Q \to P$. If we want a conditional that conforms better to intuitive constraints of relevance, then we need to block that proof: is 'weakening' the culprit? The investigation of what happens if we vary rules such as weakening belongs to 'substructural logic', whose concerns are outlined in

2. Greg Restall, 'Substructural logics', *The Stanford Encyclopedia of Philosophy*, tinyurl.com/sep-subs

And the place to continue exploring these themes at length is the same author's

3. Greg Restall, *An Introduction to Substructural Logics* (Routledge, 2000), which will also teach you more about proof theory generally in a very accessible way. Do try at least the first seven chapters.

(c) Another path forward picks up from Gentzen's proof of the consistency of arithmetic. Recall, that depends on transfinite induction along ordinals up to ε_0; and the fact that it requires just this much transfinite induction to prove the consistency of first-order PA is an important characterization of the strength of the theory.

The more general project of 'ordinal analysis' in proof theory aims to provide comparable characterizations of other theories in terms of the amount of transfinite induction that is needed to prove *their* consistency. Things do get quite hairy quite quickly, however. But you can start from two very useful sets of notes for mini courses:

4. Michael Rathjen, 'The realm of ordinal analysis' and 'Proof theory: from arithmetic to set theory', downloadable from tinyurl.com/rath-art and tinyurl.com/rath-ast.

(d) Finally, here are a couple more books of notable interest:

5. Wolfram Pohlers, *Proof Theory: The First Step into Impredicativity* (Springer 2009). This book officially has introductory ambitions, focusing on ordinal analysis. However, I would judge that it requires quite an amount of mathematical sophistication from its reader. From the blurb: "As a 'warm up' Gentzen's classical analysis of pure number theory is presented in a more modern terminology, followed by an explanation and proof of the famous result of Feferman and Schütte on the limits of predicativity." The first half of the book is probably manageable if (but only if) you already have done some of the other reading. But then the going gets pretty tough.
6. H. Schwichtenberg and S. Wainer, *Proofs and Computations* (Association of Symbolic Logic/CUP 2012) "studies fundamental interactions between proof-theory and computability". The first four chapters, at any rate, will be of wide interest, giving another take on some basic material and should be manageable given enough background. However, to my surprise, I found the book to be not particularly well written and I wonder if it sometimes makes heavier weather of its material than seems really necessary. Still, worth getting to grips with.

12.7 Higher-order logic, the lambda calculus, and type theory

(a) The logical grammar of first-order logic is very restricted. We assume a domain of objects that we can quantify over; we can have names for some of these objects; we can express properties and relations defined over those objects; and can express (total) functions from one or more objects as inputs to objects as outputs. In informal mathematics, by contrast, we quantify not only over a given domain of objects but over their properties and relations and over functions between objects too (as in second-order logic). And we also consider e.g. properties of relations (like being symmetric), relations between functions (like being asymptotically equal), functions from one function to another (e.g. differentiation), and more.

Now, as is familiar, we can trade in properties of relations, relations between functions, functions of functions, etc. for *sets*. So we can compensate for the expressive limitations of first-order logic by adopting enough set theory. Still, we might reasonably look for a more expressive logical framework in which we can talk directly about more types of things, and quantify over more types of things, without playing the set-theory card. And exploring such a higher-order logic might even offer the prospect of an alternative, non-set-theoretic, foundation for mathematics.

We looked at a small fragment of higher-order logic in Chapter 4 on second-order logic. But now we want to explore theories with a richer type-structure. Such a theory of types goes back at least until Bertrand Russell's 1908 paper 'Mathematical logic as based on the theory of types'. Its history since Russell has been rather chequered. But particularly in the hands of theoretical computer scientists, type theories have come back into considerable prominence. And in the recent guise of homotopy type theory, one particular version is advertised as a new foundation for mathematics. But where to start?

You could first take a quick look at

1. Jouko Väänänen, 'Second-order and higher-order logic', *The Stanford Encyclopedia of Philosophy*, tinyurl.com/sep-vaan.
2. Thierry Coquand, 'Type theory', *The Stanford Encyclopedia of Philosophy*, tinyurl.com/sep-type.

But the first of these mostly revisits second-order logic at a probably quite unnecessarily sophisticated level for now, so don't get bogged down. The second gives us pointers forward, but is perhaps also rather too rushed.

Still, as you'll see from Coquand, basic topics to pursue include Simple Type Theory and the lambda calculus. For a clear and gentle introduction to the latter, see the first seven chapters of the following welcome short book which doesn't assume much mathematical background:

3. Chris Hankin, *An Introduction to Lambda Calculus for Computer Scientists** (College Publications 2004).

Next, as a spur to keep going, you might find this advocacy interesting:

4. William M. Farmer, 'The seven virtues of simple type theory', *Journal of Applied Logic* 6 (2008) 267–286. Available at tinyurl.com/farm-STT.

And then for a bit more on Simple Type Theory/Church's Type Theory, though once more this is less than ideal, you could look at

5. Christoph Benzmüller and Peter Andrews, 'Church's type theory', *The Stanford Encyclopedia of Philosophy*, tinyurl.com/sep-CTT.

But then where to go next will depend on your interests and on how much more you want to know.

And a complicating factor is that a lot of current work on type theory is bound up with constructivist ideas developing the BHK conception that ties the content of a proposition to its proofs (for example, an implication $A \to C$ corresponds to a type of function taking a proof A to a proof of C). This correspondence between propositions and types of functions gets developed into the so-called Curry-Howard correspondence or isomorphism. See

6. Peter Dybjer and Erik Palmgren, 'Intuitionistic type theory', *The Stanford Encyclopedia of Philosophy*, tinyurl.com/sep-ITT.

Again, although this is supposed to be introductory, this certainly isn't easy going. So I hoped that this very short book would work better:

7. John L. Bell, *Higher-Order Logic and Type Theory* (CUP: Elements in Philosophy and Logic, 2022).

   From the blurb: This book "begins with a presentation of the syntax and semantics of classical second-order logic .... This leads to a discussion of higher-order logic based on the concept of a type. The second Section contains an account of the origins and nature of type theory, and its relationship to set theory. Section 3 introduces Local Set Theory (also known as higher-order intuitionistic logic), an important form of type theory based on intuitionistic logic. In Section 4 a number of contemporary forms of type theory are described, all of which are based on the so-called 'doctrine of propositions as types'."

But again this is harder going than the author intended and I can't enthuse.

(b) If you want to explore further, here are a number of suggestions to explore, depending on your interests. As already remarked, type theories have been a major concern of computer scientists, and some of the books I'll mention are coming from that angle. In order of publication date:

8. Henk P. Barendregt, *The Lambda Calculus: Its Syntax and Semantics** (Originally 1980, reprinted by College Publications 2012). This is the very weighty standard text: but the opening chapters — say, the first eight, are moderately accessible.
9. Peter Andrews, *An Introduction to Mathematical Logic and Type Theory: To Truth Through Proof* (Academic Press, 1986). Chapter 5, under 50 pages, is a classic introduction to a version of Church's type theory developed by Andrews. It is often recommended, and worth battling through; but it *is* a rather terse bit of old-school exposition.
10. J. Roger Hindley, *Basic Simple Type Theory* (CUP, 1997). This short book is another classic, but again it is pretty terse. Worth making a start, but perhaps, in the end, mostly for those whose main interest is in computer science applications of type theory in the design of higher-level programming languages like ML.
11. Benjamin C. Pierce, *Types and Programming Languages* (MIT Press, 2002). A frequently-recommended text for computer scientists, and readable by others if you skip over some parts about implementation in ML. The first dozen or so shortish chapters are relatively discursive and accessible.
12. Morten Heine Sørensen and Pawel Urzyczyn, *Lectures on the Curry-Howard Isomorphism* (Elsevier, 2006). This engaging book and often-recommended book ranges much more widely than the title might suggest! The early chapters, at least, are reasonably accessible too.
13. J. Roger Hindley and Jonathan P. Seldin, *Lambda-Calculus and Combinators: An Introduction* (CUP 2008). Attractively and clearly written,

aiming to avoid excess technicalities. More of the feel of a modern maths book. Recommended.

14. Rob Nederpelt and Hedman Geuvers, *Type Theory and Formal Proof: An Introduction* (CUP 2014). Focuses, the authors say, "on the use of types and lambda terms for the complete formalisation of mathematics", so promises to be of particular interest to mathematical logicians. It is also attractively and clearly written (as these things go!). Recommended.

15. Samuel Mimram, *PROGRAM = PROOF** (Amazon 2020), and downloadable at tinyurl.com/smimram. A substantial and attractively written book originating from a course for computer scientists: you will need to know a bit about functional programming to get the most out of this, but the chapters on logic and the lambda calculus are good and more generally accessible.

Harold Simmons has a book *Derivation and Computation: Taking the Curry-Howard Correspondence Seriously* (CUP 2000) which I found disappointingly opaque (surprisingly so, as Simmons usually writes well and accessibly). I was even more disappointed by *A Modern Perspective on Type Theory: From its Origins until Today* by Fairouz Kamareddine, Twan Laan and Rob Nederpelt (Kluwer 2004) which might be of interest to those with a computer-science interest in proof-checkers, but isn't for the rest of us.

Finally, I suppose I should finish by mentioning again one particular new incarnation of type theory:

16. The Univalent Foundations Program, *Homotopy Type Theory: Univalent Foundations of Mathematics* (2013), tinyurl.com/HOTT-book.

I leave it to you to make what you will of that program!

Index of authors