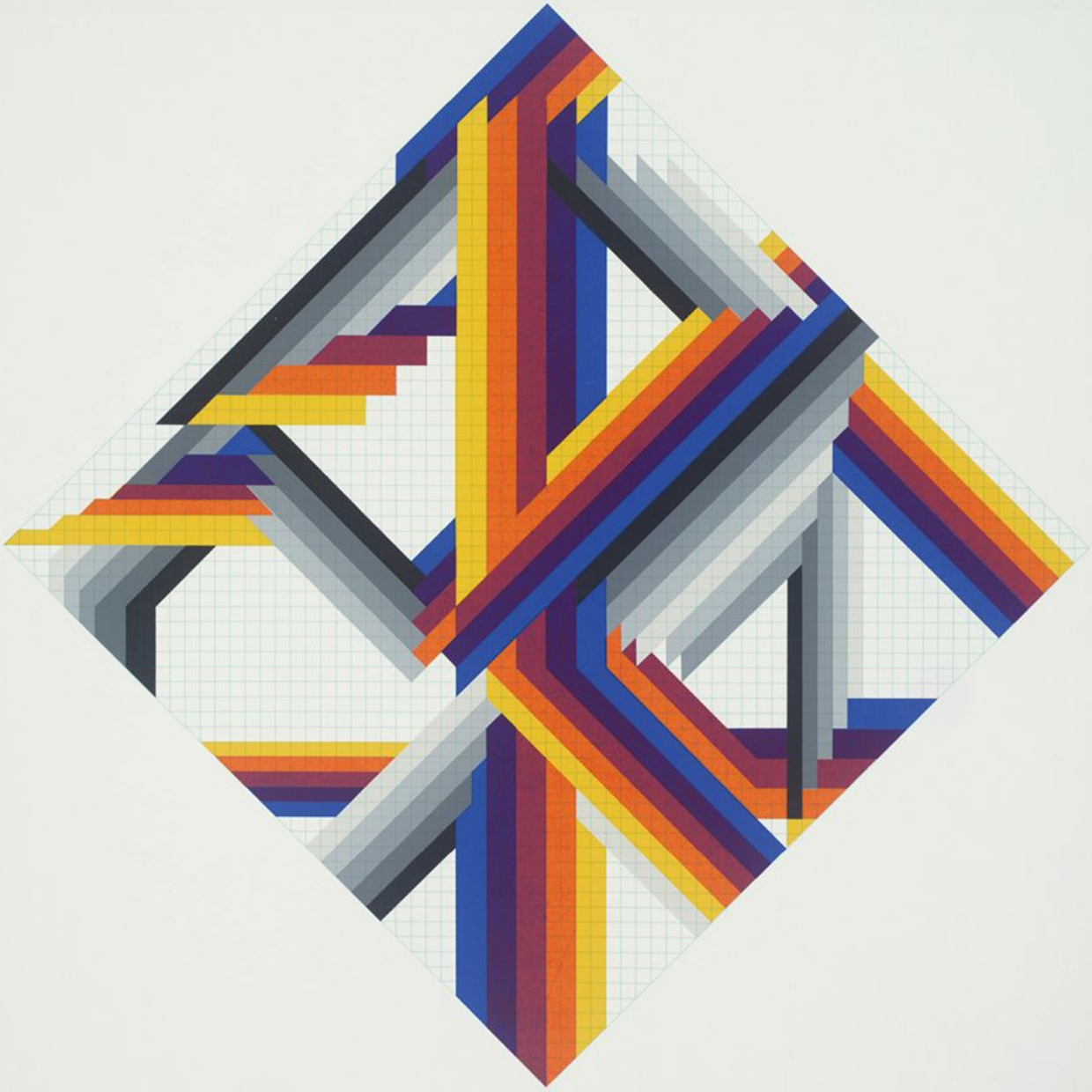
philosophy and model theory

TIM BUTTON | SEAN WALSH



OXFORD

Philosophy and Model Theory

Philosophy and Model Theory

Tim Button and Sean Walsh
with a historical appendix by Wilfrid Hodges

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2018

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2017959066

ISBN: 978–0–19–879039–6 (hbk.)
978–0–19–879040–2 (pbk.)

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Preface

Philosophy and model theory frequently meet one another. This book aims to understand their interactions.

Model theory is used in every 'theoretical' branch of analytic philosophy: in philosophy of mathematics; in philosophy of science; in philosophy of language; in philosophical logic; and in metaphysics. But these wide-ranging appeals to model theory have created a highly fragmented literature. On the one hand, many philosophically significant results are found only in mathematics textbooks: these are aimed squarely at mathematicians; they typically presuppose that the reader has a serious background in mathematics; and little clue is given as to their philosophical significance. On the other hand, the philosophical applications of these results are scattered across disconnected pockets of papers.

The first aim of our book, then, is to consider the *philosophical uses of model theory*. We state and prove the best versions of results for philosophical purposes. We then probe their philosophical significance. And we show how similar dialectical situations arise repeatedly across fragmented debates in different areas.

The second aim of our book, though, is to consider the *philosophy of model theory*. Model theory itself is rarely taken as the subject matter of philosophising (contrast this with the philosophy of biology, or the philosophy of set theory). But model theory is a beautiful part of pure mathematics, and worthy of philosophical study in its own right.

Both aims give rise to challenges. On the one hand: the philosophical uses of model theory are scattered across a disunified literature. And on the other hand: there is scarcely any literature on the philosophy of model theory.

All of which is to say: *philosophy and model theory isn't really 'a thing' yet*. This book aims to start carving out such a thing. We want to chart the rock-face and trace its dialectical contours. But we present this book, not as a final word on what philosophically inclined model theorists and model-theoretically inclined philosophers should do, but as an invitation to join in.

So. This is not a book in which a single axe is ground, page by page, into an increasingly sharp blade. No fundamental line of argument—arching from Chapter 1 through to Chapter 17—serves as the spine of the book. What knits the chapters together into a single book is not a single thesis, but a sequence of overlapping, criss-crossing themes.

Topic selection

Precisely because philosophy and model theory isn't yet 'a thing', we have had to make some difficult decisions about what topics to discuss.

On the one hand, we aimed to pick topics which should be of fairly mainstream philosophical concern. So, when it comes to the philosophical *uses* of model theory, we have largely considered topics concerning *reference*, *realism*, and *doxology* (a term we introduce in Chapter 6). But, even when we have considered questions which fall squarely within in the philosophy *of* model theory, the questions that we have focussed on are clearly instances of 'big questions'. We look at questions of *sameness* of theories/structure (Chapter 5); of taking *diverse perspectives* on the same concept (in Chapter 14); of how to draw *boundaries* of logic (in Chapter 16); and of *classification* of mathematical objects (Chapter 17).

On the other hand, we also wanted to give you a decent bang for your buck. We figured that if you were going to have to wrestle with some new philosophical idea or model-theoretic result, then you should get to see it put to decent use. (This explains why, for example, the Push-Through Construction, the just-more theory manoeuvre, supervaluational semantics, and the ideas of moderation and modelism, occur so often in this book.) Conversely, we have had to set aside debates—no matter how interesting—which would have taken too long to set up.

All of which is to say: this book is not comprehensive. Not even close.

Although we consider models of set theory in Chapters 8 and 11, we scarcely scratch the surface. Although we discuss infinitary logics in Chapters 15–16, we only use them in fairly limited ways.[1] Whilst we mention Tennenbaum's Theorem in Chapter 7, that is as close as we get to computable model theory. We devote only one brief section to o-minimality, namely §4.10. And although we consider quantifiers in Chapter 16 and frequently touch on issues concerning logical consequence, we never address the latter topic head on.[2]

The grave enormity, though, is that we have barely scratched the surface of model theory itself. As the table of contents reveals, the vast majority of the book considers model theory as it existed before Morley's Categoricity Theorem.

Partially correcting for this, Wilfrid Hodges' wonderful historical essay appears as Part D of this book. Wilfrid's essay treats Morley's Theorem as a pivot-point for the subject of model theory. He looks back critically to the history, to uncover the notions at work in Morley's Theorem and its proof, and he looks forward to the riches that have followed from it. We have both learned so much from Wilfrid's *Model Theory*,[3] and we are delighted to include his 'short history' here.

[1] We do not, for example, consider the connection between infinitary logics and supervenience, as Glanzberg (2001) and Bader (2011) do.

[2] For that, we would point the reader to Blanchette (2001) and Shapiro (2005a).

[3] Hodges (1993).

Still, concerning all those topics which we have omitted: we intend no slight against them. But there is only so much one book can do, and this book is already much (much) longer than we originally planned. We are sincere in our earlier claim, that this book is not offered as a final word, but as an invitation to take part.

Structuring the book

Having selected our topics, we needed to arrange them into a book. At this point, we realised that these topics have no natural linear ordering.

As such, we have tried to strike a balance between three aims that did not always point in the same direction: to order by *philosophical theme*, to order by increasing *philosophical sophistication*, and to order by increasing *mathematical sophistication*. The book's final structure of represents our best compromise between these three aims. It is divided into three main parts: *Reference and realism*, *Categoricity*, and *Indiscernibility and classification*.

Each part has an introduction, and those who want to dip in and out of particular topics, rather than reading cover-to-cover, should read the three part-introductions after they have finished reading this preface. The part-introductions provide thematic overviews of each chapter, and they also contain diagrams which depict the dependencies between each section of the book. In combination with the table of contents, these diagrams will allow readers to take shortcuts to their favourite destinations, without having to stop to smell every rose along the way.

Presuppositions and proofs

So far as possible, the book assumes only that you have completed a 101-level logic course, and so have some familiarity with first-order logic.

Inevitably, there are some exceptions to this: we were forced to assume some familiarity with analysis when discussing infinitesimals in Chapter 4, and equally some familiarity with topology when discussing Stone spaces in Chapter 14. We do not prove Gödel's incompleteness results, although we do state versions of them in §5.A. A book can only be so self-contained.

By and large, though, this book *is* self-contained. When we invoke a model-theoretic notion, we almost always define the notion formally in the text. When it comes to proofs, we follow these rules of thumb.

The *main text* includes both brief proofs, and also those proofs which we wanted to discuss directly.

The *appendices* include proofs which we wanted to include in the book, but which were too long to feature in the main text. These include: proofs concerning elementary topics which our readers should come to understand (at least one

day); proofs which are difficult to access in the existing literature; proofs of certain folk-lore results; and proofs of new results.

But the *book omits* all proofs which are both readily accessed and too long to be self-contained. In such cases, we simply provide readers with citations.

The quick moral for readers to extract is this. If you encounter a proof in the main text of a chapter, you should follow it through. But we would add a note for readers whose primary background is in philosophy. If you really want to understand a mathematical concept, you need to see it in action. Read the appendices!

Acknowledgements

The book arose from a seminar series on philosophy and model theory that we ran in Birkbeck in Autumn 2011. We turned the seminar into a paper, but it was vastly too long. An anonymous referee for *Philosophia Mathematica* suggested the paper might form the basis for a book. So it did.

We have presented topics from this book several times. It would not be the book it is, without the feedback, questions and comments we have received. So we owe thanks to: an anonymous referee for *Philosophia Mathematica*, and James Studd for OUP; and to Sarah Acton, George Anegg, Andrew Arana, Bahram Assadian, John Baldwin, Kyle Banick, Neil Barton, Timothy Bays, Anna Bellomo, Liam Bright, Chloé de Canson, Adam Caulton, Catrin Campbell-Moore, John Corcoran, Radin Dardashti, Walter Dean, Natalja Deng, William Demopoulos, Michael Detlefsen, Fiona Doherty, Cian Dorr, Stephen Duxbury, Sean Ebels-Duggan, Sam Eklund, Hartry Field, Branden Fitelson, Vera Flocke, Salvatore Florio, Peter Fritz, Michael Gabbay, Haim Gaifman, J. Ethan Galebach, Marcus Giaquinto, Peter Gibson, Tamara von Glehn, Owen Griffiths, Emmylou Haffner, Bob Hale, Jeremy Heis, Will Hendy, Simon Hewitt, Kate Hodesdon, Wilfrid Hodges, Luca Incurvati, Douglas Jesseph, Nicholas Jones, Peter Koellner, Brian King, Eleanor Knox, Johannes Korbmacher, Arnold Koslow, Hans-Christoph Kotzsch, Greg Lauro, Sarah Lawsky, Øystein Linnebo, Yang Liu, Pen Maddy, Kate Manion, Tony Martin, Guillaume Massas, Vann McGee, Toby Meadows, Richard Mendelsohn, Christopher Mitsch, Stella Moon, Adrian Moore, J. Brian Pitts, Jonathan Nassim, Fredrik Nyseth, Sara Parhizgari, Charles Parsons, Jonathan Payne, Graham Priest, Michael Potter, Hilary Putnam, Paula Quinon, David Rabouin, Erich Reck, Sam Roberts, Marcus Rossberg, J. Schatz, Gil Sagi, Bernhard Salow, Chris Scambler, Thomas Schindler, Dana Scott, Stewart Shapiro, Gila Sher, Lukas Skiba, Jönne Speck, Sebastian Speitel, Will Stafford, Trevor Teitel, Robert Trueman, Jouko Väänänen, Kai Wehmeier, J. Robert G. Williams, John Wigglesworth, Hugh Woodin, Jack Woods, Crispin Wright, Wesley Wrigley, and Kino Zhao.

We owe some special debts to people involved in the original Birkbeck seminar.

First, the seminar was held under the auspices of the Department of Philosophy at Birkbeck and Øystein Linnebo's European Research Council-funded project 'Plurals, Predicates, and Paradox', and we are very grateful to all the people from the project and the department for participating and helping to make the seminar possible. Second, we were lucky to have several great external speakers visit the seminar, whom we would especially like to thank. The speakers were: Timothy Bays, Walter Dean, Volker Halbach, Leon Horsten, Richard Kaye, Jeff Ketland, Angus Macintyre, Paula Quinon, Peter Smith, and J. Robert G. Williams. Third, many of the external talks were hosted by the Institute of Philosophy, and we wish to thank Barry C. Smith and Shahrar Ali for all their support and help in this connection.

A more distant yet important debt is owed to Denis Bonnay, Brice Halimi, and Jean-Michel Salanskis, who organised a lovely event in Paris in June 2010 called 'Philosophy and Model Theory.' That event got some of us first thinking about 'Philosophy and Model Theory' as a unified topic.

We are also grateful to various editors and publishers for allowing us to reuse previously published material. Chapter 5 draws heavily on Walsh 2014, and the copyright is held by the Association for Symbolic Logic and is being used with their permission. Chapters 7–11 draw heavily upon on Button and Walsh 2016, published by *Philosophia Mathematica*. Finally, §13.7 draws from Button 2016b, published by *Analysis*, and §15.1 draws from Button 2017, published by the *Notre Dame Journal of Formal Logic*.

Finally, though, a word from us, as individuals.

From Tim. I want to offer my deep thanks to the Leverhulme Trust: their funding, in the form of a Philip Leverhulme Prize (PLP–2014–140), enabled me to take the research leave necessary for this book. But I mostly want to thank two very special people. Without Sean, this book could not be. And without my Ben, I could not be.

From Sean. I want to thank the Kurt Gödel Society, whose funding, in the form of a Kurt Gödel Research Prize Fellowship, helped us put on the original Birkbeck seminar. I also want to thank Tim for being a model co-author and a model friend. Finally, I want to thank Kari for her complete love and support.

Contents

A

Reference and realism

Introduction to Part A

The two central themes of Part A are *reference* and *realism*.

Here is an old philosophical chestnut: *How do we (even manage to) represent the world?* Our most sophisticated representations of the world are perhaps linguistic. So a specialised—but still enormously broad—version of this question is: *How do words (even manage to) represent things?*

Enter model theory. One of the most basic ideas in model theory is that a structure assigns interpretations to bits of vocabulary, and in such a way that we can make excellent sense of the idea that the structure makes each sentence (in that vocabulary) either true or false. Squint slightly, and model theory seems to be providing us with a perfectly precise, formal way to understand certain aspects of linguistic representation. It is no surprise at all, then, that almost any philosophical discussion of linguistic representation, or reference, or truth, ends up invoking notions which are recognisably model-theoretic.

In Chapter 1, we introduce the building blocks of model theory: the notions of signature, structure, and satisfaction. Whilst the bare technical bones should be familiar to anyone who has covered a 101-level course in mathematical logic, we also discuss the philosophical question: *How should we best understand quantifiers and variables?* Here we see that philosophical issues arise at the very outset of our model-theoretic investigations. We also introduce second-order logic and its various semantics. While second-order logic is less commonly employed in contemporary model theory, it is employed frequently in philosophy of model theory, and understanding the differences between its various semantics will be important in many subsequent chapters.

In Chapter 2, we examine various concerns about the determinacy of reference and so, perhaps, the determinacy of our representations. Here we encounter famous arguments from Benacerraf and Putnam, which we explicate using the formal Push-Through Construction. Since isomorphic structures are elementarily equivalent—that is, they make exactly the same sentences true and false—this threatens the conclusion that it is radically indeterminate, which of many isomorphic structures accurately captures how language represents the world.

Now, one might think that the reference of our word 'cat' is constrained by the causal links between cats and our uses of that word. Fair enough. But there are no causal links between mathematical objects and mathematical words. So, on certain conceptions of what humans are like, we will be unable to answer the question: *How do we (even manage to) refer to any particular mathematical entity?* That is, we will have to accept that we *do not* refer to particular mathematical entities.

Whilst discussing these issues, we introduce Putnam's famous *just-more-theory manoeuvre*. It is important to do this both clearly and early, since many versions of this dialectical move occur in the philosophical literature on model theory. Indeed, they occur especially frequently in Part B of this book.

Now, philosophers have often linked the topic of reference to the topic of realism. One way to draw the connection is as follows: If reference is radically indeterminate, then my word 'cabbage' and my word 'cat' fail to pick out anything determinately. So when I say something like 'there is a cabbage and there is a cat', I have *at best* managed to say that there are at least two distinct objects. That seems to fall far short of expressing any real commitment to *cats* and *cabbages* themselves.[1] In short, radical referential indeterminacy threatens to undercut certain kinds of realism altogether. But only certain kinds: we close Chapter 2 by suggesting that some versions of mathematical platonism can live with the fact that mathematical language is radically referentially indeterminate by embracing a supervaluational semantics.

Concerns about referential indeterminacy also feature in discussions about realism within the philosophy of science. In Chapter 3, we examine a particular version of scientific realism that arises by considering Ramsey sentences. Roughly, these are sentences where all the 'theoretical vocabulary' has been 'existentially quantified away'. Ramsey sentences seem promising, since they seem to incur a kind of existential commitment to theoretical entities, which is characteristic of realism, whilst making room for a certain level referential indeterminacy. We look at the relation between Newman's objection and the Push-Through Construction of Chapter 2, and between Ramsey sentences and various model-theoretic notions of conservation. Ultimately, by combining the Push-Through Construction with these notions of conservation, we argue that the dialectic surrounding Newman's objection should track the dialectic of Chapter 2, surrounding Putnam's permutation argument in the philosophy of mathematics.

The notions of conservation we introduce in Chapter 3 are crucial to Abraham Robinson's attempt to use model theory to salvage Leibniz's notion of an 'infinitesimal'. Infinitesimals are quantities whose absolute value is smaller than that of any given positive real number. They were an important part of the historical calculus; they fell from grace with the rise of ε–δ notation; but they were given a new lease of life within model theory via Robinson's non-standard analysis. This is the topic of Chapter 4. Here we introduce the idea of *compactness* to prove that the use of infinitesimals is consistent.

Robinson believed that this vindicated the viability of the Leibnizian approach to the calculus. Against this, Bos has questioned whether Robinson's non-standard analysis is genuinely faithful to Leibniz's mathematical practice. In Chapter 4, we

[1] Cf. Putnam (1977: 491) and Button (2013: 59–60).

offer a novel defence of Robinson. By building valuations into Robinson's model theory, we prove new results which allow us to approximate more closely what we know about the Leibnizian conception of the structure of the infinitesimals. Indeed, we show that Robinson's non-standard analysis can rehabilitate various historical methods for reasoning with and about infinitesimals that have fallen far from fashion.

The question remains, of whether we should *believe* in infinitesimals. Leibniz himself was tempted to treat his infinitesimals as 'convenient fictions'; Robinson explicitly regarded his infinitesimals in the same way; and their method of introduction in model theory allows for perhaps the cleanest possible version of a fictionalist-cum-instrumentalist attitude towards 'troublesome' entities. Indeed, we can prove that reasoning *as if* there are infinitesimals will only generate results that one could have obtained *without* that assumption. One can have anti-realism, then, with a clear conscience.

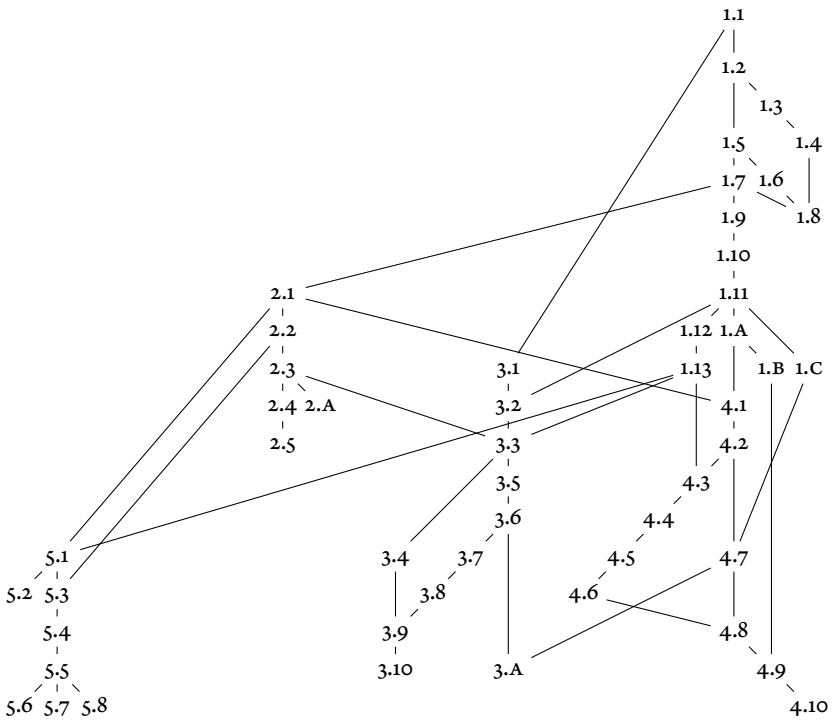
In Chapter 5, we take a step back from these specific applications of model theory, to discuss a more methodological question about the philosophical application of model theory: *under what circumstances should we call two structures 'the same'*? This question can be posed within mathematics, where its answer will depend upon the similarities and differences that matter for the mathematical purposes at hand. But the question can also be given a metaphysical gloss. In particular, consider a philosopher who thinks (for example) that: (a) there is a *single*, abstract, entity which is 'the natural number structure', and that (b) there is a *single*, abstract entity which is 'the structure of the integers'; but that (c) these two entities are distinct. Then this philosopher must provide an account of identity and distinctness between 'structures', so construed; and we show just how hard this is.

Notions of sameness of structure also induce notions of sameness of theory. After surveying a wide variety of formal notions of sameness of structure and theory, we discuss three ambitious claims concerning what sameness of theory preserves, namely: truth; arithmetical provability; and proof. We conclude that more philosophically ambitious versions of these preservation-theses generally fail.

This meta-issue of sameness of structure and theory is a good place to end Part A, though, both because (a) the discussion is enhanced by the specific examples of structures and theories discussed earlier in the text, and because (b) questions about sameness of structure and theory inform a number of the discussions and debates which we treat in later Parts of the book.

Readers who only want to dip into particular topics of Part A can consult the following Hasse diagram of dependencies between the sections of Part A, whilst referring to the table of contents. A section y depends upon a section x iff there is a path leading downwards from x to y. So, a reader who wants to get straight to the discussion of fictionalism about infinitesimals will want to leap straight to §4.7; but

they should know that this section assumes a prior understanding of §§2.1, 4.1, 4.2, and much (but not all) of Chapter 1. (We omit purely technical appendices from this diagram.)



1

Logics and languages

Model theory begins by considering the relationship between languages and structures. This chapter outlines the most basic aspects of that relationship.

One purpose of the chapter will therefore be immediately clear: we want to lay down some fairly dry, technical preliminaries. Readers with some familiarity with mathematical logic should feel free to skim through these technicalities, as there are no great surprises in store.

Before the skimming commences, though, we should flag a second purpose of this chapter. There are at least three rather different approaches to the semantics for formal languages. In a straightforward sense, these approaches are technically equivalent. Most books simply choose one of them without comment. We, however, lay down all three approaches and discuss their comparative strengths and weaknesses. Doing this highlights that there are philosophical discussions to be had from the get-go. Moreover, by considering what is invariant between the different approaches, we can better distinguish between the merely idiosyncratic features of a particular approach, and the things which really matter.

One last point, before we get going: tradition demands that we issue a caveat. Since Tarski and Quine, philosophers have been careful to emphasise the important distinction between *using* and *mentioning* words. In philosophical texts, that distinction is typically flagged with various kinds of quotation marks. But within model theory, context almost always disambiguates between use and mention. Moreover, including too much punctuation makes for ugly text. With this in mind, we follow model-theoretic practice and avoid using quotation marks except when they will be especially helpful.

1.1 Signatures and structures

We start with the idea that formal languages can have primitive vocabularies:

Definition 1.1: *A* signature*, $\mathscr{L}$, is a set of symbols, of three basic kinds: constant symbols, relation symbols, and function symbols. Each relation symbol and function symbol has an associated number of* places *(a natural number), so that one may speak of an n-place relation or function symbol.*

Throughout this book, we use script fonts for signatures. Constant symbols should be thought of as *names* for entities, and we tend to use c_1, c_2, etc. Relation symbols, which are also known as predicates, should be thought of as picking out *properties* or *relations*. A two-place relation, such as *x is smaller than y*, must be associated with a two-place relation symbol. We tend to use R_1, R_2, etc. for relation symbols. Function symbols should be thought of as picking out functions and, again, they need an associated number of places: the function of *multiplication on the natural numbers* takes two natural numbers as inputs and outputs a single natural number, so we must associate that function with a two-place function symbol. We tend to use f_1, f_2, etc. for function symbols.

The examples just given—*being smaller than*, and *multiplication on the natural numbers*—suggest that we will use our formal vocabulary to make determinate claims about certain objects, such as people or numbers. To make this precise, we introduce the notion of an $\mathscr{L}$*-structure*; that is, a structure whose signature is $\mathscr{L}$. An $\mathscr{L}$-structure, $\mathcal{M}$, is an underlying domain, M, together with an assignment of $\mathscr{L}$'s constant symbols to elements of M, of $\mathscr{L}$'s relation symbols to relations on M, and of $\mathscr{L}$'s function symbols to functions over M. We always use calligraphic fonts $\mathcal{M}$, $\mathcal{N}$, ... for structures, and M, N, ... for their underlying domains. Where s is any $\mathscr{L}$-symbol, we say that $s^{\mathcal{M}}$ is the object, relation or function (as appropriate) assigned to s in the structure $\mathcal{M}$. This informal explanation of an $\mathscr{L}$-structure is always given a set-theoretic implementation, leading to the following definition:

Definition 1.2: *An* $\mathscr{L}$-structure, $\mathcal{M}$, *consists of:*

- *a non-empty set, M, which is the underlying domain of* $\mathcal{M}$,
- *an object* $c^{\mathcal{M}} \in M$ *for each constant symbol c from* $\mathscr{L}$,
- *a relation* $R^{\mathcal{M}} \subseteq M^n$ *for each n-place relation symbol R from* $\mathscr{L}$, *and*
- *a function* $f^{\mathcal{M}} : M^n \longrightarrow M$ *for each n-place function symbol f from* $\mathscr{L}$.

As is usual in set theory, M^n is just the set of n-tuples over M, i.e.:[1]

$$M^n = \{(a_1, \ldots, a_n) : a_1 \in M \text{ and } \ldots \text{ and } a_n \in M\}$$

Likewise, we implement a function $g : M^n \longrightarrow M$ in terms of its set-theoretic graph. That is, g will be a subset of M^{n+1} such that if $(x_1, \ldots, x_n, y)$ and $(x_1, \ldots, x_n, z)$ are elements of g then $y = z$ and such that for every $(x_1, \ldots, x_n)$ in M^n there is y in M such that $(x_1, \ldots, x_n, y)$ is in g. But we continue to think about functions in the normal way, as maps sending n-tuples of the domain, M^n, to elements of the co-domain, M, so tend to write $(x_1, \ldots, x_n, y) \in g$ just as $g(x_1, \ldots, x_n) = y$.

[1] The full definition of X^n is by recursion: $X^1 = X$ and $X^{n+1} = X^n \times X$, where $A \times B = \{(a, b) : a \in A \text{ and } b \in B\}$. Likewise, we recursively define ordered n-tuples in terms of ordered pairs by setting e.g. $(a, b, c) = ((a, b), c)$.

Given the set-theoretic background, $\mathscr{L}$-structures are individuated *extensionally*: they are identical iff they have exactly the same underlying domain and make exactly the same assignments. So, where $\mathcal{M}, \mathcal{N}$ are $\mathscr{L}$-structures, $\mathcal{M} = \mathcal{N}$ iff both $M = N$ and $s^{\mathcal{M}} = s^{\mathcal{N}}$ for all s from $\mathscr{L}$. To obtain different structures, then, we can either change the domain, change the interpretation of some symbol(s), or both. Structures are, then, individuated rather finely, and indeed we will see in Chapters 2 and 5 that this individuation is too fine for many purposes. But for now, we can simply observe that there are many, *many* different structures, in the sense of Definition 1.2.

1.2 First-order logic: a first look

We know what ($\mathscr{L}$-)structures are. To move to the idea of a *model*, we need to think of a structure as making certain sentences true or false. So we must build up to the notion of a sentence. We start with their syntax.

Syntax for first-order logic

Initially, we restrict our attention to *first-order sentences*. These are the sentences we obtain by adding a basic starter-pack of logical symbols to a signature (in the sense of Definition 1.1). These logical symbols are:

- variables: u, v, w, x, y, z, with numerical subscripts as necessary
- the identity sign: $=$
- a one-place sentential connective: $\neg$
- two-place sentential connectives: $\wedge, \vee$
- quantifiers: $\exists, \forall$
- brackets: $(,)$

We now offer a recursive definition of the syntax of our language:[2]

Definition 1.3: *The following, and nothing else, are first-order* $\mathscr{L}$-terms*:*

- *any variable, and any constant symbol c from* $\mathscr{L}$

[2] A pedantic comment is in order. The symbols 't_1' and 't_2' are not being used here as expressions in the object language (i.e. first-order logic with signature $\mathscr{L}$). Rather, they are being used as expressions of the metalanguage, within which we describe the syntax of first-order $\mathscr{L}$-terms and $\mathscr{L}$-formulas. Similarly, the symbol '*x*', as it occurs in the last clause of Definition 1.3, is not being used as an expression of the object language, but in the metalanguage. So the final clause in this definition should be read as saying something like this. *For any variable and any formula φ which does not already contain a concatenation of a quantifier followed by that variable, the following concatenation is a formula: a quantifier, followed by that variable, followed by φ.* (The reason for this clause is to guarantee that e.g. $\exists v \forall v F(v)$ is not a formula.) We could flag this more explicitly, by using a different font for metalinguistic variables (for example). However, as with flagging quotation, we think the additional precision is not worth the ugliness.

- *$f(t_1, \ldots, t_n)$, for any $\mathscr{L}$-terms $t_1, \ldots, t_n$ and any n-place function symbol f from $\mathscr{L}$*

*The following, and nothing else, are first-order $\mathscr{L}$-*formulas*:*

- *$t_1 = t_2$, for any $\mathscr{L}$-terms t_1 and t_2*
- *$R(t_1, \ldots, t_n)$, for any $\mathscr{L}$-terms $t_1, \ldots, t_n$ and any n-place relation symbol R from $\mathscr{L}$*
- *$\neg\varphi$, for any $\mathscr{L}$-formula φ*
- *$(\varphi \wedge \psi)$ and $(\varphi \vee \psi)$, for any $\mathscr{L}$-formulas φ and ψ*
- *$\exists x\varphi$ and $\forall x\varphi$, for any variable x and any $\mathscr{L}$-formula φ which contains neither of the expressions $\exists x$ nor $\forall x$.*

Formulas of the first two sorts—i.e. terms appropriately concatenated either with the identity sign or an $\mathscr{L}$-predicate—are called atomic $\mathscr{L}$-formulas.

As is usual, for convenience we add two more sentential connectives, $\rightarrow$ and $\leftrightarrow$, with their usual abbreviations. So, $(\varphi \rightarrow \psi)$ abbreviates $(\neg\varphi \vee \psi)$, and $(\varphi \leftrightarrow \psi)$ abbreviates $((\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi))$. We will also use some extremely common bracketing conventions to aid readability, so we sometimes use square brackets rather than rounded brackets, and we sometimes omit brackets where no ambiguity can arise.

We say that a variable is *bound* if it occurs within the scope of a quantifier, i.e. we have something like $\exists x(\ldots x \ldots)$. A variable is *free* if it is not bound. We now say that an *$\mathscr{L}$-sentence* is an $\mathscr{L}$-formula containing no free variables. When we want to draw attention to the fact that some formula φ has certain free variables, say x and y, we tend to do this by writing the formula as $\varphi(x, y)$. We say that $\varphi(x, y)$ is a formula *with free variables displayed* iff x and y are the *only* free variables in φ. When we consider a sequence of n-variables, such as $v_1, \ldots, v_n$, we usually use overlining to write this more compactly, as $\bar{v}$, leaving it to context to determine the number of variables in the sequence. So if we say '$\varphi(\bar{x})$ is a formula with free variables displayed', we mean that all and only its free variables are in the sequence $\bar{x}$. We also use similar overlining for other expressions. For example, we could have phrased part of Definition 1.3 as follows: $f(\bar{t})$ is a term whenever each entry in $\bar{t}$ is an $\mathscr{L}$-term and f is a function symbol from $\mathscr{L}$.

Semantics: the trouble with quantifiers

We now understand the syntax of first-order sentences. Later, we will consider logics with a more permissive syntax. But first-order logic is something like the *default*, for both philosophers and model theorists. And our next task is to understand its *semantics*. Roughly, our aim is to define a relation, $\models$, which obtains between a structure and a sentence just in case (intuitively) the sentence is true in the structure. In

fact, there are many different but extensionally equivalent approaches to defining this relation, and we will consider three in this chapter.

To understand why there are several different approaches to the semantics for first-order logic, we must see why the most obvious approach fails. Our sentences have a nice, recursive syntax, so we will want to provide them with a nice, recursive semantics. The most obvious starting point is to supply semantic clauses for the two kinds of atomic sentence, as follows:

$$\mathcal{M} \vDash t_1 = t_2 \text{ iff } t_1^{\mathcal{M}} = t_2^{\mathcal{M}}$$
$$\mathcal{M} \vDash R(t_1, \ldots, t_n) \text{ iff } (t_1^{\mathcal{M}}, \ldots, t_n^{\mathcal{M}}) \in R^{\mathcal{M}}$$

Next, we would need recursion clauses for the quantifier-free sentences. So, writing $\mathcal{M} \nvDash \varphi$ for *it is not the case that* $\mathcal{M} \vDash \varphi$, we would offer:

$$\mathcal{M} \vDash \neg\varphi \text{ iff } \mathcal{M} \nvDash \varphi$$
$$\mathcal{M} \vDash (\varphi \wedge \psi) \text{ iff } \mathcal{M} \vDash \varphi \text{ and } \mathcal{M} \vDash \psi$$

So far, so good. But the problem arises with the quantifiers. Where the notation $\varphi(c/x)$ indicates the formula obtained by replacing every instance of the free variable x in $\varphi(x)$ with the constant symbol c, an obvious thought would be to try:

$$\mathcal{M} \vDash \forall x\varphi(x) \text{ iff } \mathcal{M} \vDash \varphi(c/x) \text{ for every constant symbol } c \text{ from } \mathscr{L}$$

Unfortunately, *this recursion clause is inadequate*. To see why, suppose we had a very simple signature containing a single one-place predicate R and *no* constant symbols. Then, for any structure $\mathcal{M}$ in that signature, we would *vacuously* have that $\mathcal{M} \vDash \forall v R(v)$. But this would be the case even if $R^{\mathcal{M}} = \varnothing$, that is, even if *nothing* had the property picked out by R. Intuitively, that is the wrong verdict.

The essential difficulty in defining the semantics for first-order logic therefore arises when we confront quantifiers. The three approaches to semantics which we consider present three ways to overcome this difficulty.

Why it is worth considering different approaches

In a straightforward sense, the three approaches are technically equivalent. So most books simply adopt one of these approaches, without comment, and get on with other things. In deciding to present all three approaches here, we seem to be trebling our reader's workload. So we should pause to explain our decision.

First: the three approaches to semantics are so intimately related, at a technical level, that the workload is probably only *doubled*, rather than trebled.

Second: readers who are happy ploughing through technical definitions will find nothing very tricky here. And such readers should find that the additional technical

investment gives a decent philosophical pay-off. For, as we move through the chapter, we will see that these (quite dry) technicalities can both generate and resolve philosophical controversies.

Third: we expect that even novice philosophers reading this book will have at least a rough and ready idea of what is coming next. And such readers will be better served by reading (and perhaps only partially absorbing) multiple *different* approaches to the semantics for first-order logic, than by trying to rote-learn one *specific* definition. They will thereby get a sense of what is important to supplying a semantics, and what is merely an idiosyncratic feature of a particular approach.

1.3 The Tarskian approach to semantics

We begin with the Tarskian approach.[3] Recall that the 'obvious' semantic clauses fail because $\mathscr{L}$ may not contain enough constant symbols. The Tarskian approach handles this problem by assigning interpretations to the *variables* of the language. In particular, where $\mathcal{M}$ is any $\mathscr{L}$-structure, a *variable-assignment* is any function σ from the set of variables to the underlying domain M. We then define satisfaction with respect to pairs of structures with variable-assignments.

To do this, we must first specify how the structure / variable-assignment pair determines the behaviour of the $\mathscr{L}$-terms. We do this by recursively defining an element $t^{\mathcal{M},\sigma}$ of M for a term t with free variables among $x_1, \ldots, x_n$ as follows:

$$t^{\mathcal{M},\sigma} = \sigma(x_i), \text{ if } t \text{ is the variable } x_i$$
$$t^{\mathcal{M},\sigma} = f^{\mathcal{M}}(s_1^{\mathcal{M},\sigma}, \ldots, s_k^{\mathcal{M},\sigma}), \text{ if } t \text{ is the term } f(s_1, \ldots, s_k)$$

To illustrate this definition, suppose that $\mathcal{M}$ is the natural numbers in the signature $\{0, 1, +, \times\}$, with each symbol interpreted as normal. (This licenses us in dropping the '$\mathcal{M}$'-superscript when writing the symbols.) Suppose that σ and τ are variable-assignments such that $\sigma(x_1) = 5, \sigma(x_2) = 7, \tau(x_1) = 3, \tau(x_2) = 7$, and consider the term $t(x_1, x_2) = (1 + x_1) \times (x_1 + x_2)$. Then we can compute the interpretation of the term relative to the variable-assignments as follows:

$$t^{\mathcal{M},\sigma} = (1 + x_1^{\mathcal{M},\sigma}) \times (x_1^{\mathcal{M},\sigma} + x_2^{\mathcal{M},\sigma}) = (1 + 5) \times (5 + 7) = 72$$
$$t^{\mathcal{M},\tau} = (1 + x_1^{\mathcal{M},\tau}) \times (x_1^{\mathcal{M},\tau} + x_2^{\mathcal{M},\tau}) = (1 + 3) \times (3 + 7) = 40$$

We next define the notion of satisfaction relative to a variable-assignment:

[3] See Tarski (1933) and Tarski and Vaught (1958), but also §12.a.

$$\mathcal{M}, \sigma \vDash t_1 = t_2 \text{ iff } t_1^{\mathcal{M},\sigma} = t_2^{\mathcal{M},\sigma}, \text{ for any } \mathscr{L}\text{-terms } t_1, t_2$$

$$\mathcal{M}, \sigma \vDash R(t_1, \ldots, t_n) \text{ iff } (t_1^{\mathcal{M},\sigma}, \ldots, t_n^{\mathcal{M},\sigma}) \in R^{\mathcal{M}}, \text{ for any } \mathscr{L}\text{-terms } t_1, \ldots, t_n$$
$$\text{and any } n\text{-place relation symbol } R \text{ from } \mathscr{L}$$

$$\mathcal{M}, \sigma \vDash \neg\varphi \text{ iff } \mathcal{M}, \sigma \nvDash \varphi$$

$$\mathcal{M}, \sigma \vDash (\varphi \wedge \psi) \text{ iff } \mathcal{M}, \sigma \vDash \varphi \text{ and } \mathcal{M}, \sigma \vDash \psi$$

$$\mathcal{M}, \sigma \vDash \forall x \varphi(x) \text{ iff } \mathcal{M}, \tau \vDash \varphi(x) \text{ for every variable-assignment } \tau$$
$$\text{which agrees with } \sigma \text{ except perhaps on the value of } x$$

We leave it to the reader to formulate clauses for disjunction and existential quantification. Finally, where φ is any first-order $\mathscr{L}$*-sentence*, we say that $\mathcal{M} \vDash \varphi$ iff $\mathcal{M}, \sigma \vDash \varphi$ for all variable-assignments σ.

1.4 Semantics for variables

The Tarskian approach is technically flawless. However, the apparatus of variable-assignments raises certain philosophical issues.

A variable-assignment effectively gives variables a particular interpretation. In that sense, variables are treated rather like names (or constant symbols). However, when we encounter the clause for a quantifier binding a variable, we allow ourselves to consider all of the *other* ways that the bound variable might have been interpreted. In short, the Tarskian approach treats variables as something like *varying names*.

This gives rise to a philosophical question: *should* we regard variables as varying names? With Quine, our answer is *No*: 'the "variation" connoted [by the word "variable"] belongs to a vague metaphor which is best forgotten.'[4]

To explain why we say this, we begin with a simple observation. A Tarskian variable-assignment may assign different semantic values to the formulas $x > 0$ and $y > 0$. But, on the face of it, that seems mistaken. As Fine puts the point, using one variable rather than the other 'would appear to be as clear a case as any of a mere "conventional" or "notational" difference; the difference is merely in the choice of the symbol and not in its linguistic function.'[5] And this leads Fine to say:

(a) 'Any two variables (ranging over a given domain of objects) have the same semantic role.'

[4] Quine (1981: §12). For ease of reference, we cite the 1981-edition. However, the relevant sections are entirely unchanged from the (first) 1940-edition. We owe several people thanks for discussion of material in this section. Michael Potter alerted us to Bourbaki's notation; Kai Wehmeier alerted us to Quine's (cf. Wehmeier forthcoming); and Robert Trueman suggested that we should connect all of this to Fine's antinomy of the variable.

[5] Fine (2003: 606, 2007: 7), for this and all subsequent quotes from Fine.

But, as Fine notes, this cannot be right either. For, 'when we consider the semantic role of the variables in the same expression—such as "$x > y$"—then it seems equally clear that their semantic role is different.' So Fine says:

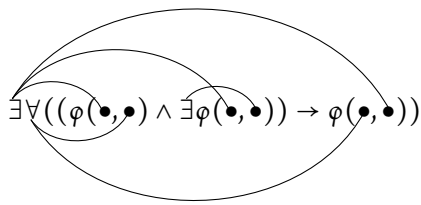
(b) 'Any two variables (ranging over a given domain of objects) have a different semantic role.'

And now we have arrived at Fine's *antinomy of the variable*.

We think that this whole antinomy gets going from the mistaken assumption that we can assign a 'semantic role' to a variable in isolation from the quantifier which binds it.[6] As Quine put the point more than six decades before Fine: 'The variables [...] serve merely to indicate cross-references to various positions of quantification.'[7] Quine's point is that $\exists x \forall y \varphi(x,y)$ and $\exists y \forall x \varphi(y,x)$ are indeed just typographical variants, but that both are importantly different from $\forall x \exists y \varphi(x,y)$. And to illustrate this graphically, Quine notes that we could use a notation which abandons typographically distinct variables altogether. For example, instead of writing:

$$\exists x \forall y((\varphi(x,y) \wedge \exists z \varphi(x,z)) \rightarrow \varphi(y,x))$$

we might have written:[8]



Bourbaki rigorously developed Quine's brief notational suggestion.[9] And the resulting *Quine–Bourbaki notation* is evidently just as expressively powerful as our ordinary notation. However, if we adopt the Quine–Bourbaki notation, then we will not even be able to *ask* whether typographically distinct variables like 'x' and 'y' have different 'semantic roles', and Fine's antinomy will dissolve away.[10]

[6] Fine (2003: 610–14, 2007: 12–16) considers this thought, but does not consider the present point.

[7] Quine (1981: 69–70). See also Curry (1933: 389–90), Quine (1981: iv, 5, 71), Dummett (1981: ch.1), Kaplan (1986: 244), Lavine (2000: 5–6), and Potter (2000: 64).

[8] Quine (1981: §12).

[9] Bourbaki (1954: ch.1), apparently independently. The slight difference is that Bourbaki uses Hilbert's epsilon operator instead of quantifiers.

[10] Pickel and Rabern (2017: 148–52) consider and criticise the Quine–Bourbaki approach to Fine's antinomy. Pickel and Rabern assume that the Quine–Bourbaki approach will be coupled with Frege's idea that one obtains the predicate '( ) ≤ ( )' by taking a sentence like '7 ≤ 7' and deleting the names. They then insist that Frege must distinguish between the case when '( ) ≤ ( )' is regarded as a one-place predicate, and the case where it is regarded as a two-place predicate. And they then maintain: 'if Frege were to introduce marks capable of typographically distinguishing between these predicates, then that mark would need its own semantic significance, which in this context means designation.' We disagree with the last part of this claim. *Brackets* are semantically significant, in that $\neg(\varphi \wedge \psi)$ is importantly different from $(\neg\varphi \wedge \psi)$; but brackets do not denote. Fregeans should simply insist that any 'marks' on predicate-positions have a similarly *non-denotational* semantic significance. After all, their ultimate purpose is just to account for the different 'cross-referencing' in $\forall x \exists y \varphi(x,y)$ and $\forall x \exists y \varphi(y,x)$.

To be clear, no one is recommending that we *should adopt* the Quine–Bourbaki notation in practice: it would be hard to read and a pain to typeset. To dissolve the antimony of the variable, it is enough to know that we *could in principle* have adopted this notation.

But there is a catch. Just as this notation leaves us unable to formulate Fine's antinomy of the variable, it leaves us unable to define the notion of a variable-assignment. So, until we can provide a non-Tarskian approach to semantics, which does *not* essentially rely upon variable-assignments, we have no guarantee that we *could* have adopted the Quine–Bourbaki notation, even in principle. Now, we can of course use the Tarskian approach to supply a semantics for Quine–Bourbaki sentences derivatively.[11] But if we were to do that that, we would lose the right to say that we could, in principle, have done away with typographically distinct variables altogether, for we would still be relying upon them in our semantic machinery.

In sum, we want an approach to semantics which (unlike Tarski's) accords variables with no more apparent significance than is suggested by the Quine–Bourbaki notation. Fortunately, such approaches are available.

1.5 The Robinsonian approach to semantics

To recall: difficulties concerning the semantics for quantifiers arise because $\mathscr{L}$ may not contain names for every object in the domain. One solution to this problem is obvious: just *add* new constants. This was essentially Robinson's approach.[12]

To define how to *add* new symbols, it is easiest to define how to *remove* them. Given a structure $\mathcal{M}$, its $\mathscr{L}$-reduct is the $\mathscr{L}$-structure we obtain by *ignoring* the interpretation of the symbols in $\mathcal{M}$'s signature which are not in $\mathscr{L}$. More precisely:[13]

Definition 1.4: *Let $\mathscr{L}^+$ and $\mathscr{L}$ be signatures with $\mathscr{L}^+ \supseteq \mathscr{L}$. Let $\mathcal{M}$ be an $\mathscr{L}^+$-structure. Then $\mathcal{M}$*'s $\mathscr{L}$-reduct*, $\mathcal{N}$, is the unique $\mathscr{L}$-structure with domain M such that $s^{\mathcal{N}} = s^{\mathcal{M}}$ for all s from $\mathscr{L}$. We also say that $\mathcal{M}$ is a* signature-expansion *of $\mathcal{N}$, and that $\mathcal{N}$ is a* signature-reduct *of $\mathcal{M}$.*

In Quinean terms, the difference between a model and its reduct is not *ontological* but *ideological*.[14] We do not add or remove any entities from the domain; we just add or remove some (interpretations of) symbols.

---

[11] Where φ is any Quine–Bourbaki sentence, let φ^{fo} be the sentence of first-order logic which results by: (*a*) inserting the variable v_n after the n^{th} quantifier in φ, counting quantifiers from left-to-right; (*b*) replacing each blob connected to the n^{th}-quantifier with the variable v_n and (*c*) deleting all the connecting wires. Then say $\mathcal{M} \vDash \varphi$ iff $\mathcal{M} \vDash \varphi^{\text{fo}}$, with $\mathcal{M} \vDash \varphi^{\text{fo}}$ defined via the Tarskian approach.

[12] A. Robinson (1951: 19–21), with a tweak that one finds in, e.g., Sacks (1972: ch.4).

[13] Cf. Hodges (1993: 9ff) and Marker (2002: 31).

[14] Quine (1951: 14).

We can now define the idea of 'adding new constants for every member of the domain'. The following definition explains how to add, for each element $a \in M$, a new constant symbol, c_a, which is taken to name a:

Definition 1.5: *Let $\mathscr{L}$ be any signature. For any set M, $\mathscr{L}(M)$ is the signature obtained by adding to $\mathscr{L}$ a new constant symbol c_a for each $a \in M$. For any $\mathscr{L}$-structure $\mathcal{M}$ with domain M, we say that $\mathcal{M}^\circ$ is the $\mathscr{L}(M)$-structure whose $\mathscr{L}$-reduct is $\mathcal{M}$ and such that $c_a^{\mathcal{M}^\circ} = a$ for all $a \in M$.*

Since $\mathcal{M}^\circ$ is flooded with constants, it is very easy to set up its semantics. We start by defining the interpretation of the $\mathscr{L}(M)$-terms which contain no variables:

$$t^{\mathcal{M}^\circ} = f^{\mathcal{M}^\circ}(s_1^{\mathcal{M}^\circ}, \ldots, s_k^{\mathcal{M}^\circ}), \text{ if } t \text{ is the variable-free } \mathscr{L}(M)\text{-term } f(s_1, \ldots, s_k)$$

For each atomic first-order $\mathscr{L}(M)$-sentence, we then define:

$$\mathcal{M}^\circ \vDash t_1 = t_2 \text{ iff } t_1^{\mathcal{M}^\circ} = t_2^{\mathcal{M}^\circ}, \text{ for any variable-free } \mathscr{L}(M)\text{-terms } t_1, t_2$$

$$\mathcal{M}^\circ \vDash R(t_1, \ldots, t_n) \text{ iff } (t_1^{\mathcal{M}^\circ}, \ldots, t_n^{\mathcal{M}^\circ}) \in R^{\mathcal{M}^\circ}, \text{ for any variable-free } \mathscr{L}(M)\text{-terms } t_1, \ldots, t_n \text{ and any } n\text{-place relation symbol } R \text{ from } \mathscr{L}(M)$$

And finally we offer:

$$\mathcal{M}^\circ \vDash \neg\varphi \text{ iff } \mathcal{M}^\circ \nvDash \varphi$$

$$\mathcal{M}^\circ \vDash (\varphi \wedge \psi) \text{ iff } \mathcal{M}^\circ \vDash \varphi \text{ and } \mathcal{M}^\circ \vDash \psi$$

$$\mathcal{M}^\circ \vDash \forall x\varphi(x) \text{ iff } \mathcal{M}^\circ \vDash \varphi(c_a/x) \text{ for every } a \in M$$

We now have what we want, in terms of $\mathcal{M}^\circ$. And, since $\mathcal{M}^\circ$ is uniquely determined by $\mathcal{M}$, we can now extract what we really wanted: definitions concerning $\mathcal{M}$ itself. Where $\varphi(\bar{v})$ is a first-order $\mathscr{L}$-formula with free variables displayed, and $\bar{a}$ are from M, we define a *three*-place relation which, intuitively, says that $\varphi(\bar{v})$ is *true of* the entities $\bar{a}$ according to $\mathcal{M}$. Here is the definition:

$$\mathcal{M} \vDash \varphi(\bar{a}) \text{ iff } \mathcal{M}^\circ \vDash \varphi(\overline{c_a}/\bar{v})$$

The notation $\varphi(\bar{c}/\bar{v})$ indicates the $\mathscr{L}(M)$-formula obtained by substituting the k^{th} constant in the sequence $\bar{c}$ for the k^{th} variable in the sequence $\bar{v}$. So we have defined a *three*-place relation between an $\mathscr{L}$-formula, entities $\bar{a}$, and a structure $\mathcal{M}$, in terms of a *two*-place relation between a structure $\mathcal{M}^\circ$ and an $\mathscr{L}(M)$-formula. For readability, we will write $\varphi(\bar{c})$ instead of $\varphi(\bar{c}/\bar{v})$, where no confusion arises.

As a limiting case, a sentence is a formula with no free variables. So for each $\mathscr{L}$-sentence φ, our definition states that $\mathcal{M} \vDash \varphi$ iff $\mathcal{M}^\circ \vDash \varphi$. And, intuitively, we can read this as saying that φ is *true* in $\mathcal{M}$.

To complete the Robinsonian semantics, we will define something similar for *terms*. So, where $\bar{a}$ are entities from M and $t(\bar{v})$ is an $\mathscr{L}$-term with free variables displayed, we define a function $t^{\mathcal{M}} : M^n \longrightarrow M$, by:

$$t^{\mathcal{M}}(\bar{a}) = \big(t(\overline{c_a}/\bar{v})\big)^{\mathcal{M}^\circ}$$

This completes the Robinsonian approach. And the approach carries no taint of the antinomy of the variable, since it clearly accords variables with no more semantic significance than is suggested by the Quine–Bourbaki notation. Indeed, it is easy to give a Robinsonian semantics directly for Quine–Bourbaki sentences, via: $\mathcal{M}^\circ$ satisfies a Quine–Bourbaki sentence beginning with '$\forall$' *iff* for every $a \in M$ the model $\mathcal{M}^\circ$ satisfies the Quine–Bourbaki sentence which results from replacing all blobs connected to the quantifier with 'c_a' and then deleting the quantifier and the connecting wires.

1.6 Straining the notion of 'language'

For all its virtues, the Robinsonian approach has some eyebrow-raising features of its own. To define satisfaction for the sentences of the first-order $\mathscr{L}$-sentences, we have considered the sentences in some *other* formal languages, namely, those with signature $\mathscr{L}(M)$ for any $\mathscr{L}$-structure $\mathcal{M}$. These languages can be *enormous*. Let $\mathcal{M}$ be an infinite $\mathscr{L}$-structure, whose domain M has size κ for some very big cardinal κ.[15] Then $\mathscr{L}(M)$ contains at least κ symbols. Can such a beast really count as a *language*, in any intuitive sense?

Of course, there is no technical impediment to defining these enormous languages. So, if model theory is just regarded as a branch of *pure* mathematics, then there is no real reason to worry about any of this. But we might, instead, want model theory to be regarded as a branch of *applied* mathematics, whose (idealised) subject matter is the languages and theories that mathematicians *actually* use. And if we regard model theory that way, then we will not want our technical notion of a 'language' to diverge too far from the kinds of things which we would ordinarily count as languages.

There is a second issue with the Robinsonian approach. In Definition 1.5, we introduced a new constant symbol, c_a, for each $a \in M$. But we did not say what, exactly, the constant symbol c_a *is*. Robinson himself suggested that the constant c_a should just be the object *a itself*.[16] In that case, every object in $\mathcal{M}^\circ$ would name itself. But this is both philosophically strange and also technically awkward.

On the philosophical front: we might want to consider a structure, $\mathcal{W}$, whose domain is the set of all living wombats. In order to work out which sentences are

[15] As is standard, we use κ to denote a cardinal; see the end of §1.B for a brief review of cardinals.

[16] A. Robinson (1951: 21).

true in $\mathcal{W}$ using Robinson's own proposal, we would have to treat each wombat as a name for itself, and so imagine a language whose syntactic parts are live wombats.[17] This stretches the ordinary notion of a language to breaking point.

There is also a technical hitch with Robinson's own proposal. Suppose that c is a constant symbol of $\mathscr{L}$. Suppose that $\mathcal{M}$ is an $\mathscr{L}$-structure where the *symbol* c is itself an *element* of $\mathcal{M}$'s underlying domain. Finally, suppose that $\mathcal{M}$ interprets c as naming some element other than c itself, i.e. $c^{\mathcal{M}} \neq c$. Now Robinson's proposal requires that $c^{\mathcal{M}^\circ} = c$. But since $\mathcal{M}^\circ$ is a signature expansion of $\mathcal{M}$, we require that $c^{\mathcal{M}^\circ} = c^{\mathcal{M}}$, which is a contradiction.

To fix this bug whilst retaining Robinson's idea that $c_a = a$, we would have to tweak the definition of an $\mathscr{L}$-structure to ensure that the envisaged situation cannot arise.[18] A better alternative—which also spares the wombats—is to abandon Robinson's suggestion that $c_a = a$, and instead define the symbol c_a so that it is guaranteed *not* to be an element of $\mathcal{M}$'s underlying domain.[19] So this is our official Robinson*ian* semantics (even if it was not exactly Robinson's).

1.7 The Hybrid approach to semantics

Tarskian and Robinsonian semantics are technically equivalent, in the following sense: they use the same notion of an $\mathscr{L}$-structure, they use the same notion of an $\mathscr{L}$-sentence, and they end up defining exactly the same relation, $\models$, between structures and sentences. But, as we have seen, neither approach is exactly ideal. So we turn to a third approach: a *hybrid* approach.

In the Robinsonian semantics, we used $\mathcal{M}^\circ$ to define the expression $\mathcal{M} \models \varphi(\bar{a})$. Intuitively, this states that φ is true of $\bar{a}$ in $\mathcal{M}$. If we *start* by defining this notation—which we can do quite easily—then we can use it to present a semantics with the following recursion clauses:

$$\begin{aligned}
\mathcal{M} \models t_1 = t_2 &\text{ iff } t_1^{\mathcal{M}} = t_2^{\mathcal{M}}\text{, for any variable-free } \mathscr{L}\text{-terms } t_1, t_2\\
\mathcal{M} \models R(t_1, \ldots, t_n) &\text{ iff } (t_1^{\mathcal{M}}, \ldots, t_n^{\mathcal{M}}) \in R^{\mathcal{M}}\text{, for any variable-free } \mathscr{L}\text{-terms}\\
&\qquad t_1, \ldots, t_n \text{ and any } n\text{-place relation symbol } R \text{ from } \mathscr{L}\\
\mathcal{M} \models \neg\varphi &\text{ iff } \mathcal{M} \not\models \varphi\\
\mathcal{M} \models (\varphi \wedge \psi) &\text{ iff } \mathcal{M} \models \varphi \text{ and } \mathcal{M} \models \psi\\
\mathcal{M} \models \forall v \varphi(v) &\text{ iff } \mathcal{M} \models \varphi(a) \text{ for all } a \in M
\end{aligned}$$

[17] Cf. Lewis (1986: 145) on 'Lagadonian languages'.

[18] We would have to add a clause: if $\mathcal{M}$ is an $\mathscr{L}$-structure and $s \in \mathscr{L} \cap M$, then $s^{\mathcal{M}} = s$.

[19] A simple way to do this is as follows: let c_a be the ordered pair (a, M). By Foundation in the background set theory within which we implement our model theory, $(a, M) \notin M$.

All that remains is to define $\mathcal{M} \vDash \varphi(\bar{a})$ without going all-out Robinsonian. And the idea here is quite simple: we just add new constant symbols when we need them, but not before. Here is the idea, rigorously developed. Let $\mathcal{M}$ be an $\mathscr{L}$-structure with $\bar{a}$ from M. For each a_i among $\bar{a}$, let c_{a_i} be a constant symbol not occurring in $\mathscr{L}$. Intuitively, we interpret each c_{a_i} as a name for a_i. More formally, we define $\mathcal{M}[\bar{a}]$ to be a structure whose signature is $\mathscr{L}$ together with the new constant symbols among $\overline{c_a}$, whose $\mathscr{L}$-reduct is $\mathcal{M}$, and such that $c_{a_i}^{\mathcal{M}[\bar{a}]} = a_i$ for each i. Where $\varphi(\bar{v})$ is an $\mathscr{L}$-formula with free variables displayed, the Hybrid approach defines:

$$\mathcal{M} \vDash \varphi(\bar{a}) \text{ iff } \mathcal{M}[\bar{a}] \vDash \varphi(\overline{c_a}/\bar{v})$$

When we combine our new definition of $\mathcal{M} \vDash \varphi(a)$ with the clause for universal quantification, we see that universal quantification effectively amounts to considering all the different ways of expanding the signature of $\mathcal{M}$ with a *new* constant symbol which could be interpreted to name *any* element of M. (So the Hybrid approach offers a semantics by simultaneous recursion over structures and languages.) Finally, we offer a similar clause for terms:

$$t^{\mathcal{M}}(\bar{a}) = t^{\mathcal{M}[\bar{a}]}(\overline{c_a}/\bar{v})$$

thereby completing the Hybrid approach.[20]

1.8 Linguistic compositionality

Unsurprisingly, the Hybrid approach is technically equivalent to the Robinsonian and Tarskian approaches. However, its philosophical merits come out when we revisit some of the potential defects of the other approaches. The Tarskian approach does not distinguish sufficiently between names and variables; the Hybrid approach has no such issues. Indeed, just like the Robinsonian approach, the Hybrid approach accords variables with no greater semantic significance than is suggested by the Quine–Bourbaki notation. But the Robinsonian approach involved vast, peculiar 'languages'; the Hybrid approach has no such issues. And, following Lavine, we will pause on this last point.[21]

It is common to insist that languages should be *compositional*, in some sense. One of the most famous arguments to this effect is due to Davidson. Because natural languages are *learnable*, Davidson insists that 'the meaning of each sentence [must be] a function of a finite number of features of the sentence'. For, on the one hand,

[20] The hybrid approach is hinted at by Geach (1962: 160), and Mates (1965: 54–7) offers something similar. But the clearest examples we can find are Boolos and Jeffrey (1974: 104–5), Boolos (1975: 513–4), and Lavine (2000: 10–12).

[21] See Lavine's (2000: 12–13) comments on compositionality and learnability.

if a language has this feature, then we 'understand how an infinite aptitude can be encompassed by finite accomplishments'. Conversely, if 'a language lacks this feature then no matter how many sentences a would-be speaker learns to produce and understand, there will remain others whose meanings are not given by the rules already mastered.'[22]

Davidson's argument is too quick. After all, it is a *wild* idealisation to suggest that any actual human can indeed understand or learn the meanings of *infinitely* many sentences: some sentences are just too long for any actual human to parse. It is unclear, then, why we should worry about the 'learnability' of such sentences.

Still, something in the *vicinity* of Davidson's argument seems right. In §1.6, we floated the idea that model theory should be regarded as a branch of *applied* mathematics, whose (idealised) subject matter is the languages and theories that (pure) mathematicians *actually* use. But here is an apparent phenomenon concerning that subject matter: once we have a fixed interpretation in mind, we tend to act as if that interpretation fixes the truth value of *any* sentence of the appropriate language, no matter how long or complicated that sentence is.[23] All three of our approaches to formal semantics accommodate this point. For, given a signature $\mathscr{L}$ and an $\mathscr{L}$-structure $\mathcal{M}$—i.e. an interpretation of the range of quantification and an interpretation of each $\mathscr{L}$-symbol—the semantic value of every $\mathscr{L}$-sentence is completely determined within $\mathcal{M}$, in the sense that, for every $\mathscr{L}$-sentence φ, either $\mathcal{M} \vDash \varphi$, or $\mathcal{M} \vDash \neg\varphi$, but not both.

But the Hybrid approach, specifically, may allow us to go a little further. For, when $\mathscr{L}$ is finite,[24] and we offer the Hybrid approach to semantics, we may gain some insight into how a finite mind might *fully understand* the rules by which an interpretation fixes the truth-value of every sentence. That understanding seems to reduce to three rather tractable components:

(a) an understanding of the finitely many recursion clauses governing satisfaction for atomic sentences (finitely many, as we assumed that $\mathscr{L}$ is finite);
(b) an understanding of the handful of recursion clauses governing sentential connectives; and
(c) an understanding of the recursion clauses governing quantification

On the Hybrid approach, point (c) reduces to an understanding of two ideas: (i) the *general* idea that names can pick out objects,[25] and (ii) the intuitive idea that, for any object, we could expand our language with a new name for that object. In short:

---

[22] Davidson (1965: 9).

[23] A theme of Part B is whether, in certain circumstances, axioms can also fix truth values.

[24] We can make a similar point if $\mathscr{L}$ can be recursively specified.

[25] There are some deep philosophical issues concerning the question of how names pick out objects (see Chapters 2 and 15). However, the general notion seems to be required by *any* model-theoretic semantics, so that there is no *special* problem here for the Hybrid approach.

the Hybrid semantics seems to provide a truly *compositional* notion of meaning. But we should be clear on what this means.

First, we are not aiming to escape what Sheffer once called the 'logocentric predicament', that '*In order to give an account of logic, we must presuppose and employ logic.*'[26] Our semantic clause for object-language conjunction, $\wedge$, always involves conjunction in the metalanguage. On the Hybrid approach, our semantic clause for object-language universal quantification, $\forall$, involved (metalinguistic) quantification over all the ways in which a new name could be added to a signature. We do not, of course, claim that anyone could read these semantic clauses and come to *understand* the very idea of conjunction or quantification from scratch. We are making a much more mundane point: to understand the hybrid approach to semantics, one need only understand a tractable number of ideas.

Second, in describing our semantics as compositional, we are *not* aiming to supply a semantics according to which the meaning of $\forall xF(x)$ depends upon the separate meanings of the expressions $\forall$, x, F, and x.[27] Not only would that involve an oddly inflexible understanding of the word 'compositional'; the discussion of §1.4 should have convinced us that variables do not have semantic values in isolation.[28] Instead, on the hybrid approach, the meaning of $\forall xF(x)$ depends upon the meanings of the quantifier-expression $\forall x \ldots x$ and the predicate-expression $F(\ )$. The crucial point is this: the Hybrid approach delivers the truth-conditions of infinitely many sentences using only a small 'starter pack' of principles.

Having aired the virtues of the Hybrid approach, though, it is worth repeating that our three semantic approaches are technically equivalent. As such, we can in good faith use whichever approach we like, whilst claiming all of the pleasant philosophical features of the Hybrid approach. Indeed, in the rest of this book, we simply use whichever approach is easiest for the purpose at hand.

This concludes our discussion of first-order logic. It also concludes the 'philosophical' component of this chapter. The remainder of this chapter sets down the purely technical groundwork for several later philosophical discussions.

1.9 Second-order logic: syntax

Having covered first-order logic, we now consider *second*-order logic. This is much less popular than first-order logic among working model-theorists. However, it has

[26] Sheffer (1926: 228).

[27] Pickel and Rabern (2017: 155) call this 'structure intrinsicalism', and advocate it.

[28] Nor would it help to suggest that the meaning of $\forall xF(x)$ depends upon the separate meanings of the two composite expressions $\forall x$ and $F(x)$. For if we think that open formulas possess semantic values (in isolation), we will obtain an exactly parallel (and exactly as confused) 'antinomy of the open formula' as follows: clearly $F(x)$ and $F(y)$ are notational variants, and so should have the same semantic value; but they cannot have the same value, since $F(x) \wedge \neg F(y)$ is not a contradiction.

certain philosophically interesting dimensions. We explore these philosophical issues in later chapters; here, we simply outline its technicalities.

First-order logic can be thought of as allowing quantification into *name* position. For example, if $\varphi(c)$ is a formula containing a constant symbol c, then we also have a formula $\forall v\varphi(v/c)$, replacing c with a variable which is bound by the quantifier. To extend the language, we can allow quantification into *relation symbol* or *function symbol* position. For example, if $\varphi(R)$ is a formula containing a relation symbol R, we would want to have a formula $\forall X\varphi(X/R)$, replacing the relation symbol R with a relation-variable, X, which is bound by the quantifier. Equally, if $\varphi(f)$ is a formula containing a function symbol f, we would want to have a formula $\forall p\varphi(p/f)$.

Let us make this precise, starting with the syntax. In addition to all the symbols of first-order logic, our language adds some new symbols:

- relation-variables: U, V, W, X, Y, Z
- function-variables: p, q

both with numerical subscripts and superscripts as necessary. In more detail: just like relation symbols and functions symbols, these higher-order variables come equipped with a number of places, indicated (where helpful) with superscripts. So, together with the subscripts, this means we have countably many relation-variables and function-symbols for each number of places. We then expand the recursive definition of a term, to allow:

- $q^n(t_1, \ldots, t_n)$, for any $\mathscr{L}$-terms $t_1, \ldots, t_n$ and n-place function-variable q^n

and we expand the notion of a formula, to allow

- $X^n(t_1, \ldots, t_n)$, for any $\mathscr{L}$-terms $t_1, \ldots, t_n$ and n-place relation-variable X^n
- $\exists X^n\varphi$ and $\forall X^n\varphi$, for any n-place relation-variable X^n and any second-order $\mathscr{L}$-formula φ which contains neither of the expressions $\exists X^n$ nor $\forall X^n$
- $\exists q^n\varphi$ and $\forall q^n\varphi$, for any n-place function-variable q^n and any second-order $\mathscr{L}$-formula φ which contains neither of the expressions $\exists q^n$ nor $\forall q^n$

We will also introduce some abbreviations which are particularly helpful in a second-order context. Where Ξ is any one-place relation symbol or relation-variable, we write $(\forall x : \Xi)\varphi$ for $\forall x(\Xi(x) \to \varphi)$, and $(\exists x : \Xi)\varphi$ for $\exists x(\Xi(x) \wedge \varphi)$. We also allow ourselves to bind multiple quantifiers at once; so $(\forall x, y, z : \Xi)\varphi$ abbreviates $\forall x\forall y\forall z((\Xi(x) \wedge \Xi(y) \wedge \Xi(z)) \to \varphi)$.

1.10 Full semantics

The syntax of second-order logic is straightforward. The semantics is more subtle; for here there are some genuinely *non*-equivalent options.

We start with *full semantics* for second-order logic (also known as *standard* semantics). This uses $\mathscr{L}$-structures, exactly as we defined them in Definition 1.2.

The trick is to add new semantic clauses for our second-order quantifiers. In fact, we can adopt any of the Tarskian, Robinsonian, or Hybrid approaches here, and we sketch all three (leaving the reader to fill in some obvious details).

Tarskian. Variable-assignments are the key to the Tarskian approach to first-order logic. So the Tarskian approach to second-order logic must expand the notion of a variable-assignment, to cover both relation-variables and function-variables. In particular, we take it that σ is a function which assigns every variable to some entity $a \in M$, every n-place relation-variable to some subset of M^n, and every function-variable to some function $M^n \longrightarrow M$. We now add clauses:

$$\begin{aligned}
\mathcal{M}, \sigma \vDash X^n(t_1, \ldots, t_n) \text{ iff } (t_1^{\mathcal{M},\sigma}, \ldots, t_n^{\mathcal{M},\sigma}) \in (X^n)^{\mathcal{M},\sigma} \text{ for any} \\
\mathscr{L}\text{-terms } t_1, \ldots, t_n \\
\mathcal{M}, \sigma \vDash \forall X^n \varphi(X^n) \text{ iff } \mathcal{M}, \tau \vDash \varphi(X^n) \text{ for every variable-assignment } \tau \\
\text{which agrees with } \sigma \text{ except perhaps on } X^n \\
\mathcal{M}, \sigma \vDash \forall q^n \varphi(q^n) \text{ iff } \mathcal{M}, \tau \vDash \varphi(q^n) \text{ for every variable-assignment } \tau \\
\text{which agrees with } \sigma \text{ except perhaps on } q^n
\end{aligned}$$

Robinsonian. The key to the Robinsonian approach to first-order logic is to introduce a new constant symbol for every entity in the domain. So the Robinsonian approach to second-order logic must introduce a new relation symbol for every possible relation on M, and a new function symbol for every possible function. Let $\mathcal{M}^\blacksquare$ be the structure which expands $\mathcal{M}$ in just this way. So, for each n and each $S \subseteq M^n$, we add a new relation symbol R_S with $S = R_S^{\mathcal{M}^\blacksquare}$, and for each function $g : M^n \longrightarrow M$ we add a new function symbol f_g with $g = f_g^{\mathcal{M}^\blacksquare}$. We can now simply rewrite the first-order semantics, replacing $\mathcal{M}^\circ$ with $\mathcal{M}^\blacksquare$, and adding:

$$\begin{aligned}
\mathcal{M}^\blacksquare \vDash \forall X^n \varphi(X^n) \text{ iff } \mathcal{M}^\blacksquare \vDash \varphi(R_S/X^n) \text{ for every } S \subseteq M^n \\
\mathcal{M}^\blacksquare \vDash \forall q^n \varphi(q^n) \text{ iff } \mathcal{M}^\blacksquare \vDash \varphi(f_g/q^n) \text{ for every function } g : M^n \longrightarrow M
\end{aligned}$$

Hybrid. The key to the Hybrid approach to second-order logic is to define, upfront, the three-place relation between $\mathcal{M}$, a formula φ, and a relation (or function) on $\mathcal{M}$.[29] We illustrate the idea for the case of relations (the case of functions is exactly similar). Let S be a relation on M^n. Let R_S be an n-place relation symbol not occurring in $\mathscr{L}$. We define $\mathcal{M}[S]$ to be a structure whose signature is $\mathscr{L}$ together with the new relation symbol R_S, such that $\mathcal{M}[S]$'s $\mathscr{L}$-reduct is $\mathcal{M}$ and $R_S^{\mathcal{M}[S]} = S$. Then where $\varphi(X)$ is an $\mathscr{L}$-formula with free relation-variable displayed, we define:

$$\begin{aligned}
\mathcal{M} \vDash \varphi(S) \text{ iff } \mathcal{M}[S] \vDash \varphi(R_S/X) \text{ for any relation symbol } R_S \notin \mathscr{L} \\
\mathcal{M} \vDash \forall X^n \varphi(X^n) \text{ iff } \mathcal{M} \vDash \varphi(S) \text{ for every relation } S \subseteq M^n
\end{aligned}$$

[29] Trueman (2012) recommends a semantics like this as a means for overcoming philosophical resistance to the use of second-order logic.

The three approaches ultimately define the same semantic relation. And we call the ensuing semantics *full* second-order semantics.

The relative merits of these three approaches are much as before. So: the Tarskian approach unhelpfully treats relation-variables as if they were varying predicates; the Robinsonian approach forces us to stretch the idea of a language to breaking point; but the Hybrid approach avoids both problems and provides us with a reasonable notion of compositionality. (It is worth noting, though, that all three approaches effectively assume that we understand notions like 'all subsets of M^n'. We revisit this point in Part B.)

1.11 Henkin semantics

The Tarskian, Robinsonian, and Hybrid approaches all yielded the same relation, $\models$. However, there is a *genuinely alternative* semantics for second-order logic. Moreover, the availability of this alternative is an important theme in Part B of this book. So we outline that alternative here.

In *full* second-order logic, universal quantification into relation-position effectively involves considering *all possible* relations on the structure. Indeed, using $\wp(A)$ for A's powerset, i.e. $\{B : B \subseteq A\}$, we have the following: if X is a one-place relation-variable, then the relevant 'domain' of quantification in $\forall X\varphi$ is $\wp(M)$; and if X is an n-place relation-variable, then the relevant 'domain' of quantification in $\forall X\varphi$ is $\wp(M^n)$. An alternative semantics naturally arises, then, by considering more *restrictive* 'domains' of quantification, as follows:

Definition 1.6: *A* Henkin $\mathscr{L}$-structure, $\mathcal{M}$, *consists of:*

- *a non-empty set, M, which is the underlying domain of* $\mathcal{M}$
- *a set* $M_n^{\text{rel}} \subseteq \wp(M^n)$ *for each* $n < \omega$
- *a set* $M_n^{\text{fun}} \subseteq \{g \in \wp(M^{n+1}) : g \text{ is a function } M^n \longrightarrow M\}$ *for each* $n < \omega$
- *an object* $c^{\mathcal{M}} \in M$ *for each constant symbol c from* $\mathscr{L}$
- *a relation* $R^{\mathcal{M}} \subseteq M^n$ *for each n-place relation symbol R from* $\mathscr{L}$
- *a function* $f^{\mathcal{M}} : M^n \longrightarrow M$ *for each n-place function symbol f from* $\mathscr{L}$.

In essence, M_n^{rel} serves as the domain of quantification for the n-place relation-variables, and M_n^{fun} serves as the domain of quantification for the n-place function-variables. As before, though, we can make this idea precise using any of our three approaches to formal semantics. We sketch all three.

Tarskian. Where $\mathcal{M}$ is a Henkin structure, we take our variable-assignments σ to be restricted in the following way: σ assigns each variable to some entity $a \in M$, each n-place relation-variable to some element of M_n^{rel}, and each n-place function-variable to some element of M_n^{fun}. We then rewrite the clauses for the full semantics,

exactly as before, but using this more restricted notion of a variable-assignment.

Robinsonian. Where $\mathcal{M}$ is a Henkin structure, we let $\mathcal{M}^\square$ be the structure which expands $\mathcal{M}$ by adding new relation symbols R_S such that $S = R_S^{\mathcal{M}^\square}$ for every relation $S \in M_n^{\text{rel}}$, and new function symbols f_g such that $g = f_g^{\mathcal{M}^\square}$ for every function $g \in M_n^{\text{fun}}$. We then offer these clauses:

$$\mathcal{M}^\square \vDash \forall X^n \varphi(X^n) \text{ iff } \mathcal{M}^\square \vDash \varphi(R_S/X^n) \text{ for every relation } S \in M_n^{\text{rel}}$$
$$\mathcal{M}^\square \vDash \forall q^n \varphi(q^n) \text{ iff } \mathcal{M}^\square \vDash \varphi(f_g/q^n) \text{ for every function } g \in M_n^{\text{fun}}$$

Hybrid. We need only tweak the recursion clauses, as follows:

$$\mathcal{M} \vDash \forall X^n \varphi(X^n) \text{ iff } \mathcal{M} \vDash \varphi(S) \text{ for every relation } S \subseteq M_n^{\text{rel}}$$
$$\mathcal{M} \vDash \forall q^n \varphi(q^n) \text{ iff } \mathcal{M} \vDash \varphi(g) \text{ for every function } g \in M_n^{\text{fun}}$$

We say that *Henkin semantics* is the semantics yielded by any of these three approaches, as applied to Henkin structures. Importantly, Henkin semantics generalises the *full* semantics of §1.10. To show this, let $\mathcal{M}$ be an $\mathscr{L}$-structure in the sense of Definition 1.2. From this, define a Henkin structure $\mathcal{N}$ by setting, for each $n < \omega$, $N_n^{\text{rel}} = \wp(N^n)$ and N_n^{fun} as the set of all functions $N^n \longrightarrow N$. Then *full* satisfaction, defined over $\mathcal{N}$, is exactly like *Henkin* satisfaction, defined over $\mathcal{N}$.

The notion of a Henkin structure may, though, be a bit *too* general. To see why, consider a Henkin $\mathscr{L}$-structure $\mathcal{M}$, and suppose that R is a one-place relation symbol of $\mathscr{L}$, so that $R^{\mathcal{M}} \subseteq M$. Presumably, we should want $\mathcal{M}$ to satisfy $\exists X \forall v(R(v) \leftrightarrow X(v))$, for $R^{\mathcal{M}}$ should *itself* provide a witness to the second-order existential quantifier. But this holds if and only if $R^{\mathcal{M}} \in M_1^{\text{rel}}$, and the definition of a Henkin structure does not guarantee this. For this reason, it is common to insist that the following axiom schema should hold in all structures:

Comprehension Schema. *$\exists X^n \forall \vec{v}(\varphi(\vec{v}) \leftrightarrow X^n(\vec{v}))$, for every formula $\varphi(\vec{v})$ which does not contain the relation-variable X^n*

We must block X^n from appearing in $\varphi(\vec{v})$, since otherwise an axiom would be $\exists X \forall v(\neg X(v) \leftrightarrow X(v))$, which will be inconsistent. However, we allow other free first-order and second-order variables, because this allows us to form new concepts from old concepts. For instance, given the two-place relation symbol R, we have as an axiom $\exists X^2 \forall v_1 \forall v_2(\neg R(v_1, v_2) \leftrightarrow X^2(v_1, v_2))$, i.e. M_2^{rel} must contain the set of all pairs not in $R^{\mathcal{M}}$, i.e. $M^2 \setminus R^{\mathcal{M}}$. So: if we insist that (all instances) of the Comprehension Schema must hold in all Henkin structures, then we are insisting on further properties concerning our various M_n^{rel}s. There is also a *predicative* version of Comprehension:

Predicative Comprehension Schema. *$\exists X^n \forall \vec{v}(\varphi(\vec{v}) \leftrightarrow X^n(\vec{v}))$, for every formula $\varphi(\vec{v})$ which neither contains the relation-variable X^n nor any second-order quantifiers*

When we want to draw the contrast, we call the (plain vanilla) Comprehension Schema the *Impredicative* Comprehension Schema. But this will happen only rarely; we only mention Predicative Comprehension in §§5.7, 10.2, 10.C, and 11.3.

We could provide a similar schema to govern functions. But it is usual to make the stronger claim, that the following should hold in all structures (for every n):[30]

Choice Schema. $\forall X^{n+1}\left(\forall \bar{v}\exists y\, X^{n+1}(\bar{v}, y) \rightarrow \exists p^{n}\forall \bar{v}\, X^{n+1}(\bar{v}, p^{n}(\bar{v}))\right)$

To understand these axioms, let S be a two-place relation on the domain, and suppose that the antecedent is satisfied, i.e. that for any x there is some y such that $S(x, y)$. The relevant Choice instance then states that there is then a one-place function, p, which 'chooses', for each x, a *particular* entity $p(x)$ such that $S(x, p(x))$. For obvious reasons, this p is known as a *choice function*. Hence, just like the Comprehension Schema, the Choice Schema guarantees that the domains of the higher-order quantifiers are well populated.

This leads to a final definition: a *faithful Henkin structure* is a Henkin structure within which both (impredicative) Comprehension and Choice hold.[31]

1.12 Consequence

We have defined satisfaction for first-order logic and for both the full- and Henkin-semantics for second-order logic. However, any definition of satisfaction induces a notion of consequence, via the following:

Definition 1.7: *A* theory *is a set of sentences in the logic under consideration. Given a structure $\mathcal{M}$ and a theory T, we say that $\mathcal{M}$ is a* model *of T, or more simply $\mathcal{M} \models T$, iff $\mathcal{M} \models \varphi$ for all sentences φ from T. We say that T has φ as a* consequence*, or that T entails φ, or more simply just $T \models \varphi$, iff: if $\mathcal{M} \models T$ then $\mathcal{M} \models \varphi$ for all structures $\mathcal{M}$.*

Note that this definition is relative to a semantics. So there are as many notions of logical consequence as there are semantics.

Here are some examples to illustrate the notation. Consider the natural numbers $\mathcal{N}$ and the integers $\mathcal{Z}$ in the signature consisting just of the symbol $<$, where this is given its natural interpretation. It is easy to see that both structures satisfy the following axioms:

$$\forall x\forall y\forall z\big((x < y \wedge y < z) \rightarrow x < z\big)$$
$$\forall x(x \not< x)$$
$$\forall x\forall y(x < y \vee x = y \vee y < x)$$

[30] For more, see Shapiro (1991: 67).

[31] See e.g. Shapiro (1991: 98–9).

These are the axioms of a *linear order*. Let T_{LO} be the theory consisting of just these three axioms. Then we would write $\mathcal{N} \vDash T_{\mathrm{LO}}$ and $\mathcal{Z} \vDash T_{\mathrm{LO}}$. But if we drop the third axiom, we obtain the related notion of a *partial order*. For an example of a partial order which is not a linear order, consider any set X with more than two elements, and consider the structure $\mathcal{P}$ whose first-order domain is the powerset $\wp(X)$ of X, with $<$ interpreted in $\mathcal{P}$ as the subset relation. If a, b are distinct elements of X, then $\mathcal{P} \vDash \{a\} \not< \{b\} \wedge \{a\} \neq \{b\} \wedge \{b\} \not< \{a\}$. So $\mathcal{P} \nvDash T_{\mathrm{LO}}$.

1.13 Definability

In addition to a notion of consequence, a semantics will induce a notion of definability, as follows:

Definition 1.8: *Let $\mathcal{M}$ be any structure and $n \geq 1$. We say that a subset X of M^n is* definable *iff there is both a formula $\varphi(v_1, \ldots, v_n, x_1, \ldots, x_m)$ with all free variables displayed and also elements $b_1, \ldots, b_m \in M$ such that:*

$$X = \{(a_1, \ldots, a_n) \in M^n : \mathcal{M} \vDash \varphi(a_1, \ldots, a_n, b_1, \ldots, b_m)\}$$

Here, the elements $b_1, \ldots, b_m$ are called *parameters*. Many authors allow parameters to be tacitly suppressed, and so say that X is definable iff $X = \{(a_1, \ldots, a_n) \in M^n : \mathcal{M} \vDash \varphi(a_1, \ldots, a_n)\}$ for some $\varphi(v_1, \ldots, v_n)$ which is (tacitly) allowed to contain further unmentioned parameters. If parameters are not allowed, such authors typically say this explicitly. We will be similarly explicit. When parameters are not allowed, the resulting sets are called *parameter-free definable sets*. Clearly a set is $\mathcal{M}$-definable iff it is parameter-free definable in some signature-expansion of $\mathcal{M}$ (see Definition 1.4).

To illustrate the idea of definability, consider again the natural numbers $\mathcal{N}$ in the signature consisting just of $<$, again with its natural interpretation. Here is a simple definable set:

$$\{0\} = \{n \in N : \mathcal{N} \vDash \neg \exists x\, x < n\}$$

As a slightly more complicated example, the graph of the successor operation in $\mathcal{N}$ is definable, since intuitively $n = m + 1$ iff m is less than n and there is no natural number strictly between m and n. More precisely:

$$G = \left\{(n, m) \in N^2 : \mathcal{N} \vDash (m < n \wedge \neg \exists z (m < z < n))\right\}$$

Now, both of these sets are *parameter-free* definable. And so it follows that *all* definable sets over $\mathcal{N}$ are parameter-free definable. For, where S is the successor function

on the natural numbers, each natural number n is equal to the term $S^n(0)$, which we define recursively as follows:

$$S^0(a) = a \qquad S^{n+1}(a) = S(S^n(a)) \qquad (\textit{numerals})$$

(We label this definition '(*numerals*)' for future reference.) Hence, to say that $2 = S^2(0)$ is just a fancy way of saying that two is the second successor of zero. The terms $S^n(0)$ are sometimes called the *numerals*, and clearly $\mathcal{N} \vDash n = S^n(0)$ for each natural number $n \geq 0$. So, we can explicitly define the numerals in terms of the less-than relation using G, any definable set on $\mathcal{N}$ is *parameter-free* definable, by the following:

$$\begin{aligned}&\{(a_1, \ldots, a_n) \in N^n : \mathcal{N} \vDash \varphi(a_1, \ldots, a_n, b_1, \ldots, b_m)\}\\ =&\{(a_1, \ldots, a_n) \in N^n : \mathcal{N} \vDash \varphi(a_1, \ldots, a_n, S^{b_1}(0), \ldots, S^{b_m}(0))\}\end{aligned}$$

For an example of a structure with definable sets which are not parameter-free definable, let $\mathscr{L}$ be a countable signature and let $\mathcal{M}$ be an uncountable $\mathscr{L}$-structure. Since there are only countably many $\mathscr{L}$-formulas, there are only countably many parameter-free definable sets. But trivially the singleton $\{a\}$ of any element a from M is definable, as $\{a\} = \{x \in M : \mathcal{M} \vDash x = a\}$. So $\mathcal{M}$ has uncountably many definable subsets which are not parameter-free definable.

Finally, it is worth mentioning a particular aspect of definability in second-order logic. Consider the natural numbers $\mathcal{N}$ in the full semantics, and consider the set $\{(n, A) \in N \times \wp(N) : \mathcal{N} \vDash A(n)\}$ consisting of all pairs of numbers and sets of numbers such that the number is in the set. It obviously makes good sense to say that this set is definable, even though it is not a subset of $N \times N$ but rather of $N \times \wp(N)$. So, in the case of second-order logic, we expand the notion of definability to include both subsets of products of the *second-order* domain, and subsets of products of the first-order domain and the second-order domain. This point holds for both the Henkin and the full semantics.

1.A First- and second-order arithmetic

We have laid down the syntax and semantics for the logics which occupy us throughout this book. However, we will frequently discuss certain specific mathematical theories. So, for ease of reference, in this appendix we lay down the usual first- and second-order axioms of arithmetic. We cover set theory in the next appendix, and reserve all philosophical commentary for later chapters.

Definition 1.9: *The theory of Robinson Arithmetic,* Q*, is given by the universal closures of the following eight axioms:*

(Q1) $S(x) \neq 0$
(Q2) $S(x) = S(y) \to x = y$
(Q3) $x \neq 0 \to \exists y\, x = S(y)$
(Q4) $x + 0 = x$
(Q5) $x + S(y) = S(x + y)$
(Q6) $x \times 0 = 0$
(Q7) $x \times S(y) = (x \times y) + x$
(Q8) $x \leq y \leftrightarrow \exists z\, x + z = y$

The theory of Peano Arithmetic, PA*, is given by adding to Robinson Arithmetic the following Induction Schema:*

$$[\varphi(0) \wedge \forall y\,(\varphi(y) \to \varphi(S(y)))] \to \forall y \varphi(y)$$

While PA obviously formalises an important part of number-theoretic practice, it was axiomatised only in 1934.[32] We now turn to second-order arithmetic:

Definition 1.10: *The theory of* second-order Peano arithmetic, PA_2*, is given by axioms (Q1)–(Q3) of Definition 1.9, the Comprehension Schema of §1.11, and the following mathematical Induction Axiom:*

$$\forall X([X(0) \wedge \forall y(X(y) \to X(S(y)))] \to \forall y X(y))$$

With the exception of the Comprehension Schema, the axioms of PA_2 were first explicitly written down by Dedekind.[33] The Choice Schema is typically not built into axiomatisations of PA_2, although it is valid on the standard semantics.[34]

Note that the signature of PA_2 is just $\{0, S\}$, whereas the signature of the first-order theory PA is $\{0, S, <, +, \times\}$. However, in the setting of PA_2, order, addition and multiplication are explicitly definable in the sense of Definition 1.8. For instance, the graph of the addition function is the unique three-place relation which is the union of all three-place relations satisfying the following condition, which intuitively describes an initial segment of the graph of addition:

$$\Phi(B) := \forall x B(x, 0, x) \wedge \forall x \forall y \forall w [B(x, S(y), w) \to \exists z\,(w = S(z) \wedge B(x, y, z))]$$

By Comprehension, there is a three-place relation A satisfying $A(a, b, c)$ iff $\exists B(\Phi(B) \wedge B(a, b, c))$. If we then define $a + b = c$ by $A(a, b, c)$ we can easily show by induction that this satisfies axioms (Q4)–(Q5) of Definition 1.9. An analogous definition can be presented in second-order logic for a formula which satisfies axioms (Q6)–(Q7). Finally, obviously (Q8) allows $\leq$ to be explicitly defined in terms of addition and first-order logic.

[32] Hilbert and Bernays (1934). For contemporary references on PA and its subsystems, see e.g. Kaye (1991) and Hájek and Pudlák (1998).

[33] Dedekind (1888).

[34] A contemporary reference on PA_2 and its subsystems is Simpson (2009).

1.B First- and second-order set theory

We now turn to set theory. The signature of set theory consists just of the binary relation $\in$, where we read $x \in y$ as 'x is a member of y'. We start with the following axioms, which we state slightly informally, leaving the reader to transcribe them into sentences of first-order logic if she wishes. Here and throughout, $(\forall y \in x)\varphi$ abbreviates $\forall y(y \in x \to \varphi)$ and $(\exists y \in x)\varphi$ abbreviates $\exists y(y \in x \wedge \varphi)$.

Extensionality. *For all x and y, we have: $x = y$ iff $\forall z\,(z \in x \leftrightarrow z \in y)$*
Pairing. *For all x and y, there is a unique set, $\{x, y\}$, such that for all z: $z \in \{x, y\}$ iff either $z = x$ or $z = y$*
Union. *For all x, there is a unique set, $\bigcup x$, , such that for all z: $z \in \bigcup x$ iff $(\exists y \in x)z \in y$*
Power Set. *For all x, there is a unique set, $\wp(x)$, such that for all z: $z \in \wp(x)$ iff $z \subseteq x$*
Separation Schema. *For all x and $\bar{v}$ there is a unique set, $\{y \in x : \varphi(y, \bar{v})\}$, such that for all z: $z \in \{y \in x : \varphi(y, \bar{v})\}$ iff both $z \in x$ and $\varphi(z, \bar{v})$*

In the Separation Schema, there is one axiom for each formula $\varphi(y, \bar{v})$ in the signature. It is worth noting that the uniqueness claims in Pairing, Union, Power Set, and the Separation Schema are redundant, given Extensionality,[35] and that the left-to-right directions of the biconditionals in Pairing, Union, and Power Set are redundant, given the Separation Schema. For instance, suppose that for all x and y there is some v such that if $z = x$ or $z = y$ then $z \in v$. Then $\{z \in v : z = x \vee z = y\}$ exists by Separation and is obviously equal to $\{x, y\}$.

Using these axioms, we define $\varnothing$ as the unique set with no members; the empty set. Whilst there are *philosophical* discussions to have about $\varnothing$'s existence,[36] there are no *technical* discussions to be had. The usual background axioms for first-order logic assert that there exists at least one object x, and applying Separation to the formula $z \neq z$ we obtain a set $\varnothing$ such that, $\forall z\,(z \in \varnothing \leftrightarrow (z \in x \wedge z \neq z))$, from which it follows by elementary logic that $\forall z\, z \notin \varnothing$. The uniqueness of the empty set then follows from Extensionality.

The intersection of x, written $\bigcap x$, is the set whose members elements are exactly those which are members of every element of x. This exists whenever x is non-empty, since $\bigcap x = \{y \in \bigcup x : (\forall z \in x)y \in z\}$, which exists by Union and Separation. The usual binary operations of union $x \cup y$ and intersection $x \cap y$ can then be defined via $x \cup y = \bigcup\{x, y\}$ and $x \cap y = \bigcap\{x, y\}$. Finally, the singleton $\{x\}$ is defined to be $\{x, x\}$ and is the set whose unique member is x.

We define the successor $s(x)$ of x to be the set $x \cup \{x\}$, so that $z \in s(x)$ iff either $z = x$ or $z \in x$. This notation allows us to state another axiom:

Infinity. *There is a set w such that $\varnothing \in w$ and for all x, if $x \in w$ then $s(x) \in w$*

[35] For philosophical commentary on uniqueness, see Potter (2004: 258–9).
[36] See e.g. Oliver and Smiley (2006: 126–32).

The empty set $\varnothing$ plays a role in set theory similar to the role zero plays in arithmetic, and the successor function s in set theory is similar to the successor function S from the axioms of Definition 1.9. In these terms, the Infinity Axiom says that there is a set which contains the ersatz of zero and is closed under the ersatz of successor.

Using the intersection operation, defined above, we can also state another axiom, whose role is to rule out infinite descending membership chains:

Foundation. *For every non-empty set x there is some $z \in x$ such that $z \cap x = \varnothing$*

After all, if an infinite chain $\ldots \in x_n \in \ldots \in x_2 \in x_1 \in x_0$ existed, then the non-empty set $x = \{x_0, x_1, x_2, \ldots, x_n, \ldots\}$ would violate Foundation.

Introducing the usual notation $\exists! x \varphi$ to abbreviate $\exists x \forall v (\varphi \leftrightarrow x = v)$, for any variable v not occurring in φ, we lay down an axiom schema which, intuitively, states that the image of any set under a function is a set:

Replacement Schema. *For all w and all $\bar{v}$: if $(\forall x \in w) \exists! y \varphi(x, y, \bar{v})$, then $\exists z (\forall x \in w)(\exists y \in z) \varphi(x, y, \bar{v})$*

Finally, we lay down an axiom stating that any set can be equipped with a binary relation that satisfies the axioms of a well-order:

Choice. *Any set can be well-ordered*

A well-order is a linear order such that any non-empty set of ordered elements has a least element. (The axioms of a linear order were given in §1.12.) Note that Choice, here, is a single axiom, expressed in first-order logic with an additional primitive, $\in$. This single Axiom should *not* be confused with the Choice Schema for second-order logic, as laid down in §1.11, which yields infinitely many second-order sentences. That said, there is evidently a connection between the Axiom and the Schema: the Axiom of Choice (in our model theory) entails that the full semantics for second-order logic always satisfies the Choice Schema, since one can use a well-order of the underlying domain of the model (or one of its finite products) to obtain the relevant witnesses for the Choice Schema.

Having discussed the axioms, we can finally define some theories:[37]

Definition 1.11: *The axioms of* first-order Zermelo–Fraenkel set theory, ZF, *are Extensionality, Pairing, Union, Power Set, Infinity, Foundation, the Separation Schema, and the Replacement Schema. The theory* ZFC *adds Choice to* ZF.

We can form second-order versions of these theories by replacing the first-order schemas with appropriate second-order sentences. In particular, we replace the

[37] A contemporary reference for ZFC is e.g. the monograph Kunen (1980).

Separation and Replacement *Schemas* with simple *Axioms*, i.e. individual sentences of second-order logic with an additional primitive, $\in$:

Separation. $\forall F \forall x \exists y \forall w \left[w \in y \leftrightarrow \left(w \in x \wedge F(w) \right) \right]$
Replacement. $\forall G \forall w \left[\left(\forall x \in w \right) \exists! y G(x,y) \rightarrow \exists z \left(\forall x \in w \right) \left(\exists y \in z \right) G(x,y) \right]$

We then define:

Definition 1.12: *The theory of* second-order Zermelo–Fraenkel set theory with Choice, ZFC_2, *is formed by taking the axioms of first-order* ZFC, *and replacing the Separation Schema with the Separation Axiom, and the Replacement Schema with the Replacement Axiom, and by adding on the Comprehension Schema.*

As with second-order arithmetic, the Choice Schema is not built into these theories, and should not be confused with the (set-theoretic) Axiom of Choice. The theory ZFC_2 is sometimes also called *Kelly–Morse set theory*.[38] While second-order set theory is less widely used than first-order set theory, it plays an important role in the foundations and philosophy of set theory. We discuss this in Chapters 8 and 11.

Occasionally, but especially from Chapter 7 onwards, we invoke elementary considerations about ordinals and cardinals. As is usual, we reserve $\alpha, \beta, \gamma, \delta$ for ordinals. An *ordinal* is defined to be a transitive set which is well-ordered by membership, where x is transitive iff every member of x is a subset of x. The membership relation on ordinals is usually just written with $<$, and it is provable in very weak fragments of ZFC that $<$ well-orders the ordinals. The successor operation $s(\alpha) = \alpha \cup \{\alpha\} = \alpha + 1$ on ordinals is such that $\alpha < s(\alpha)$ and there is no ordinal β with $\alpha < \beta < s(\alpha)$. We define $0 = \varnothing$, then $1 = s(0), 2 = s(1), 3 = s(2), \ldots$, and $\omega = \{0, 1, 2, 3, \ldots\}$. A limit ordinal is an ordinal β such that $\beta \neq 0$ and $\beta \neq s(\gamma)$ for any ordinal γ; and ω is the least limit ordinal.

A *cardinal* is an ordinal which is not bijective with any smaller ordinal. The finite ordinals $0, 1, 2, \ldots$ and ω are all cardinals. The aleph sequence provides the standard enumeration of infinite cardinals: $\aleph_0 = \omega$; $\aleph_{\alpha+1}$ is the least cardinal greater $\aleph_\alpha$; and when α is a limit ordinal, the cardinal $\aleph_\alpha$ is the least upper bound of $\{\aleph_\beta : \beta < \alpha\}$. Hence $\aleph_\omega$ is the least ordinal which is greater than $\aleph_0, \aleph_1, \aleph_2, \ldots$ and it too is a cardinal. We reserve κ, λ for cardinals, and we use $|X|$ for the *cardinality* of the set X, that is $|X| = \kappa$ iff X is bijective with κ but with no smaller ordinal. We frequently invoke the facts that $|X \times Y| = \max\{|X|, |Y|\}$ when one of $|X|, |Y|$ is infinite, and that the union of $\leq \kappa$-many sets of cardinality $\leq \kappa$ itself has cardinality $\leq \kappa$ when κ is infinite.[39]

[38] See Monk (1969) for an axiomatic development of set theory in this framework.

[39] These elementary facts about cardinality can be found in any set-theory textbook, such as Hrbáček and Jech (1999) or the beginning chapters of Kunen (1980) or Jech (2003).

1.C Deductive systems

In several places in this book, we will need to refer to a deductive system for first-order and second-order logics. Many different but provably equivalent deductive systems are possible, and we could compare and contrast their relative technical and philosophical merits. However, deduction is not the focus of this book, so we will simply set down a system of natural deduction without much comment.[40]

To be clear: we do not expect anyone to be able to learn how to use or manipulate natural deductions just by reading this appendix. Equally, we did not expect that anyone could learn how to do arithmetic or set theory just by reading the previous two appendices. The aim is just to lay down a particular system, so that we can refer back to it later in this book.

First, we lay down rules for the sentential connectives. In the rules ¬E, ∨E, and →I, an assumption is *discharged* at the point when the rule is applied. We mark this using square brackets, and a cross-referencing index, *n*:

$$\frac{\bot}{\varphi}\,\text{Ex} \qquad \frac{\varphi \quad \neg\varphi}{\bot}\,\text{Raa}$$

$$\frac{\begin{array}{c}[\varphi]^n\\ \vdots\\ \bot\end{array}}{\neg\varphi}\,\neg\text{I}, n \qquad \frac{\begin{array}{c}[\neg\varphi]^n\\ \vdots\\ \bot\end{array}}{\varphi}\,\neg\text{E}, n$$

$$\frac{\varphi \quad \psi}{(\varphi \wedge \psi)}\,\wedge\text{I} \qquad \frac{(\varphi \wedge \psi)}{\varphi}\,\wedge\text{E} \qquad \frac{(\varphi \wedge \psi)}{\psi}\,\wedge\text{E}$$

$$\frac{\varphi}{(\varphi \vee \psi)}\,\vee\text{I} \qquad \frac{\psi}{(\varphi \vee \psi)}\,\vee\text{I} \qquad \frac{(\varphi \vee \psi) \quad \begin{array}{c}[\varphi]^n\\ \vdots\\ \chi\end{array} \quad \begin{array}{c}[\psi]^n\\ \vdots\\ \chi\end{array}}{\chi}\,\vee\text{E}, n$$

$$\frac{\begin{array}{c}[\varphi]^n\\ \vdots\\ \psi\end{array}}{(\varphi \to \psi)}\,\to\text{I}, n \qquad \frac{\varphi \quad (\varphi \to \psi)}{\psi}\,\to\text{E}$$

We now consider the rules for first-order quantifiers. These rules are subject to the following restrictions: t can be any term; in ∀I, c must not occur in any undischarged assumption on which $\varphi(c)$ depends; in ∃I one can replace any/all occurrences of t with x, but in ∀I one must replace *all* occurrences of c with x, and in both of these rules x should not already occur in $\varphi(c)$; finally, in implementing ∃E, c must not occur in $\exists x\varphi(x)$, in ψ, or in any undischarged assumption on which ψ depends, except for $\varphi(c)$.

[40] It is essentially based on Prawitz (1965).

$$\frac{\varphi(c)}{\forall x \varphi(x)}\,\forall\mathrm{I} \qquad \frac{\forall x \varphi(x)}{\varphi(t)}\,\forall\mathrm{E}$$

$$\frac{\varphi(t)}{\exists x \varphi(x)}\,\exists\mathrm{I} \qquad \frac{\exists x \varphi(x) \qquad \begin{array}{c}[\varphi(c)]^n \\ \vdots \\ \psi\end{array}}{\psi}\,\exists\mathrm{E},\, n$$

To complete the rules for first-order logic, we have the rules for identity. Note that adopting the rule =I is equivalent to treating every instance of $t = t$ as an *axiom*, since it is licensed on any (including no) assumptions:

$$\frac{}{t = t}\,=\mathrm{I} \qquad \frac{t_1 = t_2 \quad \varphi(t_1)}{\varphi(t_2)}\,=\mathrm{E} \qquad \frac{t_2 = t_1 \quad \varphi(t_1)}{\varphi(t_2)}\,=\mathrm{E}$$

To move to a deduction system for second-order logic, we simply add rules for the quantifiers, exactly analogous to the first-order case. So, for relation-variables we have (with similar restrictions as before):

$$\frac{\varphi(R^m)}{\forall X^m \varphi(X^m)}\,\forall_2\mathrm{I} \qquad \frac{\forall X^m \varphi(X^m)}{\varphi(R^m)}\,\forall_2\mathrm{E}$$

$$\frac{\varphi(R^m)}{\exists X^m \varphi(X^m)}\,\exists_2\mathrm{I} \qquad \frac{\exists X^m \varphi(X^m) \qquad \begin{array}{c}[\varphi(R^m)]^n \\ \vdots \\ \psi\end{array}}{\psi}\,\exists_2\mathrm{E},\, n$$

The case of function symbols is exactly similar. Finally, to ensure that our deduction system aligns with *faithful* Henkin models, we also allow as axioms any instance of the Comprehension or Choice schemas, i.e. we add these rules:

$$\frac{}{\exists X^n \forall \bar{v}(\varphi(\bar{v}) \leftrightarrow X^n(\bar{v}))}\,\mathrm{Comp}$$

$$\frac{}{\forall X^{n+1}\left(\forall \bar{v} \exists y\, X^{n+1}(\bar{v}, y) \rightarrow \exists p^n \forall \bar{v}\, X^{n+1}(\bar{v}, p^n(\bar{v}))\right)}\,\mathrm{Choice}$$

These are all the rules for our deduction systems for sentential, first-order and second-order logic. When we have a deduction whose only undischarged assumptions are members of T and which ends with the line φ, we write $T \vdash \varphi$.

2

Permutations and referential indeterminacy

In Chapter 1, we introduced some of the most basic ideas in model theory: structures, signatures, and satisfaction. In this chapter, we introduce the fundamental notion of an *isomorphism*. This provides the technical basis for several philosophical issues made famous by Benacerraf and Putnam, concerning both the 'intuitive' idea of a mathematical structure and referential indeterminacy.

2.1 Isomorphism and the Push-Through Construction

One of the most fundamental ideas in model theory is *isomorphism*. We come to the idea of isomorphism via the notion of a map which 'preserves structure'.

Definition 2.1: *Let $\mathcal{M}$ and $\mathcal{N}$ be $\mathscr{L}$-structures. A bijection $h : M \longrightarrow N$ is an* isomorphism *from $\mathcal{M}$ to $\mathcal{N}$ iff: for any $\mathscr{L}$-constant symbol c, any n-place $\mathscr{L}$-relation symbol R, any n-place $\mathscr{L}$-function symbol f, and all $a_1, \ldots, a_n$ from M:*

$$h(c^{\mathcal{M}}) = c^{\mathcal{N}}$$
$$(a_1, \ldots, a_n) \in R^{\mathcal{M}} \text{ iff } (h(a_1), \ldots, h(a_n)) \in R^{\mathcal{N}}$$
$$h(f^{\mathcal{M}}(a_1, \ldots, a_n)) = f^{\mathcal{N}}(h(a_1), \ldots, h(a_n))$$

When there is an isomorphism from $\mathcal{M}$ to $\mathcal{N}$, we say that $\mathcal{M}$ and $\mathcal{N}$ are isomorphic, *and write $\mathcal{M} \cong \mathcal{N}$.*

We continue to use overlining to discuss tuples, as introduced §1.2. So $\bar{a}$ will be some sequence of elements $(a_1, \ldots, a_n)$. We also introduce some notation to allow tuples to interact easily with functions. The idea is to 'push h through' sets of elements of M and N, through sets of sets of elements of M and N, and so on:

Definition 2.2: *Let $h : M \longrightarrow N$ be any function and let $\bar{a}$ be from M. We define $\hbar(\bar{a}) = (h(a_1), \ldots, h(a_n))$. For each $X \subseteq M^n$, we define $\hbar(X) = \{\hbar(\bar{a}) : \bar{a} \in X\}$. Likewise, for each $Y \subseteq \wp(M^n)$ we define $\hbar(Y) = \{\hbar(X) : X \in Y\}$.*

In these terms, we can rewrite the last part of the definition of an isomorphism as $h(f^{\mathcal{M}}(\bar{a})) = f^{\mathcal{N}}(\hbar(\bar{a}))$. Where no ambiguity can arise, we sometimes simply write $h(\bar{a})$ rather than $\hbar(\bar{a})$. Now, we have explicitly written out the definition of $\hbar$ on the first couple of levels of the set-theoretic hierarchy above M, but the definition generalises naturally to higher levels. In each case, we simply define the action of $\hbar$ on a higher-level object X as the set which collects together the action of $\hbar$ on all of X's members.

There are many equivalent ways to define the notion of an isomorphism:

Theorem 2.3: *For any $\mathscr{L}$-structures $\mathcal{M}$ and $\mathcal{N}$ and any bijection $h : M \longrightarrow N$, the following are equivalent:*

(1) *h is an isomorphism from $\mathcal{M}$ to $\mathcal{N}$*
(2) *$\mathcal{M} \vDash \varphi(\bar{a})$ iff $\mathcal{N} \vDash \varphi(\hbar(\bar{a}))$, for all $\bar{a}$ from M^n and all atomic $\mathscr{L}$-formulas $\varphi(\bar{v})$ with free variables displayed*
(3) *$\mathcal{M} \vDash \varphi(\bar{a})$ iff $\mathcal{N} \vDash \varphi(\hbar(\bar{a}))$, for all $\bar{a}$ from M^n and all first-order $\mathscr{L}$-formulas $\varphi(\bar{v})$ with free variables displayed*
(4) *$\mathcal{M} \vDash \varphi(\bar{a})$ iff $\mathcal{N} \vDash \varphi(\hbar(\bar{a}))$, for all $\bar{a}$ from M^n and all second-order $\mathscr{L}$-formulas $\varphi(\bar{v})$ with free variables displayed, with consequence read either via the full or the Henkin semantics for second-order logic (see §§1.10–1.11)*

The proof of this result involves a lengthy induction on complexity of formulas, which we relegate to §2.B. But Theorem 2.3 has an immediate, important corollary. Since sentences are just formulas with no free variables, the entailment (1) ⇒ (3) shows that isomorphic structures make exactly the same first-order sentences true. This idea is significant enough to merit some new terminology.

Definition 2.4: *Let $\mathcal{M}$ and $\mathcal{N}$ be $\mathscr{L}$-structures. We say that $\mathcal{M}$ and $\mathcal{N}$ are* elementarily equivalent, *written $\mathcal{M} \equiv \mathcal{N}$, iff they satisfy exactly the same $\mathscr{L}$-sentences, i.e. $\mathcal{M} \vDash \varphi$ iff $\mathcal{N} \vDash \varphi$, for all $\mathscr{L}$-sentences φ.*

With this notation, the corollary of Theorem 2.3 which we just observed becomes:

Corollary 2.5: *If $\mathcal{M} \cong \mathcal{N}$, then $\mathcal{M} \equiv \mathcal{N}$.*

The *converse* to Corollary 2.5 is false. However, showing this requires a slightly different set of tools, and so we defer discussion of this point until Chapter 4.

Isomorphic—and so elementarily equivalent—structures are very easy to construct. Indeed, given any structure and any bijection whose domain is the structure's underlying domain, we can treat that bijection as an isomorphism. This is, in fact, one of the most basic constructions in model theory.

The Push-Through Construction. *Let $\mathscr{L}$ be any signature, let $\mathcal{M}$ be any $\mathscr{L}$-structure with domain M, and let $h : M \longrightarrow N$ be any bijection. We use h to define an $\mathscr{L}$-structure, $\mathcal{N}$, with domain N, by defining $s^{\mathcal{N}} = h(s^{\mathcal{M}})$ for each $\mathscr{L}$-symbol s.*[1] *So, for any $\mathscr{L}$-constant symbol c, any n-place $\mathscr{L}$-relation symbol R, any n-place $\mathscr{L}$-function symbol f, and all $\bar{a}$ from M^n:*

$$c^{\mathcal{N}} = h(c^{\mathcal{M}})$$
$$R^{\mathcal{N}} = h(R^{\mathcal{M}}) = \{h(\bar{a}) : \bar{a} \in R^{\mathcal{M}}\}$$
$$f^{\mathcal{N}} = h \circ f^{\mathcal{M}} \circ h^{-1}, \text{ so that } f^{\mathcal{N}}(h(\bar{a})) = h(f^{\mathcal{M}}(\bar{a}))$$

In this, $h^{-1}(\bar{b}) = \bar{a}$ iff $h(\bar{a}) = \bar{b}$. We may write $h : \mathcal{M} \longrightarrow \mathcal{N}$ to indicate that we are considering the function built from h which induces $\mathscr{L}$-structure.

As mentioned, this Construction is an extremely simple way to generate new structures from old ones. But it has a surprising number of rich philosophical consequences, which we explore in this chapter and the next.

2.2 Benacerraf's use of Push-Through

Since the Push-Through Construction makes isomorphic copies of structures *so* easy to come by, it is no surprise that, for many mathematical purposes, it seems not to matter which of two isomorphic models one works with.

The most famous philosophical statement of this point is due to Benacerraf. He focussed specifically on the fact that, when doing arithmetic, it makes no difference whether we think of the natural numbers as Zermelo's finite ordinals:

$$\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \{\{\{\varnothing\}\}\}, \ldots$$

or as von Neumann's finite ordinals:[2]

$$\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}, \ldots$$

Benacerraf maintained:

For arithmetical purposes, the properties of numbers which do not stem from the relations they bear to one another in virtue of being arranged in a progression are of no consequence whatsoever.[3]

[1] We can also define the construction when $\mathcal{M}$ is a Henkin $\mathscr{L}$-structure, as in Definition 1.6. The idea is to keep pushing h through the range of $\mathcal{N}$'s second-order variables. So we set $N_n^{\text{rel}} = h(M_n^{\text{rel}})$ and $N_n^{\text{fun}} = h(M_n^{\text{rel}})$. Note that there is no guarantee that $R^{\mathcal{N}} = h(R^{\mathcal{M}})$ should be a member of M_n^{rel}, even when $M = N$; the existence of $R^{\mathcal{N}}$ is guaranteed by the *ambient* set-theoretic framework within which we are working. Similarly, we should not expect that $N_n^{\text{rel}} = M_n^{\text{rel}}$ or that $N_n^{\text{fun}} = M_n^{\text{fun}}$.

[2] By our definition of $s(x)$ in §1.B, the finite von Neumann ordinals are the sets $s^n(\varnothing)$ for each n.

[3] Benacerraf (1965: 69-70).