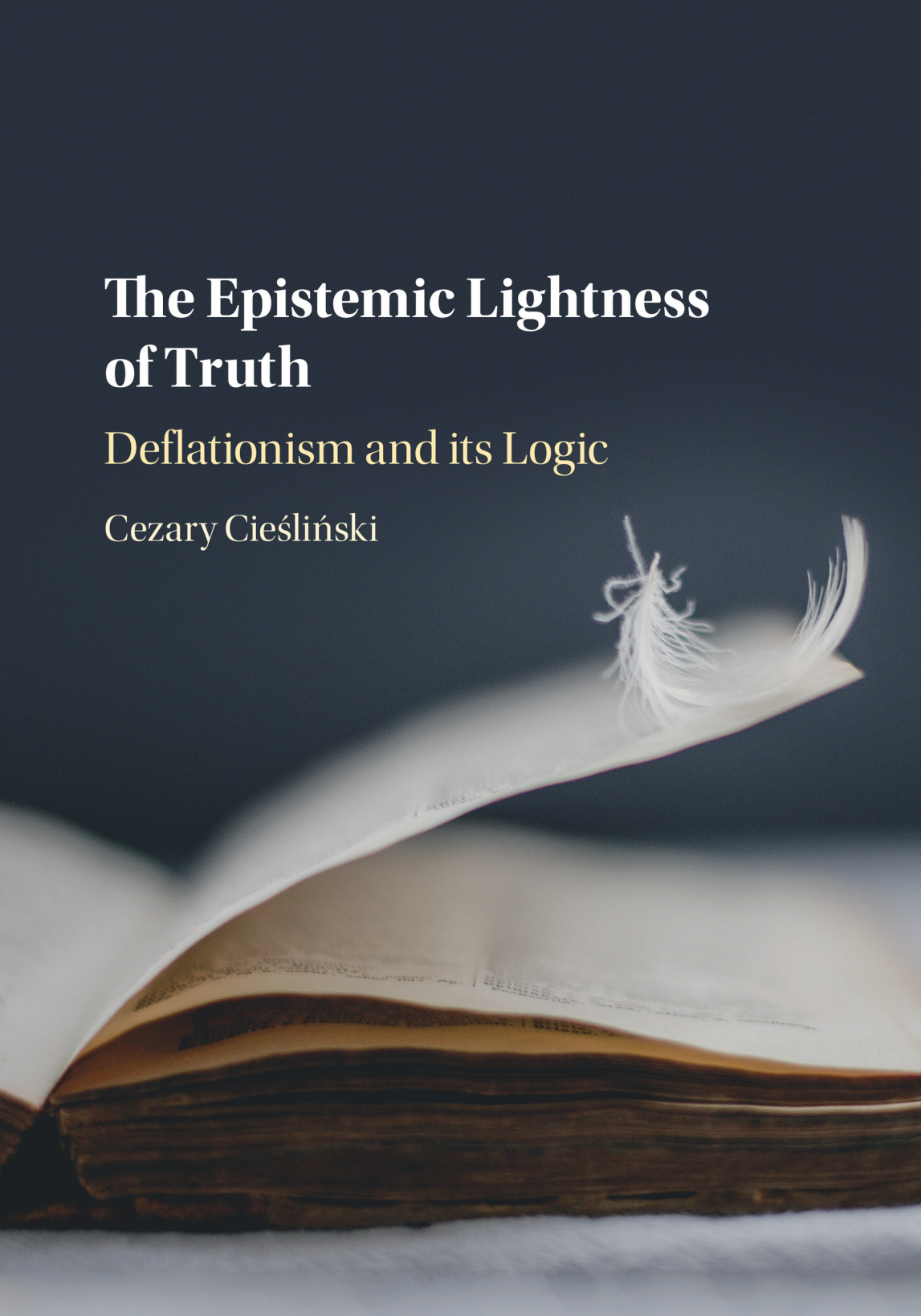
The Epistemic Lightness of Truth

Deflationism and its Logic

Cezary Cieśliński



THE EPISTEMIC LIGHTNESS OF TRUTH

This book analyses and defends the deflationist claim that there is nothing deep about our notion of truth. According to this view, truth is a 'light' and innocent concept, devoid of any essence that could be revealed by scientific inquiry. Cezary Cieśliński considers this claim in light of recent formal results on axiomatic truth theories, which are crucial for understanding and evaluating the philosophical thesis of the innocence of truth. Providing up-to-date discussion and original perspectives on this central and controversial issue, his book will be important for those with a background in logic who are interested in formal truth theories and in current philosophical debates about the deflationary conception of truth.

CEZARY CIEŚLIŃSKI is a member of the Institute of Philosophy at the University of Warsaw. His research, which focuses on truth theories, logic, and philosophy of language, has been published in journals including *Mind* and *Journal of Philosophical Logic*.

THE EPISTEMIC LIGHTNESS OF TRUTH

Deflationism and Its Logic

CEZARY CIEŚLIŃSKI
University of Warsaw

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

4843/24, 2nd Floor, Ansari Road, Daryaganj, Delhi – 110002, India

79 Anson Road, #06-04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107197657
DOI: 10.1017/9781108178600

First published 2017

Printed in the United Kingdom by Clays, St Ives plc

A catalogue record for this publication is available from the British Library.

ISBN 978-1-107-19765-7 Hardback

Contents

Acknowledgements

This book is the result of the project "How innocent is the concept of truth? Philosophical and logical analysis of deflationism" financed by the National Science Centre, Poland (NCN) based on the decision number DEC-2011/01/B/HS1/03910.

The book has emerged from the author having spent many years teaching, writing and thinking about the topic of formal theories of truth. My sincere gratitude goes to all the people who influenced me over these years of academic work. In particular, I am indebted to Ali Enayat, Martin Fischer, Volker Halbach, Leon Horsten, Jeffrey Ketland, Henryk Kotlarski, Marcin Mostowski, Rafał Urbaniak, Albert Visser and Konrad Zdanowski for lots of stimulating discussions and exchanges on the topic of philosophical and formal theories of truth.

I am much indebted to Rafał Urbaniak, who carefully read through all the versions of the manuscript and provided several valuable remarks and suggestions. I would also like to express particular appreciation and thanks to Volker Halbach, as this book owes much both to his influence and to his support.

I am grateful to the anonymous reviewers, whose suggestions and criticisms did a lot to improve the final version of the manuscript. I would also like to thank Hilary Gaskin, Daniel Brown, Sophie Taylor and the whole team from Cambridge University Press for their efficient editorial work and excellent guidance.

Last but not least, many thanks to my PhD students, in particular to Mateusz Łełyk, Bartosz Wcisło, Michał Tomasz Godziszewski and Wojciech Rostworowski. Not only have they often been the first audience permitting me to test various ideas espoused in this book, but they also actively participated in the research on formal theories of truth, enriching it with interesting new insights and theorems (indeed, some of their original results will be presented here). It has been both a pleasure and a privilege to work with such students.

Above all, my heartfelt gratitude goes to my family. I would like to thank my wife Agnieszka for her continuous support, for her energy and optimism,

for providing motivation and even for encouraging me to work in moments of doubt. Warm thanks go also to my daughter Justyna for showing a lot of patience and understanding. Without both of you this book would never have been written.

Introduction

Is there anything more familiar and obvious than the opposition of truth and falsity? It is true that the earth is round. It is false that dragons eat virgins. (As everyone knows, dragons eat only pistachio marzipan with vanilla truffle.) Elementary, is it not? However, if this is so familiar, what then is truth? When confronted with such a direct question, many of us are tempted to repeat the famous words of Saint Augustine: "If no one asks me, I know what it is. If I wish to explain it to him who asks, I do not know".

Being that no decent philosopher can rest satisfied with *ignorabimus*, some answers have naturally been proposed. Indeed, answers have proliferated, with various philosophical schools promoting their own worldviews and agendas. Unfortunately, no lasting consensus has emerged, with the only exception perhaps being the following. Philosophers seemed to agree that the task of explaining the nature of truth is a daunting one; it is hard, complicated, deep and far-reaching. However, in recent times, serious doubts have emerged even here. Some modern philosophers have reacted to the ancient puzzles with a bold claim; they have said that, in fact, truth has no nature, and the very concept of truth is, in some sense, innocent or trivial. This book is devoted to the analysis and assessment of this claim.

So, what is truth? Here is a selection of quotes giving answers to this question.

- 'To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true.' (Aristotle, *Metaphysics*, IV 7, 1011b27)
- '*Veritas est adaequatio intellectus et rei*.' ('Truth is the conformity of the intellect to the things.' Thomas Aquinas, *Summa Theologica* I, Q 16)
- 'The nominal definition of truth, namely that it is the agreement of cognition with its object, is here granted and presupposed.' (I. Kant, *Critique of Pure Reason*, A 57-8/B 82)

In one crucial respect, the first of these classical formulations is rather different from the other two. When defining truth, both Aquinas and Kant mention a special relation which is supposed to hold between the intellect (or

cognition) and its object; namely, the relation of 'conformity' or 'agreement'. In the literature, it is also customary to use the term 'correspondence' in this context. In short, the classical definition of truth consists in defining truth as the correspondence of thought (cognition) with reality.

However, once we start playing with the idea of a correspondence relation, difficult philosophical questions arise. What is the nature of this special relation between thought (or language) and reality? Does a given sentence (proposition) correspond to reality taken as a whole or to only a fragment of it? If it is the latter, then which fragment is it? Can we claim, for example, that it is the objective facts that make our sentences (propositions, thoughts) true? Here is another question concerning correspondence: in virtue of *what exactly* does this relation hold? For example, is the requirement that a truth bearer (sentence, proposition) has a similar structure to the corresponding fragment of reality (fact, state of affairs)? These are indeed troublesome questions, and many philosophers have been deeply dissatisfied with the traditional answers given to them.

On the other hand, unlike in the case of Aquinas and Kant, when reading Aristotle's explanation, it is hard to deny the impression that the notion of truth is (in some sense to be specified) simple, innocent and trivial. Aristotle's formulation is much more austere and cautious than those of the other authors quoted here. Indeed, it is worth emphasising that here Aristotle does not appeal at all to correspondence. To say 'there are dragons' is false because there are no dragons; to say 'there are horses' is true, since there are horses; and to say 'there are no electrons' is false because electrons exist – that is the underlying idea. In contrast to Aquinas and Kant, no special relation between thought (or language) and reality has been invoked.

This Aristotelian motive came to the foreground in some recent works on truth, notably by philosophers representing the popular current called 'deflationism about truth'. It is indeed the deflationary intuition that truth is in some sense insubstantial, light or metaphysically thin.[1] The

1 This is not to say that Aristotle himself should be classified as a deflationist. On the one hand, as noted by Crivelli (2004, p. 30-31), relational properties were not considered 'real' or 'genuine' by Aristotle, and since he considered truth to be a relational property, he was 'committed to the view that truth is not a genuine property. In this respect Aristotle's position is close to modern 'minimalist' theories of truth, which also claim that truth is not a genuine property'. On the other hand, a careful reconstruction of Aristotle's views leads Crivelli to the conclusion that Aristotle was, after all, an adherent of a correspondence theory of truth. For details, the reader is referred to Chapter 4 of (Crivelli 2004).

deflationists frequently repeat that when we attribute truth to a sentence (or a proposition), we might just as well assert this very sentence (or this proposition). They also say that truth has no 'essence' which could be revealed by deep scientific research. As an example, consider the following (typical) quote from Horwich:

> […] the traditional attempt to discern the essence of truth – to analyse that special quality which all truths supposedly have in common – is just a pseudo-problem based on syntactic overgeneralization. Unlike most other properties, being true is unsusceptible to conceptual or scientific analysis. No wonder that its 'underlying nature' has so stubbornly resisted philosophical elaboration; for there is simply no such thing. (Horwich 1999, p. 5)

What does it mean to claim that truth has no 'underlying nature'; that it is insubstantial, light or metaphysically thin? Truth may be a simple notion (as the deflationist wants it to be) but – as it turns out – answering the last question is still quite a demanding task. The exploration of this topic is a central theme of this book.

Here I am going to defend a certain strong version of the lightness thesis. The outline is as follows. Two explications of the lightness claim have been prominent in the literature. One of them is that truth is a disquotational notion and can be fully characterised by the so-called T-sentences or 'Tarski biconditionals'; that is, by the equivalences falling under the schema 'the sentence (or the proposition) φ is true if and only if φ'. In this view, it is the simplicity and triviality of the T-schema that gives meaning and justification to the lightness thesis. The second explication is the conservativity proposal; roughly, truth is innocent because adequate theories of truth do not establish any new non-semantic facts. A detailed discussion of these explications will be presented in Part II and Part III of this book.

Both proposals have evoked harsh criticism. In both cases, the main thrust was directed against the truth-theoretic weakness of the envisaged disquotational (or conservative) theories of truth. The critics have claimed that such theories cannot provide an adequate characterisation of truth for a very simple reason: *in fact* our knowledge about truth goes beyond such theories; in other words, facts about truth are known to us which cannot be deduced from disquotational/conservative theories of truth. In effect, the adherents to these truth theories cannot account for this additional knowledge. This is the objection.

Let me emphasise that the problem of the truth-theoretic weakness is very real. It does not rest on any misunderstanding or a flaw in the critics' reasoning. On the contrary, critics have quite correctly identified the aforementioned traits of disquotational and conservative theories of truth. Nevertheless, the main philosophical claim of this book is that an adequate theory of truth can be *both* disquotational *and* conservative. In the final chapter a solution to the problem of truth-theoretic weakness will be proposed. Namely, it will be argued that the deflationist who accepts a given disquotational and conservative theory of truth has at his disposal sufficient means to account for any additional knowledge about truth that we may possess, including facts about truth which are not provable in his initial theory. In this way, the deflationary standpoint will be vindicated.

In the discussion of innocence claims, this book will often employ formal tools of modern logic. More specifically, the claims in question will be analysed mainly within the arithmetical framework. The case of arithmetic will be treated here as a model example against which the deflationary tenets can be evaluated and tested. The assumption is that if innocence claims do not pass such a preliminary arithmetical test, then they are to be disqualified almost from the start without the need to take into consideration additional semantic phenomena. The general motivation might be global, but testing is best done on a local level; that is at least the idea. Accordingly, the book does not provide any analysis of the use of truth in science in general, nor do I purport to analyse any particular troublesome traits of natural languages, such as ambiguity, vagueness or indexicality. Instead of taking a broad-brush approach, I want to offer to the reader a detailed analysis of some quite specific issues arising in arithmetical contexts on the borderline between philosophy and formal logic.

Typically, the discussion will proceed in accordance with the following schema. Starting with some basic, philosophical idea ('truth is nothing more than disquotation' can serve as an example), I present the intuitions guiding the proponents of a given philosophical standpoint. In the next stage, formal theories are introduced, treated as attempts at a precise characterisation of the idea in question. The third stage presents the analysis of logical properties of these formal theories – it is here where formal methods will be most extensively used. Finally, the discussion returns to philosophical issues, which are analysed again in the light of mathematical results.

The plan of the book is as follows.

Chapter 1 ('Preliminaries') fixes the basic notation and terminology; I also state (without proofs) some classical formal results, which will be useful later in the book. The reader might wish to start by checking the terminology and then to use Chapter 1 as reference material, to be consulted whenever the need arises.

In Chapter 2 two general methods of characterising the notion of truth are laid out: axiomatic and model-theoretic. Being that the axiomatic method will be deemed the more suitable of the two for the purpose of defending the innocence claims, this book will focus on the axiomatic approach. It will hence deal with attempts to characterise the notion of truth *simpliciter* (the truth of sentences as we understand them in contrast to 'truth under an interpretation' or 'truth in a model') by means of simple and basic principles, with the truth predicate functioning as a primitive, undefined symbol.

Special attention will be given to disquotational and conservative truth theories; they will be discussed in Parts I and II of this book. In each of these cases I start by presenting philosophical intuitions behind both types of truth theories; the discussion will then proceed to an analysis of their formal properties. The last chapters of both Part I and Part II are devoted to the presentation of the main objections against (respectively) disquotational and conservative theories of truth. These objections are known in the literature as 'the generalisation problem' and 'the conservativeness argument'.

In the final Part III I present my uniform response to both the generalisation problem and the conservativeness argument, defending disquotational and conservative truth theories against the charge of truth-theoretic weakness. The claim will be, in effect, that such theories stay with us as formalisations of a natural and fundamentally correct approach to truth.

All of Parts I through III begin with introductory sections, which not only sketch the basic intuitions but contain also a more detailed plan of the subsequent chapters, providing the reader with a map of what is to follow. In addition, each chapter following the 'Preliminaries' ends with a summary, where the main claims are briefly listed.

I will generally avoid describing non-trivial mathematical proofs and techniques whose presentations can be found elsewhere in book format. Normally in such cases the most important theorems will be merely stated with a reference given. Nevertheless, various theorems (particularly new results, including those due to the author or his students) will be introduced with full proofs. Open mathematical problems, arising from the logical

and philosophical analysis of deflationary ideas about truth, will also be presented. It should be emphasised here that these formal parts do not just serve philosophical purposes. The additional aim is to bring the reader up to date with some of the most recent developments in formal work on truth theories and, ultimately, to convey the impression of the field as a fascinating and vibrant one worthy of further investigation. Nevertheless, for the reader's convenience, in the summaries of the formal chapters I will clearly indicate which of the theorems are of particular importance for the main philosophical theme of the book.

Let me finish by saying that the idea of translating philosophical intuitions into precise, formal claims and hypotheses is one that I find immensely appealing. This is not meant to minimise the role of intuitions, which remain absolutely crucial for our research in all of its stages. Nonetheless, it is only the precise formulations, with all the care given to the details, which permit us to test the validity of our intuitions. Certainly, there are risks, but I consider them worth taking. From my point of view, much of the value of deflationism considered as a philosophical standpoint derives from the fact that, to a substantial degree, it is susceptible to such a procedure.

1 Preliminaries

This preliminary chapter introduces the notation and basic terminology. Apart from this, I will also formulate (usually without proofs) some classical results, which will be referred to in this book. It should be emphasised at the start that this is not intended to be a comprehensive overview of any particular discipline or area of research. As a matter of fact, the main and often the only criterion motivating the choice of the material is whether a given concept (or a lemma, or a theorem) will be useful in the chapters to follow.

1.1 Peano Arithmetic

The first definition describes the language of first-order arithmetic; in the next move, a concrete arithmetical theory will be characterised: Peano arithmetic.

DEFINITION 1.1.1. The language of first-order arithmetic, denoted here as L_{PA}, contains the usual logical vocabulary (quantifiers, connectives, brackets, and variables $v_0, v_1 \ldots$). The set of primitive extralogical symbols of L_{PA} is defined as $\{\text{'}+\text{'}, \text{'}\times\text{'}, \text{'}0\text{'}, \text{'}S\text{'}\}$; in effect, it contains symbols for addition, multiplication, zero, and the successor function, respectively.

Terms, formulas and sentences of L_{PA} are defined in the usual style (in particular, sentences of L_{PA} are defined as formulas without any free variables). The expressions Var, Tm, Tm^c, and $Sent_{L_{PA}}$ will be used as referring (respectively) to the sets of variables, terms, constant terms, and sentences of L_{PA}. In general, for a theory Th, the expressions L_{Th} and $Sent_{L_{Th}}$ will refer to the language of Th and to the set of sentences of the language of Th.

The next definition introduces Peano arithmetic.

DEFINITION 1.1.2. Peano arithmetic (PA) is defined as the theory with the following arithmetical axioms:[1]

1 Apart from that, the set of axioms of PA will contain the axioms of first-order logic.

1. $\forall x\ S(x) \neq 0$
2. $\forall x,y\ [S(x) = S(y) \rightarrow x = y]$
3. $\forall x\ x + 0 = x$
4. $\forall x,y\ x + S(y) = S(x + y)$
5. $\forall x\ x \times 0 = 0$
6. $\forall x,y\ x \times S(y) = (x \times y) + x$
7. $\{[\varphi(0) \wedge \forall x\ (\varphi(x) \rightarrow \varphi(S(x)))] \rightarrow \forall x\ \varphi(x) : \varphi(x) \in L_{PA}\}$

The last item is the set of arithmetical sentences falling under the schema of mathematical induction. Since there are infinitely many such sentences, the axiomatisation given here is patently not finite.[2]

The language of first-order arithmetic, as characterised in Definition 1.1.1, does not contain any numerals except for the symbol '0' (that is, it does not contain terms '1', '2' etc.). However, the notion of a numeral – a canonical term denoting a number – can be defined in the following way:

DEFINITION 1.1.3. A numeral is an arbitrary term of L_{PA} of the form '$S \ldots S(0)$', i.e. a term obtained by preceding a symbol 'o' with (arbitrarily many) successor symbols. If the number of successor symbols in a numeral equals n, the numeral will be abbreviated as $\overline{n}$.

Some schema of coding (or Gödel numbering) will be tacitly assumed throughout the book. It is possible to define a procedure, which starts with assigning numbers to primitive expressions of L_{PA} and then extending the assignment to cover more complex syntactical objects. Eventually unique natural numbers become assigned to terms, formulas, and sequences of formulas (including proofs).[3] In effect it becomes possible to view some statements of first-order arithmetic as assertions about syntax.[4]

Truth predicate will be understood in this book as applying to syntactic objects, namely, to sentences.[5] Accordingly, a theory of syntax forms a

2 Moreover, in this respect the axiomatisation cannot be improved: it is known that Peano arithmetic is not finitely axiomatisable. See (Hájek and Pudlák 1993, p. 164), Corollary 2.24.

3 The classical method employs prime factorisation: a finite sequence of numbers $(n_1 \ldots n_k)$ will be coded by the number $2^{(n_1+1)} \times 3^{(n_2+1)} \times \ldots \times p_k^{(n_k+1)}$, with p_k being the k-th prime.

4 I will not describe the details of coding here; they can be found, e.g. in (Kaye 1991).

5 Choosing sentences instead of propositions brings simplicity, although it should be admitted that this is not a philosophically innocent decision. In particular, Halbach (2011, p. 12) observes that the modal status of disquotation sentences (like "'Snow is white' is true if and only if snow is white") depends on whether truth is ascribed to a proposition or to a sentence, with some philosophers arguing that only with the first option the disquotation sentences become *necessary*.

necessary base for the theory of truth. Peano arithmetic is one of the theories suitable for this role, with the reason being that basic syntactic properties and relations are recursive, and Peano arithmetic is strong enough to represent them. The exact definition of the notion of a recursive set will not be given here; let me emphasise only that, in intuitive terms, a set is recursive if there is an algorithm which decides, for an arbitrary number n, whether or not n belongs to this set. In what follows, I will describe only the important notion of representability together with its basic properties, treating the concept of a recursive set as given.

DEFINITION 1.1.4. A set of natural numbers Z is representable in an arithmetical theory Th iff there is a formula $\varphi(x)$ of the language of Th, with one free variable, such that for every natural number n:

1. if $n \in Z$, then $Th \vdash \varphi(\overline{n})$,
2. if $n \notin Z$, then $Th \vdash \neg\varphi(\overline{n})$.

With these conditions satisfied, we say also that $\varphi(x)$ represents Z in Th.

Before formulating the representability theorem, let me introduce the familiar arithmetical hierarchy.

DEFINITION 1.1.5 (Arithmetical hierarchy).

- A bounded quantifier is a quantifier of the form '$Qx < y$', for $Q \in \{\forall, \exists\}$.
- A formula φ belongs to the class Δ_0 iff all the quantifiers in φ are bounded. (We stipulate also that, by definition, $\Delta_0 = \Sigma_0 = \Pi_0$.)
- A formula φ belongs to the class Σ_{n+1} iff for some $\psi \in \Pi_n$ and for some sequence of variables a, φ has a form '$\exists a \psi$'.
- A formula φ belongs to the class Π_{n+1} iff for some $\psi \in \Sigma_n$ and for some sequence of variables a, φ has a form '$\forall a \psi$'.

Σ_n and Π_n classes were characterised here as containing only formulas of a rather special syntactic type. Observe in particular that Definition 1.1.5 does not introduce any closure of these classes under provable equivalence, and for this reason Σ_n and Π_n classes do not exhaust the set of all formulas (clearly there exist formulas whose syntactic form is altogether different, for example '$\exists x\, x = x \wedge \exists x\, x = x$' is neither Σ_n nor Π_n). Nevertheless, it is possible to show that every formula is provably (in PA) equivalent to some Σ_n (or Π_n) formula.

The following theorem is crucial for appreciating Peano arithmetic's role as a theory of syntax.

THEOREM 1.1.6 (Representability of recursive sets). For every recursive set X of natural numbers, there is a Σ_1 formula representing X in PA.[6]

Since a lot of basic syntactic properties are recursive, this gives us the means to build a theory of syntax inside *PA*. In particular, the following properties and relations are recursive:

- x is a negation of y,
- x is a conjunction of y and z,
- x is a variable, x is a term, x is a formula,
- x is a numeral denoting a number y,
- x is the result of substituting a term t for a variable v in a formula z.

Accordingly, Theorem 1.1.6 guarantees the existence of arithmetical formulas representing these syntactical properties and relations (they will be denoted, respectively, as $x = neg(y)$, $x = Conj(y,z)$, $Var(x)$, $Tm(x)$, $Fm(x)$, $x = name(y)$, and $x = sub(z,v,t)$). The road is open to building a theory of syntax inside *PA*.

The following application of the representability theorem will be of particular importance.

DEFINITION 1.1.7. Given a fixed recursive set $Ax(Th)$ axiomatising a theory Th, '$Prov_{Th}(x,y)$' is a formula of the language of *PA* which represents in *PA* the recursive relation 'd is a proof of φ from $Ax(Th)$'. Given a formula '$Prov_{Th}(x,y)$', '$Pr_{Th}(y)$' is defined as the formula '$\exists x Prov_{Th}(x,y)$'.[7]

It should be stressed that by this definition, '$Prov_{Th}(x,y)$' is just *any* formula representing the relation of being a proof. For a given axiomatisation of Th, there will be many such formulas, sometimes with importantly different properties. The same concerns the provability formulas '$Pr_{Th}(y)$' – it is often important to keep in mind that it is not a uniquely determined single expression of L_{PA}.

In what follows I am not going to distinguish between formulas and their Gödel numbers (for all practical aims, I will just assume that formulas *are* Gödel numbers). Sometimes in this book square corners will be used for

6 For the proof, see (Kaye 1991, pp. 36–37).

7 Strictly speaking, for two different axiomatisations $Ax_1(Th)$ and $Ax_2(Th)$ of one and the same theory Th we would need two different formulas '$Prov_{Ax_1(Th)}(x,y)$' and '$Prov_{Ax_2(Th)}(x,y)$', representing the relations of being a proof from the respective sets of axioms. I skip here this complication, noting only that the notation '$Prov_{Th}(x,y)$' presupposes a concrete, fixed axiomatisation of Th.

numerals denoting syntactic objects. Thus, if φ is a formula, the notation $\ulcorner\varphi\urcorner$ is reserved for a numeral denoting φ. In addition, Feferman's dot notation will be occasionally employed. Thus, let $\varphi(x)$ and $\psi(x)$ be formulas. The expression:

$$\varphi(\ulcorner\psi(\dot{x})\urcorner)$$

will be treated as an abbreviation of

$$\exists y,z[y = name(x) \wedge z = sub(\ulcorner\psi(x)\urcorner, x, y) \wedge \varphi(z)].$$[8]

In some contexts, what is needed is not an arbitrary provability formula (build over an arbitrary proof predicate), but a predicate with some special properties. In such cases this will be stipulated explicitly. Some important constraints are listed in the next definition.

DEFINITION 1.1.8 (Derivability conditions). Given an axiomatisable theory Th (in the language L_{Th}) extending PA, the following three statements will be called 'derivability conditions' for the predicate '$Pr_{Th}(x)$':

(D_1) For every $\psi \in L_{Th}$, if $Th \vdash \psi$, then $PA \vdash Pr_{Th}(\ulcorner\psi\urcorner)$,
(D_2) $\forall\psi,\varphi \in L_{Th}\ PA \vdash (Pr_{Th}(\ulcorner\varphi \to \psi\urcorner) \wedge Pr_{Th}(\ulcorner\varphi\urcorner)) \to Pr_{Th}(\ulcorner\psi\urcorner)$,
(D_3) $\forall\varphi \in L_{Th}\ PA \vdash Pr_{Th}(\ulcorner\varphi\urcorner) \to Pr_{Th}(\ulcorner Pr_{Th}(\varphi)\urcorner)$.

Any provability predicate $Pr_{Th}(x)$ satisfying all three derivability conditions will be called 'standard'.

It is possible to show that the 'natural' provability predicate, defined in PA in a way which closely mimics the usual, external definition of provability, is standard.[9]

8 Informally, this could be expressed as 'φ is true about the (Gödel number of the) result of substituting a numeral denoting x for a free variable in ψ'. Observe that, in effect, the expression '$\varphi(\ulcorner\psi(\dot{x})\urcorner)$' contains x as a free variable. If we used '$\varphi(\ulcorner\psi(x)\urcorner)$' instead, we would not obtain the same effect, as '$\ulcorner\psi(x)\urcorner$' is just a numeral – a constant term without any free variable inside.

9 For such a predicate, the basic formula '$Prov_{Th}(x,y)$' can be defined as stating (roughly): 'x is a finite sequence such that every element of x is either an axiom of Th or a logical axiom or it can be obtained from earlier elements of the sequence by a given rule of inference'.

The following lemma is crucial in many applications.

LEMMA 1.1.9 (Diagonal lemma). Let Th be an extension of PA (possibly in a richer language). For every formula $\varphi(x)$ of the language of Th, there is a sentence ψ of the language of Th such that:

$$Th \vdash \psi \equiv \varphi(\ulcorner \psi \urcorner).$$[10]

It should be stressed that the formulation given here covers also cases in which the theory in question is formulated in a language richer than that of first-order arithmetic. In particular, the possibility of applying the diagonal lemma to truth theories (in the language with the truth predicate) will be important to us. It is worth mentioning that in such a case the theory needed to prove the biconditional '$\psi \equiv \varphi(\ulcorner \psi \urcorner)$' is a very weak extension of PA, obtained by adding to the axioms of PA just the logical axioms in the extended language.

The diagonal lemma is employed in typical proofs of two famous incompleteness theorems, which are formulated below.

THEOREM 1.1.10 (Gödel-Rosser first incompleteness theorem). Let Th be a consistent, axiomatisable extension of PA. Then there is a sentence $\psi \in L_{PA}$ such that neither ψ nor its negation is provable in Th.

The theorem gives the information that no axiomatisable, consistent extension of Peano arithmetic will decide all arithmetical sentences. The sentence ψ, independent from Th, is obtained by diagonalising Rosser's provability predicate. Given a provability predicate $Prov_{Th}(x,y)$, define:

$$Prov^{R}_{Th}(x,y) =_{def} Prov_{Th}(x,y) \wedge \forall z < x \neg Prov_{Th}(z, \neg y).$$

Rosser's provability predicate can be defined by the condition:

$$Pr^{R}_{Th}(y) =_{def} \exists x Prov^{R}_{Th}(x,y).$$

It turns out that a sentence ψ provably (in Th) equivalent to $\neg Pr^{R}_{Th}(\ulcorner \psi \urcorner)$ will be independent of Th.

A somewhat weaker result is obtained by diagonalising on an arbitrary predicate $Pr_{Th}(x)$ from Definition 1.1.7. It is known that any sentence ψ provably equivalent to $\neg Pr_{Th}(\ulcorner \psi \urcorner)$ is not provable in Th if only Th is consistent; however, the negation of such a ψ might be provable if Th is ω-inconsistent.[11] The meaning of this last notion is explained in what follows.

10 For more details and the proof, see (Hájek and Pudlák 1993, p. 158ff).

11 For details, the reader is referred to (Smoryński 1977).

DEFINITION 1.1.11. A theory Th containing PA is ω-consistent iff for every formula $\varphi(x)$ of the language of Th:

$$\text{if for every natural number } n,\ Th \vdash \varphi(\overline{n}),\ \text{then } Th \nvdash \exists x \neg \varphi(x).$$

As it happens, ω-inconsistency of a theory does not imply that the theory in question is inconsistent. However, the basic problem with ω-inconsistent theories is that even if consistent, they admit no standard interpretation – they cannot be interpreted in the standard model of arithmetic (see Observation 1.2.4).

In this book the name 'Gödel sentence' will be reserved for an arbitrary G satisfying the following condition.

DEFINITION 1.1.12. Let Th be an axiomatisable extension of PA. A Gödel sentence for Th will be an arbitrary sentence G such that

$$Th \vdash G \equiv \neg Pr_{Th}(\ulcorner G \urcorner).$$

Gödel's second incompleteness theorem concerns the unprovability of consistency. The formulation is given next.

THEOREM 1.1.13 (Gödel's second incompleteness theorem). Let Th be any axiomatisable, consistent extension of PA. Let $Pr_{Th}(x)$ be a standard provability predicate for Th (under a chosen recursive axiomatisation of Th). Denote as 'Con_{Th}' the sentence '$\neg Pr_{Th}(\ulcorner 0 = 1 \urcorner)$'. Then $Th \nvdash Con_{Th}$.

Given that the derivability conditions (see Definition 1.1.8) are satisfied, the choice of '$0 = 1$' for the characterisation of the sentence 'Con_{Th}' is not important, and any contradiction would be just as suitable.[12] The restriction to standard provability predicates (satisfying derivability conditions) in the formulation of the theorem is important. On the one hand, if the provability predicate is standard, then Con_{Th} will be equivalent (provably in Th) to an arbitrary Gödel sentence for Th, and since the latter is not provable in a consistent theory Th, the same holds for Con_{Th}. On the other hand, without such a restriction counterexamples to Theorem 1.1.13 could be given. It is known, for example, that if we take $Pr^R_{Th}(x)$ as our starting point and define 'Con^R_{Th}' as the sentence '$\neg Pr^R_{Th}(\ulcorner 0 = 1 \urcorner)$', then $Th \vdash Con^R_{Th}$.[13]

12 For an arbitrary sentence φ disprovable in Th, we have: $PA \vdash Pr_{Th}(\ulcorner \varphi \urcorner) \equiv Pr_{Th}(\ulcorner 0 = 1 \urcorner)$.

13 For more details about the second incompleteness theorem, see, e.g. (Boolos et al. 2002, p. 247ff); see also (Cieśliński 2002) and (Cieśliński and Urbaniak 2013). For the provability of Rosser consistency, see, e.g. (Smoryński 1977, p. 841).

From the incompleteness phenomena we move now to *completeness*. The next two theorems characterise an important completeness property of arithmetical theories.

THEOREM 1.1.14 (Σ_1-completeness). Every Σ_1 sentence true in the standard model of arithmetic is provable in Peano arithmetic.

For the proof, see (Rautenberg 2006, p. 186).[14] In addition, it turns out that Theorem 1.1.14 can be formalised in *PA*.

THEOREM 1.1.15 (Formalised Σ_1-completeness). There is a standard provability predicate $Pr_{PA}(x)$ such that for every Σ_1 sentence $\psi \in L_{PA}$, $PA \vdash \psi \rightarrow Pr_{PA}(\psi)$.

For details the reader is referred to Section 7.1 of (Rautenberg 2006)[15] – one of the few textbooks giving a detailed proof of the derivability conditions and formalised Σ_1-completeness of Peano arithmetic.

Let us end this section with another useful classical theorem where the assumption of the standardness of the provability predicate is essential again.

THEOREM 1.1.16 (Löb's theorem). Let *Th* be an axiomatisable, consistent extension of *PA* and let $Pr_{Th}(x)$ be a standard provability predicate. Then for every formula β of the language of *Th*:

$$Th \vdash Pr_{Th}(\ulcorner \beta \urcorner) \rightarrow \beta \text{ iff } Th \vdash \beta.$$[16]

1.2 *Model Theory*

The reader is assumed to be familiar with the concept of a mathematical structure and with the notion of truth in a model. In this book I will not use separate symbols for models and their universes. In particular, the symbol N will be employed as referring to the standard model of arithmetic but also to the set of natural numbers.

Two definitions given in what follows introduce some basic terminology. A signature (or a type) of a given mathematical structure is the information about the number and the arity of the relations, the operations and the constant elements of the structure.[17] Signatures can be assigned also to

14 Theorem 3.1 in Rautenberg's book is even stronger than that: it attributes Σ_1 completeness to Robinson's arithmetic, which is a finitely axiomatisable subtheory of *PA*.

15 See especially Theorem 1.2 on p. 215.

16 For the proof see (Boolos et al. 2002, p. 237); see also (Cieśliński 2003) for a discussion of Löb's theorem in set theory.

17 For a full definition, see (Adamowicz and Zbierski 2011, pp. 11–12).

languages and if a given language L has the same signature as a mathematical structure S, we say that S is a model of L.

DEFINITION 1.2.1. A set X is definable with parameters in a model M of the language L iff there is a formula $\varphi(x, y_1 \dots y_k) \in L$ and $a_1 \dots a_k \in M$ such that $X = \{z : M \models \varphi(z, a_1 \dots a_k)\}$.

DEFINITION 1.2.2. Let M be a structure with the same signature as a given first-order language L. We define:

- $Th(M) = \{\psi \in L : M \models \psi\}$. The set $Th(M)$ is called the theory of M.
- $L(M)$ – the language of M – is an extension of L with a set of new constants, corresponding to all elements of M. (In effect, we enrich L with the set of constants $\{c_a : a \in M\}$.)
- $ElDiag(M)$ – the elementary diagram of M – is defined as the set $\{\psi \in L(M) : M \models \psi\}$.[18]

The next definition characterises the notions of an extension and an expansion of a model. Roughly, extensions add new elements; expansions leave the old model intact, adding only interpretations of new symbols in the old model.

DEFINITION 1.2.3.

- A model M is an extension of a model K (or: K is a submodel of M) iff the universe of K is a subset of the universe of M and the relations and functions of K are just relations and functions of M restricted to the universe of K.
- A model M is an expansion of a model K iff the only difference between M and K is that M contains new relations, functions or constant elements, absent in K.

Truth-expansions of models of PA will be particularly important. Given a model $(M, +_M, \times_M, S_M, 0_M)$ of PA, I will abbreviate as (M, T) the expansion $(M, +_M, \times_M, S_M, 0_M, T_M)$ of the initial model. In such a context T will be a subset of M which serves as an interpretation of the truth predicate.

Definition 1.1.11 introduced the notion of an ω-consistent theory. We noticed that ω-inconsistency does not imply inconsistency: if ω-inconsistent theories are not attractive, it is not because they are inconsistent. The reason

18 The definition of $ElDiag(M)$ resembles that of $Th(M)$; the only difference lies in taking into account all sentences of $L(M)$ instead of L.

to be dissatisfied with ω-inconsistent theories is given in the observation that follows.

OBSERVATION 1.2.4. If Th is ω-inconsistent, then the standard model of arithmetic cannot be expanded to a model of Th.

PROOF. Assume that for all $n \in N$, $Th \vdash \varphi(\overline{n})$ but $Th \vdash \exists x \neg \varphi(x)$. Let N^* be an expansion of N such that $N^* \models Th$. Pick an a such that $N^* \models \neg\varphi(a)$. Then $a \in N$ (since N^* is an expansion of N), but this is impossible, because then by assumption $N^* \models \varphi(a)$. ⊣

Since the standard model of arithmetic is typically meant to provide the intended interpretation for theories extending PA, the lack of such an interpretation is a quite undesirable trait.

Later on I will sometimes make use of the soundness properties of PA and its extensions. In general, soundness of a theory means that theoremhood implies truth or validity. Here the emphasis will be mostly on truth of arithmetical sentences in the standard model. The definition that follows introduces the notion of soundness with respect to a given class of sentences.

DEFINITION 1.2.5. Let Γ be a class of arithmetical sentences. A theory Th is Γ-sound iff for every arithmetical sentence ψ belonging to Γ, if $Th \vdash \psi$, then ψ is true in the standard model of arithmetic.

A discussion of sets, even infinite ones, can be sometimes carried out in an arithmetical language inside a given (nonstandard) model of Peano arithmetic. Let '$y = p_x$' be an arithmetical formula with the meaning 'y is the xth prime number'; abbreviate as '$x|y$' the arithmetical formula 'x divides y'. Then we define:

DEFINITION 1.2.6. For every M, for every $a \in M$, for every set of natural numbers Z, a codes Z in M iff

$$Z = \{n : M \models p_n|a\}.$$

Instead of '$p_x|a$' I will usually write: '$x \in a$', treating the latter formula as belonging to the language of arithmetic.

This idea of coding permits to reproduce some set theory inside models of arithmetic. Observe that in the standard model of arithmetic, only finite

sets of natural numbers will be coded. On the other hand, every nonstandard model of arithmetic codes some infinite sets.[19]

The next definition will be used in the formulation of the overspill lemma further in what follows.

DEFINITION 1.2.7. A set I is an initial segment of a model M of PA iff $I \subseteq M$ and:

$$\forall x,y[(x \in I \wedge M \models y < x) \rightarrow y \in I].$$

If in addition I is a proper subset of M, it will be called a proper initial segment of M.

In a couple of places in this book the following lemma will be employed.

LEMMA 1.2.8 (Overspill). Let Th be a fully inductive extension of PA (with axioms of induction for all formulas of the language of Th), and let M be a model of Th whose arithmetical reduct is nonstandard.[20] Let I be a proper initial segment of M closed under the successor operation of M, and let $\varphi(x,\overline{a})$ be a formula of the language of Th, with $\overline{a}$ being a finite sequence of parameters from M. If

$$\text{For all } b \in I,\ M \models \varphi(b,\overline{a}),$$

then there is an element $c \in M$ such that $c > I$ (that is, for every $x \in I$, $M \models c > x$) and

$$M \models \forall x \leq c\varphi(x,\overline{a}).$$

PROOF. Fix Th, M, I and $\varphi(x,\overline{a})$ as in the formulation of the lemma. Assume that for all $b \in I$, $M \models \varphi(b,\overline{a})$. For an indirect proof, assume that for no $c > I$ the condition '$M \models \forall x \leq c\varphi(x,\overline{a})$' is satisfied. Consider the formula $\psi(x,\overline{a})$ defined as:

$$\forall y < x\varphi(y,\overline{a}).$$

Then $\psi(x,\overline{a})$ defines I in M. Since I is closed under successor, it is possible to show by induction in M that $M \models \forall x\psi(x,\overline{a})$. However, this means that $M = I$ which contradicts our assumption that I is a proper subset of M. ⊣

19 It is a known fact that sets of natural numbers coded in every nonstandard model of PA are exactly the recursive sets (see [Kaye 1991, p. 142], lemmas 11.1 and 11.2).

20 The language of Th might be richer than L_{PA}, so a model of Th might contain interpretations of some additional (non-arithmetical) symbols. Removing these additional interpretations leaves us with the arithmetical reduct of a given model.

It is important to stress the assumption of inductiveness of Th. The language of the theory Th may be richer than that of first-order arithmetic – it may contain additional relational, functional and constant symbols. The crucial assumption is that in Th induction for formulas of the extended language is available (this is what is meant by Th being a 'fully inductive extension of PA'). Without the extended induction the preceding proof does not go through; namely, we will not be able to show that $M = I$.

The next definition introduces the notion of an elementary extension.

DEFINITION 1.2.9. Let M and K be structures with the same signature as a given first-order language L. We say that M is an elementary extension of K (in symbols: $K < M$) iff M is an extension of K and for every formula $\varphi(x_1 \dots x_n) \in L$, the following condition is satisfied:

$$\forall a_1 \dots a_n \in K \; [K \models \varphi(a_1 \dots a_n) \equiv M \models \varphi(a_1 \dots a_n)].^{21}$$

Instead of 'M is an elementary extension of K', we will also say (equivalently) that K is an elementary submodel of M.

It follows in particular that if M is an elementary extension of K, then both models satisfy exactly the same sentences. This conclusion is obtained by omitting the parameters in the definition, and the point is that since sentences do not contain free variables, in the case of sentences parameters *can* be omitted.

A useful technique of building elementary submodels employs definable elements in models of arithmetic.

DEFINITION 1.2.10. For $M \models PA$ and $A \subseteq M$, we define:

- $K(M, A)$ is a model whose universe is the set of all elements of M definable with parameters from A.
- If $A = \emptyset$, the notation $K(M)$ instead of $K(M, \emptyset)$ will be used. The model $K(M)$ will be called the prime model of $Th(M)$.[22]

The reader is referred to (Kaye 1991, p. 91), where the proof is given that $K(M, A)$ is closed under the operations of the model M (that is, that $K(M, A)$ is a substructure of M). In addition, the following theorem will be useful in a couple of places in this book.

21 The expression '$K \models \varphi(a_1 \dots a_n)$' is an abbreviation of '$K \models \varphi(x_1 \dots x_n)[a_1 \dots a_n]$', which means that the formula in question is satisfied in K under a valuation assigning objects $a_1 \dots a_n$ to the variables $x_1 \dots x_n$.

22 It is possible to show that this definition of a prime model depends on $Th(M)$ but not on the choice of M. For details, see (Kaye 1991, pp. 92–93), Theorem 8.2.

THEOREM 1.2.11. For every $M \models PA$, for every $A \subseteq M$, $K(M, A) < M$.[23]

It easily follows that each element of $K(M, A)$ is definable in $K(M, A)$ with parameters from A. Observe that if M satisfies some false arithmetical sentences (that is, if $Th(M)$ is not identical with the theory of the standard model of arithmetic), then $K(M, A)$ is nonstandard.

Two definitions that follow introduce the notion of a recursive type and the concept of a recursively saturated model, crucial in many contexts in discussions concerning truth theories.

DEFINITION 1.2.12. Let Z be a set of formulas with one free variable x and with parameters $a_1 \ldots a_n$ from a model M. We say that:

(a) Z is realised in M iff there is an $s \in M$ such that every formula in Z is satisfied in M under a valuation assigning s to x.
(b) Z is a type of M iff every finite subset of Z is realised in M.
(c) Z is a recursive type of M iff apart from being a type of M, Z is also recursive.

DEFINITION 1.2.13. M is recursively saturated iff every recursive type of M is realised in M.

One of the basic facts about recursively saturated models is formulated in what follows.

FACT 1.2.14. Every infinite model M has a recursively saturated elementary extension of the same cardinality as M.[24]

Fact 1.2.14 is important: it means that it is possible (in a sense) to restrict one's attention to recursively saturated models while arguing for general conclusions. Imagine that you start with an arbitrary model of Peano arithmetic. If it is the theory of this model (the set of sentences true in the model) that matters for your aims, you could just as well pick a recursively saturated model which makes exactly the same sentences true – that is the moral.
The next definition introduces the notion of an elementary chain of models and the operation of union of such a chain.

DEFINITION 1.2.15.

1. An elementary chain of models is a family of models $\{M_n : n \in N\}$ such that for every $k, n \in N$, if $k < n$, then $M_k < M_n$.

23 For the proof, see (Kaye 1991, p. 91ff).
24 For the proof see, e.g. (Kaye 1991, p. 14), Proposition 11.4.

2. The union of the chain $\{M_n : n \in N\}$ is defined as the model M such that:
   - The universe of M is the union of all the universes of M_n-s.
   - The relations in M are the unions of the corresponding relations in M_n-s.
   - The functions in M are the unions of the corresponding functions in M_n-s.
   - The constant elements in M are the same as in M_n-s (all of these models have the same constant elements).

The theorem formulated next will be of crucial importance in Chapter 7.

THEOREM 1.2.16 (Elementary chain theorem). If $\{M_n : n \in N\}$ is an elementary chain of models, then for every $n \in N$, M_n is an elementary submodel of the union of this chain.[25]

1.3 *Conservativity*

In many places in this book the notion of a conservative extension will be used. The key definition is provided here.

DEFINITION 1.3.1. Let T_1 and T_2 be theories in languages L_1 and L_2 (with $L_1 \subseteq L_2$). Then:

(a) T_2 is syntactically conservative over T_1 iff $T_1 \subseteq T_2$ and $\forall \psi \in L_1 [T_2 \vdash \psi \rightarrow T_1 \vdash \psi]$.
(b) T_2 is semantically conservative over T_1 iff every model M of T_1 can be expanded to a model of T_2 (i.e. interpretations for new expressions of L_2 can be provided in M so that T_2 is true in the expansion of M).

If T_2 is semantically conservative over T_1, syntactical conservativity also follows. For a proof, assume that T_2 is not syntactically conservative over T_1. Then there is a sentence $\psi \in L_1$ such that $T_2 \vdash \psi$ but $T_1 \nvdash \psi$. Picking such a ψ, we see that $T_1 + \neg\psi$ is consistent, so T_1 has a model M, in which $\neg\psi$ is true. By semantic conservativity, M is expandable to a model of T_2. But then, since $T_2 \vdash \psi$, the sentence ψ must be true in M, and we obtain a contradiction.

In spite of this, these two notions of conservativeness do not coincide. Semantic conservativeness is more strict. The opposite implication does not hold, which means that it is not possible to derive semantic conservativity

25 For the proof, see (Chang and Keisler 1990, pp. 140–141).

from the mere assumption that T_2 is syntactically conservative over T_1. Illustrations in terms of truth theories will be presented in the chapters to follow; here a simple example will be given, one which does not involve truth.

Let L^c_{PA} be the language obtained by extending L_{PA} with a new constant c. Let PA^c be a theory in the language L^c_{PA} which is recursively axiomatised by the set of axioms of PA enlarged with sentences of the form '$c \neq \overline{n}$' for each $n \in N$. In other words:

$$Ax(PA^c) = Ax(PA) \cup \{c \neq \overline{n} : n \in N\}.$$

We claim that:

FACT 1.3.2. *PA^c is syntactically but not semantically conservative over PA.*

PROOF. In order to prove syntactic conservativity, fix $\psi \in L_{PA}$ and assume that $PA^c \vdash \psi$; for an indirect proof, assume also that $PA \nvdash \psi$. Then $PA + \neg\psi$ is consistent. We will show that $PA^c + \neg\psi$ is also consistent. This will end the proof, since by assumption $PA^c \vdash \psi$, therefore the negation of ψ cannot be consistently added to PA^c.

Let M be an arbitrary model of $PA + \neg\psi$ and let S be an arbitrary finite subset of $Ax(PA^c) \cup \{\neg\psi\}$. We are going to show how to interpret S in M. Let k be the largest natural number such that '$c \neq \overline{k}$' belongs to S. It is easy to observe that with c interpreted as $k+1$, all sentences in S become true in M. This shows that every finite subset of $Ax(PA^c) \cup \{\neg\psi\}$ has a model, so by compactness $Ax(PA^c) \cup \{\neg\psi\}$ has a model, hence $PA^c + \neg\psi$ is consistent. This ends the proof of the syntactic conservativity property.

For the semantic non-conservativity of PA^c over PA, it is enough to observe that the standard model N of arithmetic cannot be expanded to a model of PA^c – no interpretation of the new constant c can be found, making all the sentences $c \neq n$ true in N. ⊣

Fact 1.3.2, together with its proof, provides one of the simplest illustrations known to me of a difference between the two notions of conservativity.

Below I formulate another useful fact, providing a model-theoretic characterisation of the notion of syntactic conservativity.

FACT 1.3.3. *Let Th_1 and Th_2 be first-order theories in languages L_{Th_1} and L_{Th_2} such that $Th_1 \subseteq Th_2$. Then Th_2 is syntactically conservative over Th_1 iff for every model M of Th_1 there is a model K such that:*

- *$M \equiv_{L_{Th_1}} K$; that is, for every sentence $\psi \in L_{Th_1}$, $M \models \psi$ iff $K \models \psi$,*
- *$K \models Th_2$.*

PROOF. Fix Th_1 and Th_2 as in the formulation of the Fact. Proving the implication from left to right, assume that Th_2 is syntactically conservative over Th_1. Fixing a model M of Th_1, for the indirect proof let us assume that:

$$\forall K[K \models Th_2 \rightarrow \neg(M \equiv_{L_{Th_1}} K)].$$

Then $Th(M) \cup Th_2$ is inconsistent (with $Th(M) =$ the set of all L_{Th_1}-sentences true in M). By compactness, let us choose a finite subset A of $Th(M)$ inconsistent with Th_2. Then $Th_2 \vdash \neg A$ (since A is finite, we may treat it as a single sentence of L_{Th_1}). However, $Th_1 \nvdash \neg A$, because $M \models Th_1 \cup A$. Therefore Th_2 is not syntactically conservative over Th_1, and we obtain a contradiction.

For the opposite implication, assume that for every model M of Th_1 there is a model K satisfying the conditions from Fact 1.3.3. For an indirect proof, fix $\psi \in L_{Th_1}$ such that $Th_2 \vdash \psi$ but $Th_1 \nvdash \psi$. Then $Th_1 \cup \neg\psi$ is consistent, so there is a model M of $Th_1 \cup \neg\psi$. By assumption, we have then a model K such that $K \models Th_2$ and $K \equiv_{L_{Th_1}} M$. Therefore $K \models \neg\psi$; but since $K \models Th_2$, we have also: $K \models \psi$, which is a contradiction ending the proof. ⊣

1.4 Truth

The first definition introduces the basic notation for the language with the truth predicate.

DEFINITION 1.4.1. L_T is the language obtained from L_{PA} by enriching it with a new one-place predicate T.

After adding a new predicate to the arithmetical language, the basic theory (that is, Peano arithmetic) becomes modified as well, since it starts functioning as a theory in the new language. From now on, the expression '*PAT*' will be used to denote Peano arithmetic as formulated in L_T.

DEFINITION 1.4.2. *PAT* is a theory in the language L_T, whose axioms contain those of *PA*, together with all the logical axioms in L_T and all the substitutions of the induction schema by formulas of L_T.

It should be stressed that although some axioms of *PAT* contain the new predicate '*T*', there is absolutely no reason to consider it a *truth* predicate. In fact, in the axioms of *PAT* the new predicate is merely idling: for all we know (from the axioms), *T* could even express one of the arithmetically definable properties, including the empty one.