Incompleteness and Computability
An Open Logic Text

Spring 2017 bis
Incompleteness and Computability
The Open Logic Project

Instigator

Richard Zach, University of Calgary

Editorial Board

Aldo Antonelli, † University of California, Davis
Andrew Arana, Université Paris I Panthéon–Sorbonne
Jeremy Avigad, Carnegie Mellon University
Walter Dean, University of Warwick
Gillian Russell, University of North Carolina
Nicole Wyatt, University of Calgary
Audrey Yap, University of Victoria

Contributors

Samara Burns, University of Calgary
Dana Hägg, University of Calgary
Incompleteness and Computability

An Open Logic Text

Remixed by Richard Zach

Spring 2017 bis
The Open Logic Project would like to acknowledge the generous support of the Faculty of Arts and the Taylor Institute of Teaching and Learning of the University of Calgary, and the Alberta Open Educational Resources (ABOER) Initiative, which is made possible through an investment from the Alberta government.

Cover illustrations by Matthew Leadbeater, used under a Creative Commons Attribution-NonCommercial 4.0 International License.

Typeset in Baskervald X and Universalis ADF Standard by \LaTeX{}.

This version of phil479 is revision a3270c8 (2018-02-22), with content generated from OpenLogicProject revision 7553196 (2018-04-25).

Incompleteness and Computability by Richard Zach is licensed under a Creative Commons Attribution 4.0 International License. It is based on The Open Logic Text by the Open Logic Project, used under a Creative Commons Attribution 4.0 International License.
# Contents

1 Introduction to Incompleteness  
1.1 Historical Background ........................................... 1  
1.2 Definitions ............................................................. 7  
1.3 Overview of Incompleteness Results .............................. 14  
1.4 Undecidability and Incompleteness ............................. 16  
Summary ................................................................. 18  
Problems ................................................................. 19

2 Recursive Functions ......................................................... 20  
2.1 Introduction ............................................................. 20  
2.2 Primitive Recursion .................................................. 21  
2.3 Primitive Recursive Functions are Computable .............. 26  
2.4 Examples of Primitive Recursive Functions .................. 27  
2.5 Primitive Recursive Relations .................................... 29  
2.6 Bounded Minimization .............................................. 31  
2.7 Primes ................................................................. 33  
2.8 Sequences ............................................................. 34  
2.9 Other Recursions .................................................... 36  
2.10 Non-Primitive Recursive Functions ............................ 38  
2.11 Partial Recursive Functions ..................................... 41  
2.12 The Normal Form Theorem ....................................... 43  
2.13 The Halting Problem .............................................. 44  
2.14 General Recursive Functions .................................... 45  
Summary ................................................................. 46
CONTENTS

Problems ................................................. 47

3 Arithmetization of Syntax 49
  3.1 Introduction ...................................... 49
  3.2 Coding Symbols .................................. 51
  3.3 Coding Terms .................................... 53
  3.4 Coding Formulas .................................. 56
  3.5 Substitution ...................................... 58
  3.6 Derivations in Natural Deduction ............. 59
  Summary ................................................ 65
  Problems .............................................. 66

4 Representability in \( \mathbb{Q} \) 67
  4.1 Introduction ...................................... 67
  4.2 Functions Representable in \( \mathbb{Q} \) are Computable 70
  4.3 The Beta Function Lemma ....................... 71
  4.4 Simulating Primitive Recursion .................. 76
  4.5 Basic Functions are Representable in \( \mathbb{Q} \) ....... 77
  4.6 Composition is Representable in \( \mathbb{Q} \) .......... 80
  4.7 Regular Minimization is Representable in \( \mathbb{Q} \) .. 82
  4.8 Computable Functions are Representable in \( \mathbb{Q} \) .. 86
  4.9 Representing Relations ........................... 87
  4.10 Undecidability ................................. 88
  Summary ............................................. 90
  Problems ............................................. 90

5 Incompleteness and Provability 92
  5.1 Introduction ...................................... 92
  5.2 The Fixed-Point Lemma ......................... 94
  5.3 The First Incompleteness Theorem ............... 97
  5.4 Rosser’s Theorem ................................ 99
  5.5 Comparison with Gödel’s Original Paper ...... 101
  5.6 The Provability Conditions for \( \mathbf{PA} \) ........ 101
  5.7 The Second Incompleteness Theorem .......... 103
  5.8 Löb’s Theorem ................................... 105
  5.9 The Undefinability of Truth .................... 109
CONTENTS

A.2 Terms and Formulas ............................. 162
A.3 Free Variables and Sentences ..................... 165
A.4 Substitution .................................. 167
A.5 Structures for First-order Languages ............ 169
A.6 Satisfaction of a Formula in a Structure ........ 171
A.7 Variable Assignments ........................... 176
A.8 Extensionality .................................. 180
A.9 Semantic Notions ............................... 182
A.10 Theories ...................................... 185
Summary ............................................. 186
Problems ............................................. 187

B Natural Deduction ................................. 190
B.1 Natural Deduction ............................... 190
B.2 Rules and Derivations ........................... 192
B.3 Propositional Rules ............................. 193
B.4 Quantifier Rules ................................. 194
B.5 Derivations ..................................... 195
B.6 Examples of Derivations ......................... 197
B.7 Derivations with Quantifiers .................... 201
B.8 Derivations with Identity predicate .............. 206
B.9 Proof-Theoretic Notions ......................... 207
Summary ............................................. 210
Problems ............................................. 210

C Biographies ....................................... 212
C.1 Alonzo Church ................................... 212
C.2 Kurt Gödel ..................................... 213
C.3 Rózsa Péter ..................................... 215
C.4 Julia Robinson .................................. 217
C.5 Alfred Tarski .................................... 219

Photo Credits ....................................... 221

Bibliography ....................................... 223
1.1 Historical Background

In this section, we will briefly discuss historical developments that will help put the incompleteness theorems in context. In particular, we will give a very sketchy overview of the history of mathematical logic; and then say a few words about the history of the foundations of mathematics.

The phrase “mathematical logic” is ambiguous. One can interpret the word “mathematical” as describing the subject matter, as in, “the logic of mathematics,” denoting the principles of mathematical reasoning; or as describing the methods, as in “the mathematics of logic,” denoting a mathematical study of the principles of reasoning. The account that follows involves mathematical logic in both senses, often at the same time.

The study of logic began, essentially, with Aristotle, who lived approximately 384–322 BCE. His Categories, Prior analytics, and
Posterior analytics include systematic studies of the principles of scientific reasoning, including a thorough and systematic study of the syllogism.

Aristotle’s logic dominated scholastic philosophy through the middle ages; indeed, as late as eighteenth century Kant maintained that Aristotle’s logic was perfect and in no need of revision. But the theory of the syllogism is far too limited to model anything but the most superficial aspects of mathematical reasoning. A century earlier, Leibniz, a contemporary of Newton’s, imagined a complete “calculus” for logical reasoning, and made some rudimentary steps towards designing such a calculus, essentially describing a version of propositional logic.

The nineteenth century was a watershed for logic. In 1854 George Boole wrote *The Laws of Thought*, with a thorough algebraic study of propositional logic that is not far from modern presentations. In 1879 Gottlob Frege published his *Begriffsschrift* (Concept writing) which extends propositional logic with quantifiers and relations, and thus includes first-order logic. In fact, Frege’s logical systems included higher-order logic as well, and more. In his *Basic Laws of Arithmetic*, Frege set out to show that all of arithmetic could be derived in his Begriffsschrift from purely logical assumption. Unfortunately, these assumptions turned out to be inconsistent, as Russell showed in 1902. But setting aside the inconsistent axiom, Frege more or less invented modern logic singlehandedly, a startling achievement. Quantificational logic was also developed independently by algebraically-minded thinkers after Boole, including Peirce and Schröder.

Let us now turn to developments in the foundations of mathematics. Of course, since logic plays an important role in mathematics, there is a good deal of interaction with the developments I just described. For example, Frege developed his logic with the explicit purpose of showing that all of mathematics could be based solely on his logical framework; in particular, he wished to show that mathematics consists of a priori analytic truths instead of, as Kant had maintained, a priori synthetic ones.

Many take the birth of mathematics proper to have occurred
with the Greeks. Euclid’s *Elements*, written around 300 B.C., is already a mature representative of Greek mathematics, with its emphasis on rigor and precision. The definitions and proofs in Euclid’s *Elements* survive more or less in tact in high school geometry textbooks today (to the extent that geometry is still taught in high schools). This model of mathematical reasoning has been held to be a paradigm for rigorous argumentation not only in mathematics but in branches of philosophy as well. (Spinoza even presented moral and religious arguments in the Euclidean style, which is strange to see!)

Calculus was invented by Newton and Leibniz in the seventeenth century. (A fierce priority dispute raged for centuries, but most scholars today hold that the two developments were for the most part independent.) Calculus involves reasoning about, for example, infinite sums of infinitely small quantities; these features fueled criticism by Bishop Berkeley, who argued that belief in God was no less rational than the mathematics of his time. The methods of calculus were widely used in the eighteenth century, for example by Leonhard Euler, who used calculations involving infinite sums with dramatic results.

In the nineteenth century, mathematicians tried to address Berkeley’s criticisms by putting calculus on a firmer foundation. Efforts by Cauchy, Weierstrass, Bolzano, and others led to our contemporary definitions of limits, continuity, differentiation, and integration in terms of “epsilons and deltas,” in other words, devoid of any reference to infinitesimals. Later in the century, mathematicians tried to push further, and explain all aspects of calculus, including the real numbers themselves, in terms of the natural numbers. (Kronecker: “God created the whole numbers, all else is the work of man.”) In 1872, Dedekind wrote “Continuity and the irrational numbers,” where he showed how to “construct” the real numbers as sets of rational numbers (which, as you know, can be viewed as pairs of natural numbers); in 1888 he wrote “Was sind und was sollen die Zahlen” (roughly, “What are the natural numbers, and what should they be?”) which aimed to explain the natural numbers in purely “logical” terms. In 1887 Kro-
necker wrote “Über den Zahlbegriff” (“On the concept of number”) where he spoke of representing all mathematical object in terms of the integers; in 1889 Giuseppe Peano gave formal, symbolic axioms for the natural numbers.

The end of the nineteenth century also brought a new boldness in dealing with the infinite. Before then, infinitary objects and structures (like the set of natural numbers) were treated gingerly; “infinitely many” was understood as “as many as you want,” and “approaches in the limit” was understood as “gets as close as you want.” But Georg Cantor showed that it was possible to take the infinite at face value. Work by Cantor, Dedekind, and others help to introduce the general set-theoretic understanding of mathematics that is now widely accepted.

This brings us to twentieth century developments in logic and foundations. In 1902 Russell discovered the paradox in Frege’s logical system. In 1904 Zermelo proved Cantor’s well-ordering principle, using the so-called “axiom of choice”; the legitimacy of this axiom prompted a good deal of debate. Between 1910 and 1913 the three volumes of Russell and Whitehead’s *Principia Mathematica* appeared, extending the Fregean program of establishing mathematics on logical grounds. Unfortunately, Russell and Whitehead were forced to adopt two principles that seemed hard to justify as purely logical: an axiom of infinity and an axiom of “reducibility.” In the 1900’s Poincaré criticized the use of “impredicative definitions” in mathematics, and in the 1910’s Brouwer began proposing to refound all of mathematics in an “intuitionistic” basis, which avoided the use of the law of the excluded middle ($A \lor \neg A$).

Strange days indeed! The program of reducing all of mathematics to logic is now referred to as “logicism,” and is commonly viewed as having failed, due to the difficulties mentioned above. The program of developing mathematics in terms of intuitionistic mental constructions is called “intuitionism,” and is viewed as posing overly severe restrictions on everyday mathematics. Around the turn of the century, David Hilbert, one of the most influential mathematicians of all time, was a strong sup-
porter of the new, abstract methods introduced by Cantor and Dedekind: “no one will drive us from the paradise that Cantor has created for us.” At the same time, he was sensitive to foundational criticisms of these new methods (oddly enough, now called “classical”). He proposed a way of having one’s cake and eating it too:

1. Represent classical methods with formal axioms and rules; represent mathematical questions as formulas in an axiomatic system.

2. Use safe, “finitary” methods to prove that these formal deductive systems are consistent.

Hilbert’s work went a long way toward accomplishing the first goal. In 1899, he had done this for geometry in his celebrated book *Foundations of geometry*. In subsequent years, he and a number of his students and collaborators worked on other areas of mathematics to do what Hilbert had done for geometry. Hilbert himself gave axiom systems for arithmetic and analysis. Zermelo gave an axiomatization of set theory, which was expanded on by Fraenkel, Skolem, von Neumann, and others. By the mid-1920s, there were two approaches that laid claim to the title of an axiomatization of “all” of mathematics, the *Principia mathematica* of Russell and Whitehead, and what came to be known as Zermelo-Fraenkel set theory.

In 1921, Hilbert set out on a research project to establish the goal of proving these systems to be consistent. He was aided in this project by several of his students, in particular Bernays, Ackermann, and later Gentzen. The basic idea for accomplishing this goal was to cast the question of the possibility of a derivation of an inconsistency in mathematics as a combinatorial problem about possible sequences of symbols, namely possible sequences of sentences which meet the criterion of being a correct derivation of, say, \( A \land \neg A \) from the axioms of an axiom system for arithmetic, analysis, or set theory. A proof of the impossibility
of such a sequence of symbols would—since it is itself a mathematical proof—be formalizable in these axiomatic systems. In other words, there would be some sentence $\text{Con}$ which states that, say, arithmetic is consistent. Moreover, this sentence should be provable in the systems in question, especially if its proof requires only very restricted, “finitary” means.

The second aim, that the axiom systems developed would settle every mathematical question, can be made precise in two ways. In one way, we can formulate it as follows: For any sentence $A$ in the language of an axiom system for mathematics, either $A$ or $\neg A$ is provable from the axioms. If this were true, then there would be no sentences which can neither be proved nor refuted on the basis of the axioms, no questions which the axioms do not settle. An axiom system with this property is called complete. Of course, for any given sentence it might still be a difficult task to determine which of the two alternatives holds. But in principle there should be a method to do so. In fact, for the axiom and derivation systems considered by Hilbert, completeness would imply that such a method exists—although Hilbert did not realize this. The second way to interpret the question would be this stronger requirement: that there be a mechanical, computational method which would determine, for a given sentence $A$, whether it is derivable from the axioms or not.

In 1931, Gödel proved the two “incompleteness theorems,” which showed that this program could not succeed. There is no axiom system for mathematics which is complete, specifically, the sentence that expresses the consistency of the axioms is a sentence which can neither be proved nor refuted.

This struck a lethal blow to Hilbert’s original program. However, as is so often the case in mathematics, it also opened up exciting new avenues for research. If there is no one, all-encompassing formal system of mathematics, it makes sense to develop more circumscribed systems and investigate what can be proved in them. It also makes sense to develop less restricted methods of proof for establishing the consistency of these systems, and to find ways to measure how hard it is to prove their consistency. Since Gödel
showed that (almost) every formal system has questions it cannot settle, it makes sense to look for “interesting” questions a given formal system cannot settle, and to figure out how strong a formal system has to be to settle them. To the present day, logicians have been pursuing these questions in a new mathematical discipline, the theory of proofs.

1.2 Definitions

In order to carry out Hilbert’s project of formalizing mathematics and showing that such a formalization is consistent and complete, the first order of business would be that of picking a language, logical framework, and a system of axioms. For our purposes, let us suppose that mathematics can be formalized in a first-order language, i.e., that there is some set of constant symbols, function symbols, and predicate symbols which, together with the connectives and quantifiers of first-order logic, allow us to express the claims of mathematics. Most people agree that such a language exists: the language of set theory, in which \( \in \) is the only non-logical symbol. That such a simple language is so expressive is of course a very implausible claim at first sight, and it took a lot of work to establish that practically all mathematics can be expressed in this very austere vocabulary. To keep things simple, for now, let’s restrict our discussion to arithmetic, so the part of mathematics that just deals with the natural numbers \( \mathbb{N} \). The natural language in which to express facts of arithmetic is \( \mathcal{L}_A \). \( \mathcal{L}_A \) contains a single two-place predicate symbol \( < \), a single constant symbol \( 0 \), one one-place function symbol \( \cdot \), and two two-place function symbols \( + \) and \( \times \).
**Definition 1.1.** A set of sentences $\Gamma$ is a *theory* if it is closed under entailment, i.e., if $\Gamma = \{A : \Gamma \models A\}$.

There are two easy ways to specify theories. One is as the set of sentences true in some structure. For instance, consider the structure for $\mathcal{L}_A$ in which the domain is $\mathbb{N}$ and all non-logical symbols are interpreted as you would expect.

**Definition 1.2.** The *standard model of arithmetic* is the structure $\mathcal{N}$ defined as follows:

1. $|\mathcal{N}| = \mathbb{N}$
2. $0^\mathcal{N} = 0$
3. $\cdot^n\mathcal{N}(n) = n + 1$ for all $n \in \mathbb{N}$
4. $\cdot^\mathcal{N}(n, m) = n + m$ for all $n, m \in \mathbb{N}$
5. $\times^\mathcal{N}(n, m) = n \cdot m$ for all $n, m \in \mathbb{N}$
6. $<^\mathcal{N} = \{\langle n, m \rangle : n \in \mathbb{N}, m \in \mathbb{N}, n < m\}$

**Definition 1.3.** The theory of *true arithmetic* is the set of sentences satisfied in the standard model of arithmetic, i.e.,

$$\mathcal{T}_A = \{A : \mathcal{N} \models A\}.$$  

$\mathcal{T}_A$ is a theory, for whenever $\mathcal{T}_A \models A$, $A$ is satisfied in every structure which satisfies $\mathcal{T}_A$. Since $M \models \mathcal{T}_A$, $M \models A$, and so $A \in \mathcal{T}_A$.

The other way to specify a theory $\Gamma$ is as the set of sentences entailed by some set of sentences $\Gamma_0$. In that case, $\Gamma$ is the “closure” of $\Gamma_0$ under entailment. Specifying a theory this way is only interesting if $\Gamma_0$ is explicitly specified, e.g., if the elements of $\Gamma_0$ are listed. At the very least, $\Gamma_0$ has to be decidable, i.e., there
has to be a computable test for when a sentence counts as an element of \( \Gamma_0 \) or not. We call the sentences in \( \Gamma_0 \) axioms for \( \Gamma \), and \( \Gamma \) axiomatized by \( \Gamma_0 \).

**Definition 1.4.** A theory \( \Gamma \) is *axiomatized* by \( \Gamma_0 \) iff 
\[
\Gamma = \{ A : \Gamma_0 \models A \}
\]

**Definition 1.5.** The theory \( Q \), axiomatized by the following sentences is known as “Robinson’s \( Q \)” and is a very simple theory of arithmetic.
\[
\begin{align*}
\forall x \forall y (x' = y' \to x = y) & \quad (Q_1) \\
\forall x (x \neq 0) & \quad (Q_2) \\
\forall x (x \neq 0 \to \exists y x = y') & \quad (Q_3) \\
\forall x (x + 0) = x & \quad (Q_4) \\
\forall x \forall y (x + y') = (x + y)' & \quad (Q_5) \\
\forall x (x \times 0) = 0 & \quad (Q_6) \\
\forall x \forall y (x \times y') = ((x \times y) + x) & \quad (Q_7) \\
\forall x \forall y (x < y \leftrightarrow \exists z (x + z' = y)) & \quad (Q_8)
\end{align*}
\]

The set of sentences \( \{Q_1, \ldots, Q_8\} \) are the axioms of \( Q \), so \( Q \) consists of all sentences entailed by them:
\[
Q = \{ A : \{Q_1, \ldots, Q_8\} \models A \}.
\]

**Definition 1.6.** Suppose \( A(x) \) is a formula in \( \mathcal{L}_A \) with free variables \( x \) and \( y_1, \ldots, y_n \). Then any sentence of the form
\[
\forall y_1 \ldots \forall y_n ((A(0) \land \forall x (A(x) \to A(x'))) \to \forall x A(x))
\]
is an instance of the *induction schema.*
Peano arithmetic PA is the theory axiomatized by the axioms of Q together with all instances of the induction schema.

Every instance of the induction schema is true in N. This is easiest to see if the formula A only has one free variable x. Then A(x) defines a subset X_A of N in N. X_A is the set of all n ∈ N such that N, s ⊨ A(x) when s(x) = n. The corresponding instance of the induction schema is

\[(\text{A(0)} \land \forall x (A(x) \rightarrow A(x'))) \rightarrow \forall x A(x)\]

If its antecedent is true in N, then 0 ∈ X_A and, whenever n ∈ X_A, so is n + 1. Since 0 ∈ X_A, we get 1 ∈ X_A. With 1 ∈ X_A we get 2 ∈ X_A. And so on. So for every n ∈ N, n ∈ X_A. But this means that \(\forall x A(x)\) is satisfied in N.

Both Q and PA are axiomatized theories. The big question is, how strong are they? For instance, can PA prove all the truths about N that can be expressed in \(\mathcal{L}_A\)? Specifically, do the axioms of PA settle all the questions that can be formulated in \(\mathcal{L}_A\)?

Another way to put this is to ask: Is PA = TA? For TA obviously does prove (i.e., it includes) all the truths about N, and it settles all the questions that can be formulated in \(\mathcal{L}_A\), since if A is a sentence in \(\mathcal{L}_A\), then either N ⊨ A or N ⊨ \neg A, and so either TA ⊨ A or TA ⊨ \neg A. Call such a theory complete.

**Definition 1.7.** A theory Γ is complete iff for every sentence A in its language, either \(\Gamma \models A\) or \(\Gamma \models \neg A\).

By the Completeness Theorem, \(\Gamma \models A\) iff \(\Gamma \vdash A\), so Γ is complete iff for every sentence A in its language, either \(\Gamma \vdash A\) or \(\Gamma \vdash \neg A\).

Another question we are led to ask is this: Is there a computational procedure we can use to test if a sentence is in TA, in PA, or even just in Q? We can make this more precise by defining when a set (e.g., a set of sentences) is decidable.
1.2. DEFINITIONS

**Definition 1.8.** A set $X$ is *decidable* iff its characteristic function $\chi_X : X \rightarrow \{0, 1\}$, with

$$
\chi_X(x) = \begin{cases} 
1 & \text{if } x \in X \\
0 & \text{if } x \notin X
\end{cases}
$$

is computable.

So our question becomes: Is $\text{TA} (\text{PA}, \text{Q})$ decidable?

The answer to all these questions will be: no. None of these theories is decidable. However, this phenomenon is not specific to these particular theories. In fact, *any* theory that satisfies certain conditions is subject to the same results. One of these conditions, which $\text{Q}$ and $\text{PA}$ satisfy, is that they are axiomatized by a decidable set of axioms.

**Definition 1.9.** A theory is *axiomatizable* if it is axiomatized by a decidable set of axioms.

**Example 1.10.** Any theory axiomatized by a finite set of sentences is axiomatizable, since any finite set is decidable. Thus, $\text{Q}$, for instance, is axiomatizable.

Schematically axiomatized theories like $\text{PA}$ are also axiomatizable. For to test if $B$ is among the axioms of $\text{PA}$, i.e., to compute the function $\chi_X$ where $\chi_X(B) = 1$ if $B$ is an axiom of $\text{PA}$ and $= 0$ otherwise, we can do the following: First, check if $B$ is one of the axioms of $\text{Q}$. If it is, the answer is “yes” and the value of $\chi_X(B) = 1$. If not, test if it is an instance of the induction schema. This can be done systematically; in this case, perhaps it’s easiest to see that it can be done as follows: Any instance of the induction schema begins with a number of universal quantifiers, and then a sub-formula that is a conditional. The consequent of that conditional is $\forall x A(x, y_1, \ldots, y_n)$ where $x$ and $y_1, \ldots, y_n$ are all the free variables of $A$ and the initial quantifiers of $B$ bind the variables $y_1, \ldots, y_n$. Once we have extracted this $A$ and checked
that its free variables match the variables bound by the universal quantifiers at the front and $\forall x$, we go on to check that the antecedent of the conditional matches

$$A(0, y_1, \ldots, y_n) \land \forall x (A(x, y_1, \ldots, y_n) \rightarrow A(x', y_1, \ldots, y_n))$$

Again, if it does, $B$ is an instance of the induction schema, and if it doesn’t, $B$ isn’t.

In answering this question—and the more general question of which theories are complete or decidable—it will be useful to consider also the following definition. Recall that a set $X$ is countable iff it is empty or if there is a surjective function $f : \mathbb{N} \rightarrow X$. Such a function is called an enumeration of $X$.

**Definition 1.11.** A set $X$ is called *computably enumerable* (c.e. for short) iff it is empty or it has a computable enumeration.

In addition to axiomatizability, another condition on theories to which the incompleteness theorems apply will be that they are strong enough to prove basic facts about computable functions and decidable relations. By “basic facts,” we mean sentences which express what the values of computable functions are for each of their arguments. And by “strong enough” we mean that the theories in question count these sentences among its theorems. For instance, consider a prototypical computable function: addition. The value of $+\!$ for arguments 2 and 3 is 5, i.e., $2 + 3 = 5$. A sentence in the language of arithmetic that expresses that the value of $+\!$ for arguments 2 and 3 is 5 is: $(2 + 3) = 5$. And, e.g., $\mathcal{Q}$ proves this sentence. More generally, we would like there to be, for each computable function $f(x_1, x_2)$ a formula $A_f(x_1, x_2, y)$ in $\mathcal{L}_A$ such that $\mathcal{Q} \vdash A_f(n_1, n_2, m)$ whenever $f(n_1, n_2) = m$. In this way, $\mathcal{Q}$ proves that the value of $f$ for arguments $n_1, n_2$ is $m$. In fact, we require that it proves a bit more, namely that no other number is the value of $f$ for arguments $n_1, n_2$. And the same goes for decidable relations. This is made precise in the following two definitions.
**Definition 1.12.** A formula $A(x_1, \ldots, x_k, y)$ represents the function $f: \mathbb{N}^k \to \mathbb{N}$ in $\Gamma$ iff whenever $f(n_1, \ldots, n_k) = m$, then

1. $\Gamma \vdash A(\overline{n_1}, \ldots, \overline{n_k}, \overline{m})$, and
2. $\Gamma \vdash \forall y(A(\overline{n_1}, \ldots, \overline{n_k}, y) \rightarrow y = \overline{m})$.

**Definition 1.13.** A formula $A(x_1, \ldots, x_k)$ represents the relation $R \subseteq \mathbb{N}^k$ iff,

1. whenever $R(n_1, \ldots, n_k)$, $\Gamma \vdash A(\overline{n_1}, \ldots, \overline{n_k})$, and
2. whenever not $R(n_1, \ldots, n_k)$, $\Gamma \vdash \neg A(\overline{n_1}, \ldots, \overline{n_k})$.

A theory is “strong enough” for the incompleteness theorems to apply if it represents all computable functions and all decidable relations. $Q$ and its extensions satisfy this condition, but it will take us a while to establish this—it’s a non-trivial fact about the kinds of things $Q$ can prove, and it’s hard to show because $Q$ has only a few axioms from which we’ll have to prove all these facts. However, $Q$ is a very weak theory. So although it’s hard to prove that $Q$ represents all computable functions, most interesting theories are stronger than $Q$, i.e., prove more than $Q$ does. And if $Q$ proves something, any stronger theory does; since $Q$ represents all computable functions, every stronger theory does. This means that many interesting theories meet this condition of the incompleteness theorems. So our hard work will pay off, since it shows that the incompleteness theorems apply to a wide range of theories. Certainly, any theory aiming to formalize “all of mathematics” must prove everything that $Q$ proves, since it should at the very least be able to capture the results of elementary computations. So any theory that is a candidate for a theory of “all of mathematics” will be one to which the incompleteness theorems apply.
1.3 Overview of Incompleteness Results

Hilbert expected that mathematics could be formalized in an axiomatizable theory which it would be possible to prove complete and decidable. Moreover, he aimed to prove the consistency of this theory with very weak, “finitary,” means, which would defend classical mathematics against the challenges of intuitionism. Gödel’s incompleteness theorems showed that these goals cannot be achieved.

Gödel’s first incompleteness theorem showed that a version of Russell and Whitehead’s *Principia Mathematica* is not complete. But the proof was actually very general and applies to a wide variety of theories. This means that it wasn’t just that *Principia Mathematica* did not manage to completely capture mathematics, but that no acceptable theory does. It took a while to isolate the features of theories that suffice for the incompleteness theorems to apply, and to generalize Gödel’s proof to apply make it depend only on these features. But we are now in a position to state a very general version of the first incompleteness theorem for theories in the language $L_A$ of arithmetic.

**Theorem 1.14.** If $\Gamma$ is a consistent and axiomatizable theory in $L_A$ which represents all computable functions and decidable relations, then $\Gamma$ is not complete.

To say that $\Gamma$ is not complete is to say that for at least one sentence $A$, $\Gamma \nvdash A$ and $\Gamma \nvdash \neg A$. Such a sentence is called *independent* (of $\Gamma$). We can in fact relatively quickly prove that there must be independent sentences. But the power of Gödel’s proof of the theorem lies in the fact that it exhibits a *specific example* of such an independent sentence. The intriguing construction produces a sentence $G_{\Gamma}$, called a *Gödel sentence* for $\Gamma$, which is unprovable because in $\Gamma$, $G_{\Gamma}$ is equivalent to the claim that $G_{\Gamma}$ is unprovable in $\Gamma$. It does so *constructively*, i.e., given an axiomatization of $\Gamma$ and a description of the proof system, the proof gives a method for actually writing down $G_{\Gamma}$. 
The construction in Gödel’s proof requires that we find a way to express in $\mathcal{L}_A$ the properties of and operations on terms and formulas of $\mathcal{L}_A$ itself. These include properties such as “$A$ is a sentence,” “$\delta$ is a derivation of $A$,” and operations such as $A[t/x]$. This way must (a) express these properties and relations via a “coding” of symbols and sequences thereof (which is what terms, formulas, derivations, etc. are) as natural numbers (which is what $\mathcal{L}_A$ can talk about). It must (b) do this in such a way that $\Gamma$ will prove the relevant facts, so we must show that these properties are coded by decidable properties of natural numbers and the operations correspond to computable functions on natural numbers. This is called “arithmetization of syntax.”

Before we investigate how syntax can be arithmetized, however, we will consider the condition that $\Gamma$ is “strong enough,” i.e., represents all computable functions and decidable relations. This requires that we give a precise definition of “computable.” This can be done in a number of ways, e.g., via the model of Turing machines, or as those functions computable by programs in some general-purpose programming language. Since our aim is to represent these functions and relations in a theory in the language $\mathcal{L}_A$, however, it is best to pick a simple definition of computability of just numerical functions. This is the notion of recursive function. So we will first discuss the recursive functions. We will then show that $\mathsf{Q}$ already represents all recursive functions and relations. This will allow us to apply the incompleteness theorem to specific theories such as $\mathsf{Q}$ and $\mathsf{PA}$, since we will have established that these are examples of theories that are “strong enough.”

The end result of the arithmetization of syntax is a formula $\text{Prov}_\Gamma(x)$ which, via the coding of formulas as numbers, expresses provability from the axioms of $\Gamma$. Specifically, if $A$ is coded by the number $n$, and $\Gamma \vdash A$, then $\Gamma \vdash \text{Prov}_\Gamma(n)$. This “provability predicate” for $\Gamma$ allows us also to express, in a certain sense, the consistency of $\Gamma$ as a sentence of $\mathcal{L}_A$: let the “consistency statement” for $\Gamma$ be the sentence $\neg \text{Prov}_\Gamma(n)$, where we take $n$ to be the code of a contradiction, e.g., of $\bot$. The second incom-
pleteteness theorem states that consistent axiomatizable theories also do not prove their own consistency statements. The conditions required for this theorem to apply are a bit more stringent than just that the theory represents all computable functions and decidable relations, but we will show that $\mathbf{PA}$ satisfies them.

### 1.4 Undecidability and Incompleteness

Gödel’s proof of the incompleteness theorems require arithmetization of syntax. But even without that we can obtain some nice results just on the assumption that a theory represents all decidable relations. The proof is a diagonal argument similar to the proof of the undecidability of the halting problem.

**Theorem 1.15.** If $\Gamma$ is a consistent theory that represents every decidable relation, then $\Gamma$ is not decidable.

**Proof.** Suppose $\Gamma$ were decidable. We show that if $\Gamma$ represents every decidable relation, it must be inconsistent.

Decidable properties (one-place relations) are represented by formulas with one free variable. Let $A_0(x), A_1(x), \ldots$, be a computable enumeration of all such formulas. Now consider the following set $D \subseteq \mathbb{N}$:

$$D = \{ n : \Gamma \vdash \neg A_n(\overline{n}) \}$$

The set $D$ is decidable, since we can test if $n \in D$ by first computing $A_n(x)$, and from this $\neg A_n(\overline{n})$. Obviously, substituting the term $\overline{n}$ for every free occurrence of $x$ in $A_n(x)$ and prefixing $A(\overline{n})$ by $\neg$ is a mechanical matter. By assumption, $\Gamma$ is decidable, so we can test if $\neg A(\overline{n}) \in \Gamma$. If it is, $n \in D$, and if it isn’t, $n \notin D$. So $D$ is likewise decidable.

Since $\Gamma$ represents all decidable properties, it represents $D$. And the formulas which represent $D$ in $\Gamma$ are all among $A_0(x), A_1(x), \ldots$. So let $d$ be a number such that $A_d(x)$ represents $D$ in $\Gamma$. If $d \notin D$, then, since $A_d(x)$ represents $D$, $\Gamma \vdash \neg A_d(\overline{d})$. 

But that means that \( d \) meets the defining condition of \( D \), and so \( d \in D \). This contradicts \( d \notin D \). So by indirect proof, \( d \in D \).

Since \( d \in D \), by the definition of \( D \), \( \Gamma \vdash \neg A_d(d) \). On the other hand, since \( A_d(x) \) represents \( D \) in \( \Gamma \), \( \Gamma \vdash A_d(d) \). Hence, \( \Gamma \) is inconsistent.

The preceding theorem shows that no theory that represents all decidable relations can be decidable. We will show that \( Q \) does represent all decidable relations; this means that all theories that include \( Q \), such as \( \text{PA} \) and \( \text{TA} \), also do, and hence also are not decidable.

We can also use this result to obtain a weak version of the first incompleteness theorem. Any theory that is axiomatizable and complete is decidable. Consistent theories that are axiomatizable and represent all decidable properties then cannot be complete.

**Theorem 1.16.** If \( \Gamma \) is axiomatizable and complete it is decidable.

**Proof.** Any inconsistent theory is decidable, since inconsistent theories contain all sentences, so the answer to the question “is \( A \in \Gamma \)” is always “yes,” i.e., can be decided.

So suppose \( \Gamma \) is consistent, and furthermore is axiomatizable, and complete. Since \( \Gamma \) is axiomatizable, it is computably enumerable. For we can enumerate all the correct derivations from the axioms of \( \Gamma \) by a computable function. From a correct derivation we can compute the sentence it derives, and so together there is a computable function that enumerates all theorems of \( \Gamma \). A sentence is a theorem of \( \Gamma \) iff \( \neg A \) is not a theorem, since \( \Gamma \) is consistent and complete. We can therefore decide if \( A \in \Gamma \) as follows. Enumerate all theorems of \( \Gamma \). When \( A \) appears on this list, we know that \( \Gamma \vdash A \). When \( \neg A \) appears on this list, we know that \( \Gamma \nvdash A \). Since \( \Gamma \) is complete, one of these cases eventually obtains, so the procedure eventually produces and answer. \( \square \)
**Corollary 1.17.** If $\Gamma$ is consistent, axiomatizable, and represents every decidable property, it is not complete.

*Proof.* If $\Gamma$ were complete, it would be decidable by the previous theorem (since it is axiomatizable and consistent). But since $\Gamma$ represents every decidable property, it is not decidable, by the first theorem. □

Once we have established that, e.g., $Q$, represents all decidable properties, the corollary tells us that $Q$ must be incomplete. However, its proof does not provide an example of an independent sentence; it merely shows that such a sentence must exist. For this, we have to arithmetize syntax and follow Gödel’s original proof idea. And of course, we still have to show the first claim, namely that $Q$ does, in fact, represent all decidable properties.

**Summary**

Hilbert’s program aimed to show that all of mathematics could be formalized in an axiomatized theory in a formal language, such as the language of arithmetic or of set theory. He believed that such a theory would be **complete**. That is, for every sentence $A$, either $T \vdash A$ or $T \vdash \neg A$. In this sense then, $T$ would have settled every mathematical question: it would either prove that it’s true or that it’s false. If Hilbert had been right, it would also have turned out that mathematics is **decidable**. That’s because any axiomatizable theory is **computably enumerable**, i.e., there is a computable function that lists all its theorems. We can test if a sentence $A$ is a theorem by listing all of them until we find $A$ (in which it is a theorem) or $\neg A$ (in which case it isn’t). Alas, Hilbert was wrong. Gödel proved that no axiomatizable, consistent theory that is “strong enough” is complete. That’s the **first incompleteness theorem**. The requirement that the theory be “strong enough” amounts to it representing all computable functions and relations. Specifically, the very weak theory $Q$ satisfies
this property, and any theory that is at least as strong as $\mathbf{Q}$ also does. He also showed—that is the second incompleteness theorem—that the sentence that expresses the consistency of the theory is itself undecidable in it, i.e., the theory proves neither it nor its negation. So Hilbert’s further aim of finding “finitary” consistency proof of all of mathematics cannot be realized. For any finitary consistency proof would, presumably, be formalizable in a theory that captures all of mathematics. Finally, we established that theories that represent all computable functions and relations are not decidable. Note that although axiomatizability and completeness implies decidability, incompleteness does not imply undecidability. So this result shows that the second of Hilbert’s goals, namely that there be a procedure that decides if $T \vdash A$ or not, can also not be achieved, at least not for theories at least as strong as $\mathbf{Q}$.

**Problems**

**Problem 1.1.** Show that $TA = \{A : N \models A\}$ is not axiomatizable. You may assume that $TA$ represents all decidable properties.
Introduction

In order to develop a mathematical theory of computability, one has to first of all develop a model of computability. We now think of computability as the kind of thing that computers do, and computers work with symbols. But at the beginning of the development of theories of computability, the paradigmatic example of computation was numerical computation. Mathematicians were always interested in number-theoretic functions, i.e., functions \( f : \mathbb{N}^n \to \mathbb{N} \) that can be computed. So it is not surprising that at the beginning of the theory of computability, it was such functions that were studied. The most familiar examples of computable numerical functions, such as addition, multiplication, exponentiation (of natural numbers) share an interesting feature: they can be defined recursively. It is thus quite natural to attempt a general definition of computable function on the basis of recursive definitions. Among the many possible ways to define number-theoretic functions recursively, one particularly simple pattern of definition here becomes central: so-called primitive recursion.
In addition to computable functions, we might be interested in computable sets and relations. A set is computable if we can compute the answer to whether or not a given number is an element of the set, and a relation is computable iff we can compute whether or not a tuple \( \langle n_1, \ldots, n_k \rangle \) is an element of the relation. By considering the characteristic function of a set or relation, discussion of computable sets and relations can be subsumed under that of computable functions. Thus we can define primitive recursive relations as well, e.g., the relation “\( n \) evenly divides \( m \)” is a primitive recursive relation.

Primitive recursive functions—those that can be defined using just primitive recursion—are not, however, the only computable number-theoretic functions. Many generalizations of primitive recursion have been considered, but the most powerful and widely-accepted additional way of computing functions is by unbounded search. This leads to the definition of partial recursive functions, and a related definition to general recursive functions. General recursive functions are computable and total, and the definition characterizes exactly the partial recursive functions that happen to be total. Recursive functions can simulate every other model of computation (Turing machines, lambda calculus, etc.) and so represent one of the many accepted models of computation.

## 2.2 Primitive Recursion

Suppose we specify that a certain function \( l \) from \( \mathbb{N} \) to \( \mathbb{N} \) satisfies the following two clauses:

\[
\begin{align*}
l(0) &= 1 \\
l(x + 1) &= 2 \cdot l(x).
\end{align*}
\]

It is pretty clear that there is only one function, \( l \), that meets these two criteria. This is an instance of a definition by primitive recursion. We can define even more fundamental functions like addition and multiplication by

\[
\begin{align*}
f(x, 0) &= x
\end{align*}
\]
\[ f(x, y + 1) = f(x, y) + 1 \]

and

\[
\begin{align*}
g(x, 0) &= 0 \\
g(x, y + 1) &= f(g(x, y), x).
\end{align*}
\]

Exponentiation can also be defined recursively, by

\[
\begin{align*}
h(x, 0) &= 1 \\
h(x, y + 1) &= g(h(x, y), x).
\end{align*}
\]

We can also compose functions to build more complex ones; for example,

\[
k(x) = x^x + (x + 3) \cdot x = f(h(x, x), g(f(x, 3), x)).
\]

Let zero\(_x\) be the function that always returns 0, regardless of what \( x \) is, and let succ\(_x\) = \( x + 1 \) be the successor function. The set of **primitive recursive functions** is the set of functions from \( \mathbb{N}^n \) to \( \mathbb{N} \) that you get if you start with zero and succ by iterating the two operations above, primitive recursion and composition. The idea is that primitive recursive functions are defined in a straightforward and explicit way, so that it is intuitively clear that each one can be computed using finite means.

**Definition 2.1.** If \( f \) is a \( k \)-place function and \( g_0, \ldots, g_{k-1} \) are \( l \)-place functions on the natural numbers, the composition of \( f \) with \( g_0, \ldots, g_{k-1} \) is the \( l \)-place function \( h \) defined by

\[
h(x_0, \ldots, x_{l-1}) = f(g_0(x_0, \ldots, x_{l-1}), \ldots, g_{k-1}(x_0, \ldots, x_{l-1})).
\]

**Definition 2.2.** If \( f \) is a \( k \)-place function and \( g \) is a \((k + 2)\)-place function function, then the function defined by **primitive recursion from \( f \)**
and \( g \) is the \((k + 1)\)-place function \( h \) defined by the equations

\[
\begin{align*}
  h(0, z_0, \ldots, z_{k-1}) &= f(z_0, \ldots, z_{k-1}) \\
  h(x + 1, z_0, \ldots, z_{k-1}) &= g(x, h(x, z_0, \ldots, z_{k-1}), z_0, \ldots, z_{k-1})
\end{align*}
\]

In addition to zero and succ, we will include among primitive recursive functions the projection functions,

\[
P^n_i(x_0, \ldots, x_{n-1}) = x_i,
\]

for each natural number \( n \) and \( i < n \). These are not terribly exciting in themselves: \( P^n_i \) is simply the \( k \)-place function that always returns its \( i \)th argument. But the allow us to define new functions by disregarding arguments or switching arguments, as we’ll see later.

In the end, we have the following:

**Definition 2.3.** The set of primitive recursive functions is the set of functions from \( \mathbb{N}^n \) to \( \mathbb{N} \), defined inductively by the following clauses:

1. zero is primitive recursive.
2. succ is primitive recursive.
3. Each projection function \( P^n_i \) is primitive recursive.
4. If \( f \) is a \( k \)-place primitive recursive function and \( g_0, \ldots, g_{k-1} \) are \( l \)-place primitive recursive functions, then the composition of \( f \) with \( g_0, \ldots, g_{k-1} \) is primitive recursive.
5. If \( f \) is a \( k \)-place primitive recursive function and \( g \) is a \( k + 2 \)-place primitive recursive function, then the function defined by primitive recursion from \( f \) and \( g \) is primitive recursive.

Put more concisely, the set of primitive recursive functions is the smallest set containing zero, succ, and the projection func-
tions $P^n_j$, and which is closed under composition and primitive recursion.

Another way of describing the set of primitive recursive functions keeps track of the “stage” at which a function enters the set. Let $S_0$ denote the set of starting functions: zero, succ, and the projections. Once $S_i$ has been defined, let $S_{i+1}$ be the set of all functions you get by applying a single instance of composition or primitive recursion to functions in $S_i$. Then

$$S = \bigcup_{i \in \mathbb{N}} S_i$$

is the set of all primitive recursive functions.

Our definition of composition may seem too rigid, since $g_0, \ldots, g_{k-1}$ are all required to have the same arity $l$. (Remember that the *arity* of a function is the number of arguments; an $l$-place function has arity $l$.) But adding the projection functions provides the desired flexibility. For example, suppose $f$ and $g$ are 3-place functions and $h$ is the 2-place function defined by

$$h(x, y) = f(x, g(x, x, y), y).$$

The definition of $h$ can be rewritten with the projection functions, as

$$h(x, y) = f(P^2_0(x, y), g(P^2_0(x, y), P^2_0(x, y), P^2_1(x, y)), P^2_1(x, y)).$$

Then $h$ is the composition of $f$ with $P^2_0$, $l$, and $P^2_1$, where

$$l(x, y) = g(P^2_0(x, y), P^2_0(x, y), P^2_1(x, y)),$$

i.e., $l$ is the composition of $g$ with $P^2_0$, $P^2_0$, and $P^2_1$.

For another example, let us again consider addition. This is described recursively by the following two equations:

$$x + 0 = x$$

$$x + (y + 1) = \text{succ}(x + y).$$
In other words, addition is the function $\text{add}$ defined recursively by the equations

\[
\begin{align*}
\text{add}(0, x) &= x \\
\text{add}(y + 1, x) &= \text{succ}(\text{add}(y, x)).
\end{align*}
\]

But even this is not a strict primitive recursive definition; we need to put it in the form

\[
\begin{align*}
\text{add}(0, x) &= f(x) \\
\text{add}(y + 1, x) &= g(y, \text{add}(y, x), x)
\end{align*}
\]

for some 1-place primitive recursive function $f$ and some 3-place primitive recursive function $g$. We can take $f$ to be $P^1_0$, and we can define $g$ using composition,

\[
g(y, w, x) = \text{succ}(P^3_1(y, w, x)).
\]

The function $g$, being the composition of basic primitive recursive functions, is primitive recursive; and hence so is $h$. (Note that, strictly speaking, we have defined the function $g(y, x)$ meeting the recursive specification of $x + y$; in other words, the variables are in a different order. Luckily, addition is commutative, so here the difference is not important; otherwise, we could define the function $g'$ by

\[
g'(x, y) = g(P^2_1(y, x)), P^2_0(y, x)) = g(y, x),
\]

using composition.

One advantage to having the precise description of the primitive recursive functions is that we can be systematic in describing them. For example, we can assign a “notation” to each such function, as follows. Use symbols zero, succ, and $P^n_i$ for zero, successor, and the projections. Now suppose $f$ is defined by composition from a $k$-place function $h$ and $l$-place functions $g_0, \ldots, g_{k-1}$, and we have assigned notations $H, G_0, \ldots, G_{k-1}$ to the latter functions. Then, using a new symbol $\text{Comp}_{k,l}$, we can denote
the function \( f \) by \( \text{Comp}_{k,l}[H, G_0, \ldots, G_{k-1}] \). For the functions
defined by primitive recursion, we can use analogous notations
of the form \( \text{Rec}_k[G, H] \), where \( k \) denotes that arity of the func-
tion being defined. With this setup, we can denote the addition
function by
\[
\text{Rec}_2[P^1_0, \text{Comp}_{1,3}[\text{succ}, P^3_1]].
\]
Having these notations sometimes proves useful.

### 2.3 Primitive Recursive Functions are Computable

Suppose a function \( h \) is defined by primitive recursion
\[
\begin{align*}
  h(0, \vec{z}) &= f(\vec{z}) \\
  h(x + 1, \vec{z}) &= g(x, h(x, \vec{z}), \vec{z})
\end{align*}
\]
and suppose the functions \( f \) and \( g \) are computable. Then \( h(0, \vec{z}) \)
can obviously be computed, since it is just \( f(\vec{z}) \) which we assume
is computable. \( h(1, \vec{z}) \) can then also be computed, since \( 1 = 0 + 1 \)
and so \( h(1, \vec{z}) \) is just
\[
g(0, h(0, \vec{z}), \vec{z}) = g(0, f(\vec{z}), \vec{z}).
\]
We can go on in this way and compute
\[
\begin{align*}
  h(2, \vec{z}) &= g(1, g(0, f(\vec{z}), \vec{z}), \vec{z}) \\
  h(3, \vec{z}) &= g(2, g(1, g(0, f(\vec{z}), \vec{z}), \vec{z}), \vec{z}) \\
  h(4, \vec{z}) &= g(3, g(2, g(1, g(0, f(\vec{z}), \vec{z}), \vec{z}), \vec{z}), \vec{z}) \\
  &\vdots
\end{align*}
\]
Thus, to compute \( h(x, \vec{z}) \) in general, successively compute \( h(0, \vec{z}) \),
\( h(1, \vec{z}) \), \ldots, until we reach \( h(x, \vec{z}) \).

Thus, primitive recursion yields a new computable function if
the functions \( f \) and \( g \) are computable. Composition of functions
also results in a computable function if the functions \( f \) and \( g_i \) are computable.

Since the basic functions zero, succ, and \( P^n_i \) are computable, and composition and primitive recursion yield computable functions from computable functions, this means that every primitive recursive function is computable.

### 2.4 Examples of Primitive Recursive Functions

Here are some examples of primitive recursive functions:

1. Constants: for each natural number \( n \), the function that always returns \( n \) primitive recursive function, since it is equal to \( \text{succ} (\text{succ}(\ldots \text{succ}(\text{zero}(x)))) \).

2. The identity function: \( \text{id}(x) = x \), i.e. \( P^1_0 \)

3. Addition, \( x + y \)

4. Multiplication, \( x \cdot y \)

5. Exponentiation, \( x^y \) (with \( 0^0 \) defined to be 1)

6. Factorial, \( x! = 1 \cdot 2 \cdot 3 \cdots x \)

7. The predecessor function, \( \text{pred}(x) \), defined by
   \[
   \text{pred}(0) = 0, \quad \text{pred}(x + 1) = x
   \]

8. Truncated subtraction, \( x \div y \), defined by
   \[
   x \div 0 = x, \quad x \div (y + 1) = \text{pred}(x \div y)
   \]

9. Maximum, \( \text{max}(x, y) \), defined by
   \[
   \text{max}(x, y) = x + (y \div x)
   \]
10. Minimum, \( \text{min}(x, y) \)

11. Distance between \( x \) and \( y \), \( |x - y| \)

In our definitions, we’ll often use constants \( n \). This is ok because the constant function \( \text{const}_{n}(x) \) is primitive recursive (defined from zero and succ). So if, e.g., we want to define the function \( f(x) = 2 \cdot x \) can obtain it by composition from \( \text{const}_{n}(x) \) and multiplication as \( f(x) = \text{const}_{2}(x) \cdot P_{0}^{1}(x) \). We’ll make use of this trick from now on.

You’ll also have noticed that the definition of \( \text{pred} \) does not, strictly speaking, fit into the pattern of definition by primitive recursion, since that pattern requires an extra argument. It is also odd in that it does not actually \( \text{pred}(x) \) in the definition of \( \text{pred}(x + 1) \). But we can define \( \text{pred}'(x, y) \) by

\[
\text{pred}'(0, y) = \text{zero}(y) = 0 \\
\text{pred}'(x + 1, y) = P_{0}^{3}(x, \text{pred}'(x, y), y) = x
\]

and then define \( \text{pred} \) from it by composition, e.g., as \( \text{pred}(x) = \text{pred}'(P_{0}^{1}(x), \text{zero}(x)) \).

The set of primitive recursive functions is further closed under the following two operations:

1. Finite sums: if \( f(x, z) \) is primitive recursive, then so is the function

\[
g(y, z) = \sum_{x=0}^{y} f(x, z).
\]

2. Finite products: if \( f(x, z) \) is primitive recursive, then so is the function

\[
h(y, z) = \prod_{x=0}^{y} f(x, z).
\]

For example, finite sums are defined recursively by the equations

\[
g(0, z) = f(0, z), \quad g(y + 1, z) = g(y, z) + f(y + 1, z).
\]
We can also define boolean operations, where 1 stands for true, and 0 for false:

1. Negation, not($x$) = $1 \div x$
2. Conjunction, and($x$, $y$) = $x \cdot y$

Other classical boolean operations like or($x$, $y$) and ifthen($x$, $y$) can be defined from these in the usual way.

2.5 Primitive Recursive Relations

**Definition 2.4.** A relation $R(\vec{x})$ is said to be primitive recursive if its characteristic function,

$$
\chi_{R}(\vec{x}) = \begin{cases} 
1 & \text{if } R(\vec{x}) \\
0 & \text{otherwise} 
\end{cases}
$$

is primitive recursive.

In other words, when one speaks of a primitive recursive relation $R(\vec{x})$, one is referring to a relation of the form $\chi_{R}(\vec{x}) = 1$, where $\chi_{R}$ is a primitive recursive function which, on any input, returns either 1 or 0. For example, the relation IsZero($x$), which holds if and only if $x = 0$, corresponds to the function $\chi_{\text{IsZero}}$, defined using primitive recursion by

$$
\chi_{\text{IsZero}}(0) = 1, \quad \chi_{\text{IsZero}}(x + 1) = 0.
$$

It should be clear that one can compose relations with other primitive recursive functions. So the following are also primitive recursive:

1. The equality relation, $x = y$, defined by IsZero($|x - y|$)
2. The less-than relation, $x \leq y$, defined by IsZero($x - y$)

Furthermore, the set of primitive recursive relations is closed under boolean operations:
1. Negation, \( \neg P \)
2. Conjunction, \( P \land Q \)
3. Disjunction, \( P \lor Q \)
4. If \ldots\ then, \( P \rightarrow Q \)

are all primitive recursive, if \( P \) and \( Q \) are. For suppose \( \chi_P(z) \) and \( \chi_Q(z) \) are primitive recursive. Then the relation \( R(z) \) that holds iff both \( P(z) \) and \( Q(z) \) hold has the characteristic function \( \chi_R(z) = \text{and}(\chi_P(z), \chi_Q(z)) \).

One can also define relations using bounded quantification:

1. Bounded universal quantification: if \( R(x, z) \) is a primitive recursive relation, then so is the relation
   \[
   (\forall x < y) R(x, z)
   \]
   which holds if and only if \( R(x, z) \) holds for every \( x \) less than \( y \).

2. Bounded existential quantification: if \( R(x, z) \) is a primitive recursive relation, then so is
   \[
   (\exists x < y) R(x, z).
   \]

By convention, we take \((\forall x < 0) R(x, z)\) to be true (for the trivial reason that there \textit{are} no \( x \) less than 0) and \((\exists x < 0) R(x, z)\) to be false. A universal quantifier functions just like a finite product; it can also be defined directly by

\[
g(0, z) = 1, \quad g(y + 1, z) = \text{and}(g(y, z), \chi_R(y, z)).
\]

Bounded existential quantification can similarly be defined using or. Alternatively, it can be defined from bounded universal quantification, using the equivalence, \((\exists x < y) A(x) \leftrightarrow \neg(\forall x < y) \neg A(x)\). Note that, for example, a bounded quantifier of the form \((\exists x \leq y) \ldots x \ldots\) is equivalent to \((\exists x < y + 1) \ldots x \ldots\).

Another useful primitive recursive function is:
1. The conditional function, \( \text{cond}(x, y, z) \), defined by

\[
\text{cond}(x, y, z) = \begin{cases} 
  y & \text{if } x = 0 \\
  z & \text{otherwise}
\end{cases}
\]

This is defined recursively by

\[
\text{cond}(0, y, z) = y, \quad \text{cond}(x + 1, y, z) = z.
\]

One can use this to justify:

1. Definition by cases: if \( g_0(x), \ldots, g_m(x) \) are functions, and \( R_1(x), \ldots, R_{m-1}(x) \) are relations, then the function \( f \) defined by

\[
f(x) = \begin{cases} 
  g_0(x) & \text{if } R_0(x) \\
  g_1(x) & \text{if } R_1(x) \text{ and not } R_0(x) \\
  \vdots & \\
  g_{m-1}(x) & \text{if } R_{m-1}(x) \text{ and none of the previous hold} \\
  g_m(x) & \text{otherwise}
\end{cases}
\]

is also primitive recursive.

When \( m = 1 \), this is just the function defined by

\[
f(x) = \text{cond}(x \rightarrow R_0(x), g_0(x), g_1(x)).
\]

For \( m \) greater than 1, one can just compose definitions of this form.

### 2.6 Bounded Minimization

It is often useful to define a function as the least number satisfying some property or relation \( P \). If \( P \) is decidable, we can compute this function simply by trying out all the possible numbers, 0, 1, 2, \ldots, until we find the least one satisfying \( P \). This kind of unbounded search takes us out of the realm of primitive
recursive functions. However, if we’re only interested in the least number less than some independently given bound, we stay primitive recursive. In other words, and a bit more generally, suppose we have a primitive recursive relation $R(x, z)$. Consider the function that maps $y$ and $z$ to the least $x < y$ such that $R(x, z)$. It, too, can be computed, by testing whether $R(0, z)$, $R(1, z)$, \ldots, $R(y - 1, z)$. But why is it primitive recursive?

**Proposition 2.5.** If $R(x, z)$ is primitive recursive, so is the function $m_R(y, z)$ which returns the least $x$ less than $y$ such that $R(x, z)$ holds, if there is one, and $0$ otherwise. We will write the function $m_R$ as

$$\left(\min x < y\right) R(x, z),$$

*Proof.* Note than there can be no $x < 0$ such that $R(x, z)$ since there is no $x < 0$ at all. So $m_R(x, 0) = 0$.

In case the bound is $y + 1$ we have three cases: (a) There is an $x < y$ such that $R(x, z)$, in which case $m_R(y + 1, z) = m_R(y, z)$. (b) There is no such $x$ but $R(y, z)$ holds, then $m_R(y + 1, z) = y$. (c) There is no $x < y + 1$ such that $R(x, z)$, then $m_R(y + 1, z) = 0$. So,

$$m_R(0, z) = 0$$

$$m_R(y + 1, z) = \begin{cases} m_R(y, z) & \text{if } (\exists x < y) R(x, z) \\ y & \text{otherwise, provided } R(y, z) \\ 0 & \text{otherwise.} \end{cases}$$

\[\square\]

The choice of “0 otherwise” is somewhat arbitrary. It is in fact even easier to recursively define the function $m'_R$ which returns the least $x$ less than $y$ such that $R(x, z)$ holds, and $y + 1$ otherwise. When we use min, however, we will always know that the least $x$ such that $R(x, z)$ exists and is less than $y$. Thus, in practice, we will not have to worry about the possibility that if $(\min x < y) R(x, z) = 0$ we do not know if that value indicates that $R(0, z)$ or that for no $x < y$, $R(x, z)$. As with bounded quantification, $(\min x \leq y) \ldots$ can be understood as $(\min x < y + 1) \ldots$.
2.7 Primes

Bounded quantification and bounded minimization provide us with a good deal of machinery to show that natural functions and relations are primitive recursive. For example, consider the relation relation “$x$ divides $y$”, written $x \mid y$. $x \mid y$ holds if division of $x$ by $y$ is possible without remainder, i.e., if $y$ is an integer multiple of $x$. (If it doesn’t hold, i.e., the remainder when dividing $x$ by $y$ is $>0$, we write $x \nmid y$.) In other words, $x \mid y$ iff for some $z$, $x \cdot z = y$. Obviously, any such $z$, if it exists, must be $\leq y$. So, we have that $x \mid y$ iff for some $z \leq y$, $x \cdot z = y$. We can define the relation $x \mid y$ by bounded existential quantification from = and multiplication by

$$x \mid y \iff (\exists z \leq y) (x \cdot z) = y.$$ 

We’ve thus shown that $x \mid y$ is primitive recursive.

A natural number $x$ is prime if it is neither 0 nor 1 and is only divisible by 1 and itself. In other words, prime numbers are such that, whenever $y \mid x$, either $y = 1$ or $y = x$. To test if $x$ is prime, we only have to check if $y \mid x$ for all $y \leq x$, since if $y > x$, then automatically $y \nmid x$. So, the relation $\text{Prime}(x)$, which holds iff $x$ is prime, can be defined by

$$\text{Prime}(x) \iff x \geq 2 \land (\forall y \leq x) (y \mid x \rightarrow y = 1 \lor y = x)$$

and is thus primitive recursive.

The primes are 2, 3, 5, 7, 11, etc. Consider the function $p(x)$ which returns the $x$th prime in that sequence, i.e., $p(0) = 2$, $p(1) = 3$, $p(2) = 5$, etc. (For convenience we will often write $p(x)$ as $p_x$ ($p_0 = 2$, $p_1 = 3$, etc.)

If we had a function $\text{nextPrime}(x)$, which returns the first prime number larger than $x$, $p$ can be easily defined using primitive recursion:

$$p(0) = 2$$
$$p(x + 1) = \text{nextPrime}(p(x))$$
Since nextPrime\(x\) is the least \(y\) such that \(y > x\) and \(y\) is prime, it can be easily computed by unbounded search. But it can also be defined by bounded minimization, thanks to a result due to Euclid: there is always a prime number between \(x\) and \(x! + 1\).

\[
\text{nextPrime}(x) = (\min y \leq x! + 1) (y > x \land \text{Prime}(y)).
\]

This shows, that nextPrime\(x\) and hence \(p(x)\) are (not just computable but) primitive recursive.

(If you’re curious, here’s a quick proof of Euclid’s theorem. Suppose \(p_n\) is the largest prime \(\leq x\) and consider the product \(p = p_0 \cdot p_1 \cdots \cdot p_n\) of all primes \(\leq x\). Either \(p + 1\) is prime or there is a prime between \(x\) and \(p + 1\). Why? Suppose \(p + 1\) is not prime. Then some prime number \(q \mid p + 1\) where \(q < p + 1\). None of the primes \(\leq x\) divide \(p + 1\). (By definition of \(p\), each of the primes \(p_i \leq x\) divides \(p\), i.e., with remainder 0. So, each of the primes \(p_i \leq x\) divides \(p + 1\) with remainder 1, and so \(p_i \mid p + 1\).) Hence, \(q\) is a prime \(> x\) and \(< p + 1\). And \(p \leq x!\), so there is a prime \(> x\) and \(\leq x! + 1\).)

## 2.8 Sequences

The set of primitive recursive functions is remarkably robust. But we will be able to do even more once we have developed an adequate means of handling sequences. We will identify finite sequences of natural numbers with natural numbers in the following way: the sequence \(\langle a_0, a_1, a_2, \ldots, a_k \rangle\) corresponds to the number

\[
p_0^{a_0+1} \cdot p_1^{a_1+1} \cdot p_2^{a_2+1} \cdots \cdot p_k^{a_k+1}.
\]

We add one to the exponents to guarantee that, for example, the sequences \(\langle 2, 7, 3 \rangle\) and \(\langle 2, 7, 3, 0, 0 \rangle\) have distinct numeric codes. We can take both 0 and 1 to code the empty sequence; for concreteness, let \(\Lambda\) denote 0.

Let us define the following functions:
1. **len(s)**, which returns the length of the sequence *s*: Let *R(i, s)* be the relation defined by

\[ R(i, s) \text{ iff } p_i \mid s \land (\forall j < s) (j > i \rightarrow p_j \nmid s) \]

*R* is primitive recursive. Now let

\[ \text{len}(s) = \begin{cases} 0 & \text{if } s = 0 \text{ or } s = 1 \\ 1 + (\min i < s) R(i, s) & \text{otherwise} \end{cases} \]

Note that we need to bound the search on *i*; clearly *s* provides an acceptable bound.

2. **append(s, a)**, which returns the result of appending *a* to the sequence *s*:

\[ \text{append}(s, a) = \begin{cases} 2^{a+1} & \text{if } s = 0 \text{ or } s = 1 \\ s \cdot p_{\text{len}(s)}^{a+1} & \text{otherwise} \end{cases} \]

3. **element(s, i)**, which returns the *i*th element of *s* (where the initial element is called the 0th), or 0 if *i* is greater than or equal to the length of *s*:

\[ \text{element}(s, i) = \begin{cases} 0 & \text{if } i \geq \text{len}(s) \\ \min j < s (p_j^{i+2} \nmid s) - 1 & \text{otherwise} \end{cases} \]

Instead of using the official names for the functions defined above, we introduce a more compact notation. We will use \((s)_i\) instead of **element(s, i)**, and \(<s_0, \ldots, s_k>\) to abbreviate

\[ \text{append}(\text{append}(\ldots \text{append}(\Lambda, s_0) \ldots), s_k). \]

Note that if *s* has length *k*, the elements of *s* are \((s)_0, \ldots, (s)_{k-1}\).

It will be useful for us to be able to bound the numeric code of a sequence in terms of its length and its largest element. Suppose *s* is a sequence of length *k*, each element of which is less than equal to some number *x*. Then *s* has at most *k* prime factors,
each at most $p_{k-1}$, and each raised to at most $x + 1$ in the prime factorization of $s$. In other words, if we define

$$\text{sequenceBound}(x, k) = p_{k-1}^{k-(x+1)},$$

then the numeric code of the sequence $s$ described above is at most $\text{sequenceBound}(x, k)$.

Having such a bound on sequences gives us a way of defining new functions using bounded search. For example, suppose we want to define the function $\text{concat}(s, t)$, which concatenates two sequences. One first option is to define a “helper” function $\text{hconcat}(s, t, n)$ which concatenates the first $n$ symbols of $t$ to $s$. This function can be defined by primitive recursion, as follows:

$$\text{hconcat}(s, t, 0) = s$$
$$\text{hconcat}(s, t, n + 1) = \text{append}(\text{hconcat}(s, t, n), (t)_n)$$

Then we can define $\text{concat}$ by

$$\text{concat}(s, t) = \text{hconcat}(s, t, \text{len}(t)).$$

But using bounded search, we can be lazy. All we need to do is write down a primitive recursive specification of the object (number) we are looking for, and a bound on how far to look. The following works:

$$\text{concat}(s, t) = (\min v < \text{sequenceBound}(s + t, \text{len}(s) + \text{len}(t)))$$
$$\text{len}(v) = \text{len}(s) + \text{len}(t) \land$$
$$(\forall i < \text{len}(s)) ((v)_i = (s)_i) \land$$
$$(\forall j < \text{len}(t)) ((v)_{\text{len}(s)+j} = (t)_j))$$

We will write $s \rightsquigarrow t$ instead of $\text{concat}(s, t)$.

### 2.9 Other Recursions

Using pairing and sequencing, we can justify more exotic (and useful) forms of primitive recursion. For example, it is often useful to define two functions simultaneously, such as in the following
definition:

\[
\begin{align*}
f_0(0, \bar{z}) &= k_0(\bar{z}) \\
f_1(0, \bar{z}) &= k_1(\bar{z}) \\
f_0(x + 1, \bar{z}) &= h_0(x, f_0(x, \bar{z}), f_1(x, \bar{z}), \bar{z}) \\
f_1(x + 1, \bar{z}) &= h_1(x, f_0(x, \bar{z}), f_1(x, \bar{z}), \bar{z})
\end{align*}
\]

This is an instance of *simultaneous recursion*. Another useful way of defining functions is to give the value of \( f(x + 1, \bar{z}) \) in terms of *all* the values \( f(0, \bar{z}), \ldots, f(x, \bar{z}) \), as in the following definition:

\[
\begin{align*}
f(0, \bar{z}) &= g(\bar{z}) \\
f(x + 1, \bar{z}) &= h(x, \langle f(0, \bar{z}), \ldots, f(x, \bar{z}) \rangle, \bar{z}).
\end{align*}
\]

The following schema captures this idea more succinctly:

\[
f(x, \bar{z}) = h(x, \langle f(0, \bar{z}), \ldots, f(x - 1, \bar{z}) \rangle)
\]

with the understanding that the second argument to \( h \) is just the empty sequence when \( x \) is 0. In either formulation, the idea is that in computing the “successor step,” the function \( f \) can make use of the entire sequence of values computed so far. This is known as a *course-of-values* recursion. For a particular example, it can be used to justify the following type of definition:

\[
f(x, \bar{z}) = \begin{cases} h(x, f(k(x, \bar{z}), \bar{z}), \bar{z}) & \text{if } k(x, \bar{z}) < x \\ g(x, \bar{z}) & \text{otherwise} \end{cases}
\]

In other words, the value of \( f \) at \( x \) can be computed in terms of the value of \( f \) at any previous value, given by \( k \).

You should think about how to obtain these functions using ordinary primitive recursion. One final version of primitive recursion is more flexible in that one is allowed to change the *parameters* (side values) along the way:

\[
\begin{align*}
f(0, \bar{z}) &= g(\bar{z}) \\
f(x + 1, \bar{z}) &= h(x, f(x, k(\bar{z})), \bar{z})
\end{align*}
\]
This, too, can be simulated with ordinary primitive recursion. (Doing so is tricky. For a hint, try unwinding the computation by hand.)

Finally, notice that we can always extend our “universe” by defining additional objects in terms of the natural numbers, and defining primitive recursive functions that operate on them. For example, we can take an integer to be given by a pair \( \langle m, n \rangle \) of natural numbers, which, intuitively, represents the integer \( m - n \). In other words, we say

\[
\text{Integer}(x) \Leftrightarrow \text{length}(x) = 2
\]

and then we define the following:

1. \( \text{iequal}(x, y) \)
2. \( \text{iplus}(x, y) \)
3. \( \text{iminus}(x, y) \)
4. \( \text{itimes}(x, y) \)

Similarly, we can define a rational number to be a pair \( \langle x, y \rangle \) of integers with \( y \neq 0 \), representing the value \( x/y \). And we can define \( \text{qequal} \), \( \text{qplus} \), \( \text{qminus} \), \( \text{qtimes} \), \( \text{qdivides} \), and so on.

## 2.10 Non-Primitive Recursive Functions

The primitive recursive functions do not exhaust the intuitively computable functions. It should be intuitively clear that we can make a list of all the unary primitive recursive functions, \( f_0, f_1, f_2, \ldots \) such that we can effectively compute the value of \( f_x \) on input \( y \); in other words, the function \( g(x, y) \), defined by

\[
g(x, y) = f_x(y)
\]

is computable. But then so is the function

\[
h(x) = g(x, x) + 1
\]
For each primitive recursive function $f_i$, the value of $h$ and $f_i$ differ at $i$. So $h$ is computable, but not primitive recursive; and one can say the same about $g$. This is a an “effective” version of Cantor’s diagonalization argument.

One can provide more explicit examples of computable functions that are not primitive recursive. For example, let the notation $g^n(x)$ denote $g(g(\ldots g(x)))$, with $n$ g’s in all; and define a sequence $g_0, g_1, \ldots$ of functions by

$$
g_0(x) = x + 1
$$

$$
g_{n+1}(x) = g^n(x)
$$

You can confirm that each function $g_n$ is primitive recursive. Each successive function grows much faster than the one before; $g_1(x)$ is equal to $2x$, $g_2(x)$ is equal to $2^x \cdot x$, and $g_3(x)$ grows roughly like an exponential stack of $x$ 2’s. Ackermann’s function is essentially the function $G(x) = g_x(x)$, and one can show that this grows faster than any primitive recursive function.

Let us return to the issue of enumerating the primitive recursive functions. Remember that we have assigned symbolic notations to each primitive recursive function; so it suffices to enumerate notations. We can assign a natural number $(F)$ to each notation $F$, recursively, as follows:

$$
#(0) = \langle 0 \rangle
$$

$$
#(S) = \langle 1 \rangle
$$

$$
#(P^n_l) = \langle 2, n, i \rangle
$$

$$
#(\text{Comp}_{k,l}[H, G_0, \ldots, G_{k-1}]) = \langle 3, k, l, #(H), #(G_0), \ldots, #(G_{k-1}) \rangle
$$

$$
#(\text{Rec}_{l}[G, H]) = \langle 4, l, #(G), #(H) \rangle
$$

Here I am using the fact that every sequence of numbers can be viewed as a natural number, using the codes from the last section. The upshot is that every code is assigned a natural number. Of course, some sequences (and hence some numbers) do not
correspond to notations; but we can let $f_i$ be the unary primitive recursive function with notation coded as $i$, if $i$ codes such a notation; and the constant 0 function otherwise. The net result is that we have an explicit way of enumerating the unary primitive recursive functions.

(In fact, some functions, like the constant zero function, will appear more than once on the list. This is not just an artifact of our coding, but also a result of the fact that the constant zero function has more than one notation. We will later see that one can not computably avoid these repetitions; for example, there is no computable function that decides whether or not a given notation represents the constant zero function.)

We can now take the function $g(x, y)$ to be given by $f_x(y)$, where $f_x$ refers to the enumeration we have just described. How do we know that $g(x, y)$ is computable? Intuitively, this is clear: to compute $g(x, y)$, first “unpack” $x$, and see if it a notation for a unary function; if it is, compute the value of that function on input $y$.

You may already be convinced that (with some work!) one can write a program (say, in Java or C++) that does this; and now we can appeal to the Church-Turing thesis, which says that anything that, intuitively, is computable can be computed by a Turing machine.

Of course, a more direct way to show that $g(x, y)$ is computable is to describe a Turing machine that computes it, explicitly. This would, in particular, avoid the Church-Turing thesis and appeals to intuition. But, as noted above, working with Turing machines directly is unpleasant. Soon we will have built up enough machinery to show that $g(x, y)$ is computable, appealing to a model of computation that can be simulated on a Turing machine: namely, the recursive functions.
2.11 Partial Recursive Functions

To motivate the definition of the recursive functions, note that our proof that there are computable functions that are not primitive recursive actually establishes much more. The argument was simple: all we used was the fact was that it is possible to enumerate functions \( f_0, f_1, \ldots \) such that, as a function of \( x \) and \( y \), \( f_x(y) \) is computable. So the argument applies to any class of functions that can be enumerated in such a way. This puts us in a bind: we would like to describe the computable functions explicitly; but any explicit description of a collection of computable functions cannot be exhaustive!

The way out is to allow partial functions to come into play. We will see that it is possible to enumerate the partial computable functions. In fact, we already pretty much know that this is the case, since it is possible to enumerate Turing machines in a systematic way. We will come back to our diagonal argument later, and explore why it does not go through when partial functions are included.

The question is now this: what do we need to add to the primitive recursive functions to obtain all the partial recursive functions? We need to do two things:

1. Modify our definition of the primitive recursive functions to allow for partial functions as well.

2. Add something to the definition, so that some new partial functions are included.

The first is easy. As before, we will start with zero, successor, and projections, and close under composition and primitive recursion. The only difference is that we have to modify the definitions of composition and primitive recursion to allow for the possibility that some of the terms in the definition are not defined. If \( f \) and \( g \) are partial functions, we will write \( f(x) \downarrow \) to mean that \( f \) is defined at \( x \), i.e., \( x \) is in the domain of \( f \); and \( f(x) \uparrow \) to mean the opposite, i.e., that \( f \) is not defined at \( x \). We
will use $f(x) \simeq g(x)$ to mean that either $f(x)$ and $g(x)$ are both undefined, or they are both defined and equal. We will use these notations for more complicated terms as well. We will adopt the convention that if $h$ and $g_0, \ldots, g_k$ all are partial functions, then

$$h(g_0(\bar{x}), \ldots, g_k(\bar{x}))$$

is defined if and only if each $g_i$ is defined at $\bar{x}$, and $h$ is defined at $g_0(\bar{x}), \ldots, g_k(\bar{x})$. With this understanding, the definitions of composition and primitive recursion for partial functions is just as above, except that we have to replace “=” by “$\simeq$”.

What we will add to the definition of the primitive recursive functions to obtain partial functions is the unbounded search operator. If $f(x, \bar{z})$ is any partial function on the natural numbers, define $\mu x f(x, \bar{z})$ to be

the least $x$ such that $f(0, \bar{z}), f(1, \bar{z}), \ldots, f(x, \bar{z})$ are all defined, and $f(x, \bar{z}) = 0$, if such an $x$ exists

with the understanding that $\mu x f(x, \bar{z})$ is undefined otherwise. This defines $\mu x f(x, \bar{z})$ uniquely.

Note that our definition makes no reference to Turing machines, or algorithms, or any specific computational model. But like composition and primitive recursion, there is an operational, computational intuition behind unbounded search. When it comes to the computability of a partial function, arguments where the function is undefined correspond to inputs for which the computation does not halt. The procedure for computing $\mu x f(x, \bar{z})$ will amount to this: compute $f(0, \bar{z}), f(1, \bar{z}), f(2, \bar{z})$ until a value of 0 is returned. If any of the intermediate computations do not halt, however, neither does the computation of $\mu x f(x, \bar{z})$.

If $R(x, \bar{z})$ is any relation, $\mu x R(x, \bar{z})$ is defined to be $\mu x (1 - \chi_R(x, \bar{z}))$. In other words, $\mu x R(x, \bar{z})$ returns the least value of $x$ such that $R(x, \bar{z})$ holds. So, if $f(x, \bar{z})$ is a total function, $\mu x f(x, \bar{z})$ is the same as $\mu x (f(x, \bar{z}) = 0)$. But note that our original definition is more general, since it allows for the possibility that $f(x, \bar{z})$ is not everywhere defined (whereas, in contrast, the characteristic function of a relation is always total).
Definition 2.6. The set of partial recursive functions is the smallest set of partial functions from the natural numbers to the natural numbers (of various arities) containing zero, successor, and projections, and closed under composition, primitive recursion, and unbounded search.

Of course, some of the partial recursive functions will happen to be total, i.e., defined for every argument.

Definition 2.7. The set of recursive functions is the set of partial recursive functions that are total.

A recursive function is sometimes called “total recursive” to emphasize that it is defined everywhere.

2.12 The Normal Form Theorem

Theorem 2.8 (Kleene’s Normal Form Theorem). There is a primitive recursive relation $T(e, x, s)$ and a primitive recursive function $U(s)$, with the following property: if $f$ is any partial recursive function, then for some $e$,

$$f(x) \cong U(\mu s \ T(e, x, s))$$

for every $x$.

The proof of the normal form theorem is involved, but the basic idea is simple. Every partial recursive function has an index $e$, intuitively, a number coding its program or definition. If $f(x) \downarrow$, the computation can be recorded systematically and coded by some number $s$, and that $s$ codes the computation of $f$ on input $x$ can be checked primitive recursively using only $x$ and the definition $e$. This means that $T$ is primitive recursive. Given the full record of the computation $s$, the “upshot” of $s$ is the value of $f(x)$, and it can be obtained from $s$ primitive recursively as well.
The normal form theorem shows that only a single unbounded search is required for the definition of any partial recursive function. We can use the numbers $e$ as “names” of partial recursive functions, and write $\varphi_e$ for the function $f$ defined by the equation in the theorem. Note that any partial recursive function can have more than one index—in fact, every partial recursive function has infinitely many indices.

### 2.13 The Halting Problem

The *halting problem* in general is the problem of deciding, given the specification $e$ (e.g., program) of a computable function and a number $n$, whether the computation of the function on input $n$ halts, i.e., produces a result. Famously, Alan Turing proved that this problem itself cannot be solved by a computable function, i.e., the function

$$h(e, n) = \begin{cases} 1 & \text{if computation } e \text{ halts on input } n \\ 0 & \text{otherwise,} \end{cases}$$

is not computable.

In the context of partial recursive functions, the role of the specification of a program may be played by the index $e$ given in Kleene’s normal form theorem. If $f$ is a partial recursive function, any $e$ for which the equation in the normal form theorem holds, is an index of $f$. Given a number $e$, the normal form theorem states that

$$\varphi_e(x) \simeq U(\mu s \ T(e, x, s))$$

is partial recursive, and for every partial recursive $f : \mathbb{N} \to \mathbb{N}$, there is an $e \in \mathbb{N}$ such that $\varphi_e(x) \simeq f(x)$ for all $x \in \mathbb{N}$. In fact, for each such $f$ there is not just one, but infinitely many such $e$. The *halting function* $h$ is defined by

$$h(e, x) = \begin{cases} 1 & \text{if } \varphi_e(x) \downarrow \\ 0 & \text{otherwise.} \end{cases}$$
Note that $h(e, x) = 0$ if $\varphi_e(x) \uparrow$, but also when $e$ is not the index of a partial recursive function at all.

**Theorem 2.9.** The halting function $h$ is not partial recursive.

*Proof.* If $h$ were partial recursive, we could define

$$d(y) = \begin{cases} 1 & \text{if } h(y, y) = 0 \\ \mu x \ x \neq x & \text{otherwise.} \end{cases}$$

From this definition it follows that

1. $d(y) \downarrow$ iff $\varphi_y(y) \uparrow$ or $y$ is not the index of a partial recursive function.
2. $d(y) \uparrow$ iff $\varphi_y(y) \downarrow$.

If $h$ were partial recursive, then $d$ would be partial recursive as well. Thus, by the Kleene normal form theorem, it has an index $e_d$. Consider the value of $h(e_d, e_d)$. There are two possible cases, 0 and 1.

1. If $h(e_d, e_d) = 1$ then $\varphi_{e_d}(e_d) \downarrow$. But $\varphi_{e_d} \approx d$, and $d(e_d)$ is defined iff $h(e_d, e_d) = 0$. So $h(e_d, e_d) \neq 1$.

2. If $h(e_d, e_d) = 0$ then either $e_d$ is not the index of a partial recursive function, or it is and $\varphi_{e_d}(e_d) \uparrow$. But again, $\varphi_{e_d} \approx d$, and $d(e_d)$ is undefined iff $\varphi_{e_d}(e_d) \downarrow$.

The upshot is that $e_d$ cannot, after all, be the index of a partial recursive function. But if $h$ were partial recursive, $d$ would be too, and so our definition of $e_d$ as an index of it would be admissible. We must conclude that $h$ cannot be partial recursive. \qed

### 2.14 General Recursive Functions

There is another way to obtain a set of total functions. Say a total function $f(x, z)$ is *regular* if for every sequence of natural numbers
There is an $x$ such that $f(x, \bar{z}) = 0$. In other words, the regular functions are exactly those functions to which one can apply unbounded search, and end up with a total function. One can, conservatively, restrict unbounded search to regular functions:

**Definition 2.10.** The set of general recursive functions is the smallest set of functions from the natural numbers to the natural numbers (of various arities) containing zero, successor, and projections, and closed under composition, primitive recursion, and unbounded search applied to regular functions.

Clearly every general recursive function is total. The difference between Definition 2.10 and Definition 2.7 is that in the latter one is allowed to use partial recursive functions along the way; the only requirement is that the function you end up with at the end is total. So the word “general,” a historic relic, is a misnomer; on the surface, Definition 2.10 is less general than Definition 2.7. But, fortunately, the difference is illusory; though the definitions are different, the set of general recursive functions and the set of recursive functions are one and the same.

**Summary**

In order to show that $\mathbb{Q}$ represents all computable functions, we need a precise model of computability that we can take as the basis for a proof. There are, of course, many models of computability, such as Turing machines. One model that plays a significant role historically—it's one of the first models proposed, and is also the one used by Gödel himself—is that of the recursive functions. The recursive functions are a class of arithmetical functions—that is, their domain and range are the natural numbers—that can be defined from a few basic functions using a few operations. The basic functions are zero, succ, and the projection functions. The operations are composition, primitive recursion, and regular minimization. Composition is simply a
general version of “chaining together” functions: first apply one, then apply the other to the result. Primitive recursion defines a new function $f$ from two functions $g$, $h$ already defined, by stipulating that the value of $f$ for 0 is given by $g$, and the value for any number $n + 1$ is given by $h$ applied to $f(n)$. Functions that can be defined using just these two principles are called primitive recursive. A relation is primitive recursive iff its characteristic function is. It turns out that a whole list of interesting functions and relations is primitive recursive (such as addition, multiplication, exponentiation, divisibility), and that we can define new primitive recursive functions and relations from old ones using principles such as bounded quantification and bounded minimization. In particular, this allowed us to show that we can deal with sequences of numbers in primitive recursive ways. That is, there is a way to “code” sequences of numbers as single numbers in such a way that we can compute the $i$-the element, the length, the concatenation of two sequences, etc., all using primitive recursive functions operating on these codes. To obtain all the computable functions, we finally added definition by regular minimization to composition and primitive recursion. A function $g(x, y)$ is regular iff, for every $y$ it takes the value 0 for at least one $x$. If $f$ is regular, the least $x$ such that $g(x, y) = 0$ always exists, and can be found simply by computing all the values of $g(0, y)$, $g(1, y)$, etc., until one of them is $= 0$. The resulting function $f(y) = \mu x \ g(x, y) = 0$ is the function defined by regular minimization from $g$. It is always total and computable. The resulting set of functions are called general recursive. One version of the Church-Turing Thesis says that the computable arithmetical functions are exactly the general recursive ones.

**Problems**

**Problem 2.1.** Multiplication satisfies the recursive equations

\[
0 \cdot y = y \\
(x + 1) \cdot y = (x \cdot y) + x
\]
Give the explicit precise definition of the function \( \text{mult}(x, y) = x \cdot y \), assuming that \( \text{add}(x, y) = x + y \) is already defined. Give the complete notation for \( \text{mult} \).

**Problem 2.2.** Show that

\[
f(x, y) = 2^{2^x} \text{ } y \text{ 's}
\]

is primitive recursive.

**Problem 2.3.** Show that \( d(x, y) = \lfloor x/y \rfloor \) (i.e., division, where you disregard everything after the decimal point) is primitive recursive. When \( y = 0 \), we stipulate \( d(x, y) = 0 \). Give an explicit definition of \( d \) using primitive recursion and composition. You will have detour through an auxiliary function—you cannot use recursion on the arguments \( x \) or \( y \) themselves.

**Problem 2.4.** Suppose \( R(x, z) \) is primitive recursive. Define the function \( m'_R(y, z) \) which returns the least \( x \) less than \( y \) such that \( R(x, z) \) holds, if there is one, and \( y + 1 \) otherwise, by primitive recursion from \( \chi_R \).

**Problem 2.5.** Define integer division \( d(x, y) \) using bounded minimization.

**Problem 2.6.** Show that there is a primitive recursive function \( \text{sconcat}(s) \) with the property that

\[
\text{sconcat}(\langle s_0, \ldots, s_k \rangle) = s_0 \cdots s_k.
\]
3.1 Introduction

In order to connect computability and logic, we need a way to talk about the objects of logic (symbols, terms, formulas, derivations), operations on them, and their properties and relations, in a way amenable to computational treatment. We can do this directly, by considering computable functions and relations on symbols, sequences of symbols, and other objects built from them. Since the objects of logical syntax are all finite and built from a countable sets of symbols, this is possible for some models of computation. But other models of computation—such as the recursive functions—are restricted to numbers, their relations and functions. Moreover, ultimately we also want to be able to deal with syntax within certain theories, specifically, in theories formulated in the language of arithmetic. In these cases it is necessary to arithmetize syntax, i.e., to represent syntactic objects, operations on them, and their relations, as numbers, arithmetical functions, and arithmetical relations, respectively. The idea, which goes back to Leibniz, is to assign numbers to syntactic objects.

It is relatively straightforward to assign numbers to symbols
as their “codes.” Some symbols pose a bit of a challenge, since, e.g., there are infinitely many variables, and even infinitely many function symbols of each arity $n$. But of course it’s possible to assign numbers to symbols systematically in such a way that, say, $v_2$ and $v_3$ are assigned different codes. Sequences of symbols (such as terms and formulas) are a bigger challenge. But if can deal with sequences of numbers purely arithmetically (e.g., by the powers-of-primes coding of sequences), we can extend the coding of individual symbols to coding of sequences of symbols, and then further to sequences or other arrangements of formulas, such as derivations. This extended coding is called “Gödel numbering.” Every term, formula, and derivation is assigned a Gödel number.

By coding sequences of symbols as sequences of their codes, and by choosing a system of coding sequences that can be dealt with using computable functions, we can then also deal with Gödel numbers using computable functions. In practice, all the relevant functions will be primitive recursive. For instance, computing the length of a sequence and computing the $i$-th element of a sequence from the code of the sequence are both primitive recursive. If the number coding the sequence is, e.g., the Gödel number of a formula $A$, we immediately see that the length of a formula and the (code of the) $i$-th symbol in a formula can also be computed from the Gödel number of $A$. It is a bit harder to prove that, e.g., the property of being the Gödel number of a correctly formed term, of being the Gödel number of a correct derivation is primitive recursive. It is nevertheless possible, because the sequences of interest (terms, formulas, derivations) are inductively defined.

As an example, consider the operation of substitution. If $A$ is a formula, $x$ a variable, and $t$ a term, then $A[t/x]$ is the result of replacing every free occurrence of $x$ in $A$ by $t$. Now suppose we have assigned Gödel numbers to $A$, $x$, $t$—say, $k$, $l$, and $m$, respectively. The same scheme assigns a Gödel number to $[t/x]$, say, $n$. This mapping—of $k$, $l$, $m$ to $n$—is the arithmetical analog of the substitution operation. When the substitution operation maps $A$, $x$, $t$ to $A[t/x]$, the arithmetized substitution functions
maps the Gödel numbers $k$, $l$, $m$ to the Gödel number $n$. We will see that this function is primitive recursive.

Arithmetization of syntax is not just of abstract interest, although it was originally a non-trivial insight that languages like the language of arithmetic, which do not come with mechanisms for “talking about” languages can, after all, formalize complex properties of expressions. It is then just a small step to ask what a theory in this language, such as Peano arithmetic, can **prove** about its own language (including, e.g., whether sentences are provable or true). This leads us to the famous limitative theorems of Gödel (about unprovability) and Tarski (the undefinability of truth). But the trick of arithmetizing syntax is also important in order to prove some important results in computability theory, e.g., about the computational power of theories or the relationship between different models of computability. The arithmetization of syntax serves as a model for arithmetizing other objects and properties. For instance, it is similarly possible to arithmetize configurations and computations (say, of Turing machines). This makes it possible to simulate computations in one model (e.g., Turing machines) in another (e.g., recursive functions).

### 3.2 Coding Symbols

The basic language $\mathcal{L}$ of first order logic makes use of the symbols

$$\bot \neg \lor \land \rightarrow \forall \exists = ( ) ,$$

together with countable sets of variables and constant symbols, and countable sets of function symbols and predicate symbols of arbitrary arity. We can assign *codes* to each of these symbols in such a way that every symbol is assigned a unique number as its code, and no two different symbols are assigned the same number. We know that this is possible since the set of all symbols is countable and so there is a bijection between it and the set of natural numbers. But we want to make sure that we can recover the symbol (as well as some information about it, e.g., the arity of
a function symbol) from its code in a computable way. There are many possible ways of doing this, of course. Here is one such way, which uses primitive recursive functions. (Recall that \langle n_0, \ldots, n_k \rangle is the number coding the sequence of numbers \( n_0, \ldots, n_k \).)

**Definition 3.1.** If \( s \) is a symbol of \( \mathcal{L} \), let the symbol code \( c_s \) be defined as follows:

1. If \( s \) is among the logical symbols, \( c_s \) is given by the following table:

\[
\begin{align*}
\bot & \quad \langle 0,0 \rangle \\
\neg & \quad \langle 0,1 \rangle \\
\lor & \quad \langle 0,2 \rangle \\
\land & \quad \langle 0,3 \rangle \\
\Rightarrow & \quad \langle 0,4 \rangle \\
\forall & \quad \langle 0,5 \rangle \\
\exists & \quad \langle 0,6 \rangle \\
= & \quad \langle 0,7 \rangle \\
( & \quad \langle 0,8 \rangle \\
) & \quad \langle 0,9 \rangle \\
, & \quad \langle 0,10 \rangle
\end{align*}
\]

2. If \( s \) is the \( i \)-th variable \( v_i \), then \( c_s = \langle 1, i \rangle \).

3. If \( s \) is the \( i \)-th constant symbol \( c^n_i \), then \( c_s = \langle 2, i \rangle \).

4. If \( s \) is the \( i \)-th \( n \)-ary function symbol \( f^n_i \), then \( c_s = \langle 3, n, i \rangle \).

5. If \( s \) is the \( i \)-th \( n \)-ary predicate symbol \( P^n_i \), then \( c_s = \langle 4, n, i \rangle \).

**Proposition 3.2.** The following relations are primitive recursive:

1. \( F_n(x, n) \) iff \( x \) is the code of \( f^n_i \) for some \( i \), i.e., \( x \) is the code of an \( n \)-ary function symbol.

2. \( \text{Pred}(x, n) \) iff \( x \) is the code of \( P^n_i \) for some \( i \) or \( x \) is the code of \( = \) and \( n = 2 \), i.e., \( x \) is the code of an \( n \)-ary predicate symbol.
**Definition 3.3.** If $s_0, \ldots, s_{n-1}$ is a sequence of symbols, its *Goedel number* is $\langle c_{s_0}, \ldots, c_{s_{n-1}} \rangle$.

Note that *codes* and *Goedel numbers* are different things. For instance, the variable $v_5$ has a code $c_{v_5} = \langle 1, 5 \rangle = 2^2 \cdot 3^6$. But the variable $v_5$ considered as a term is also a sequence of symbols (of length 1). The Goedel number $\# v_5$ of the term $v_5$ is $\langle c_{v_5} \rangle = 2^{c_{v_5}+1} = 2^{2^2 \cdot 3^6 + 1}$.

**Example 3.4.** Recall that if $k_0, \ldots, k_{n-1}$ is a sequence of numbers, then the code of the sequence $\langle k_0, \ldots, k_{n-1} \rangle$ in the power-of-primes coding is

$$2^{k_0+1} \cdot 3^{k_1+1} \ldots \cdot p_{n-1}^{k_{n-1}},$$

where $p_i$ is the $i$-th prime (starting with $p_0 = 2$). So for instance, the formula $v_0 = 0$, or, more explicitly, $=(v_0, c_0)$, has the Goedel number

$$\langle c_\_, c_\_, c_{v_0}, c_\_, c_{c_{v_0}}, c_\rangle.$$ 

Here, $c_\_ = \langle 0, 7 \rangle = 2^{0+1} \cdot 3^{7+1}$, $c_{v_0}$ is $\langle 1, 0 \rangle = 2^{1+1} \cdot 3^{0+1}$, etc. So $\#=(v_0, c_0)$ is

$$2^{c_{\_}+1} \cdot 3^{c_{\_}+1} \cdot 5^{c_{v_0}+1} \cdot 7^{c_{\_}+1} \cdot 11^{c_{c_{v_0}}+1} \cdot 13^{c_{\_}+1} =$$

$$2^{2^{1\cdot 3^8+1}} \cdot 3^{2^{1\cdot 3^9+1}} \cdot 5^{2^{1\cdot 3^{11}+1}} \cdot 7^{2^{1\cdot 3^{11}+1}} \cdot 11^{2^{3\cdot 3^1}+1} \cdot 13^{2^{1\cdot 3^{10}+1}} =$$

$$2^{13^{123 \cdot 3^{39^{367 \cdot 5^{13 \cdot 7^{354^{295 \cdot 11^{25}}}}}} \cdot 13^{118^{099}}}.$$ 

### 3.3 Coding Terms

A term is simply a certain kind of sequence of symbols: it is built up inductively from constants and variables according to the formation rules for terms. Since sequences of symbols can be coded as numbers—using a coding scheme for the symbols plus a way to code sequences of numbers—assigning Goedel numbers to terms is not difficult. The challenge is rather to show that the property a number has if it is the Goedel number of a correctly formed term is computable, or in fact primitive recursive.
Proposition 3.5. The relations \(\text{Term}(x)\) and \(\text{ClTerm}(x)\) which hold iff \(x\) is the Gödel number of a term or a closed term, respectively, are primitive recursive.

Proof. A sequence of symbols \(s\) is a term iff there is a sequence \(s_0, \ldots, s_{k-1} = s\) of terms which records how the term \(s\) was formed from constant symbols and variables according to the formation rules for terms. To express that such a putative formation sequence follows the formation rules it has to be the case that, for each \(i < k\), either

1. \(s_i\) is a variable \(v_j\), or
2. \(s_i\) is a constant symbol \(c_j\), or
3. \(s_i\) is built from \(n\) terms \(t_1, \ldots, t_n\) occurring prior to place \(i\) using an \(n\)-place function symbol \(f^n_j\).

To show that the corresponding relation on Gödel numbers is primitive recursive, we have to express this condition primitive recursively, i.e., using primitive recursive functions, relations, and bounded quantification.

Suppose \(y\) is the number that codes the sequence \(s_0, \ldots, s_{k-1}\), i.e., \(y = \langle \#s_0\#, \ldots, \#s_{k-1}\# \rangle\). It codes a formation sequence for the term with Gödel number \(x\) iff for all \(i < k\):

1. there is a \(j\) such that \((y)_i = \#v_j\#, or
2. there is a \(j\) such that \((y)_i = \#c_j\#, or
3. there is an \(n\) and a number \(z = \langle z_1, \ldots, z_n \rangle\) such that each \(z_l\) is equal to some \((y)_{i'}\) for \(i' < i\) and
\[
(y)_i = \#f^n_j (\# \text{flatten}(z) \#)\#
\]
and moreover \((y)_{k-1} = x\). The function \(\text{flatten}(z)\) turns the sequence \(\langle \#t_1\#, \ldots, \#t_n\# \rangle\) into \(\#t_1, \ldots, t_n\#\) and is primitive recursive.
The indices $j$, $n$, the Gödel numbers $z_l$ of the terms $t_l$, and the code $z$ of the sequence $\langle z_1, \ldots, z_n \rangle$, in (3) are all less than $y$. We can replace $k$ above with $\text{len}(y)$. Hence we can express “$y$ is the code of a formation sequence of the term with Gödel number $x$” in a way that shows that this relation is primitive recursive.

We now just have to convince ourselves that there is a primitive recursive bound on $y$. But if $x$ is the Gödel number of a term, it must have a formation sequence with at most $\text{len}(x)$ terms (since every term in the formation sequence of $s$ must start at some place in $s$, and no two subterms can start at the same place). The Gödel number of each subterm of $s$ is of course $\leq x$. Hence, there always is a formation sequence with code $\leq x^{\text{len}(x)}$.

For ClTerm, simply leave out the clause for variables. □

**Alternative proof of Proposition 3.5.** The inductive definition says that constant symbols and variables are terms, and if $t_1, \ldots, t_n$ are terms, then so is $f^n_j(t_1, \ldots, t_n)$, for any $n$ and $j$. So terms are formed in stages: constant symbols and variables at stage 0, terms involving one function symbol at stage 1, those involving at least two nested function symbols at stage 2, etc. Let’s say that a sequence of symbols $s$ is a term of level $l$ iff $s$ can be formed by applying the inductive definition of terms $l$ (or fewer) times, i.e., it “becomes” a term by stage $l$ or before. So $s$ is a term of level $l+1$ iff

1. $s$ is a variable $v_j$, or
2. $s$ is a constant symbol $c_j$, or
3. $s$ is built from $n$ terms $t_1, \ldots, t_n$ of level $l$ and an $n$-place function symbol $f^n_j$.

To show that the corresponding relation on Gödel numbers is primitive recursive, we have to express this condition primitive recursively, i.e., using primitive recursive functions, relations, and bounded quantification.

The number $x$ is the Gödel number of a term $s$ of level $l + 1$ iff
1. there is a \( j \) such that \( x = \# v_j \), or
2. there is a \( j \) such that \( x = \# c_j \), or
3. there is an \( n \), a \( j \), and a number \( z = \langle z_1, \ldots, z_n \rangle \) such that each \( z_i \) is the Gödel number of a term of level \( l \) and

\[
x = \# f^n_j (\# \text{flatten}(z) \#),
\]

and moreover \( (y)_{k-1} = x \).

The indices \( j, n \), the Gödel numbers \( z_i \) of the terms \( t_i \), and the code \( z \) of the sequence \( \langle z_1, \ldots, z_n \rangle \), in (3) are all less than \( x \). So we get a primitive recursive definition by:

\[
\text{ITerm}(x, 0) = \text{Var}(x) \lor \text{Const}(x)
\]
\[
\text{ITerm}(x, l + 1) = \text{Var}(x) \lor \text{Const}(x) \lor
(\exists z < x) ((\forall i < \text{len}(z)) \text{ITerm}((z)_i, l) \land
(\exists j < x) x = (\# f^\text{len}(z)_j (\# \text{flatten}(z) \#))
\]

We can now define Term\((x)\) by ITerm\((x, x)\), since the level of a term is always less than the Gödel number of the term. □

**Proposition 3.6.** The function \( \text{num}(n) = \# n \) is primitive recursive.

**Proof.** We define \( \text{num}(n) \) by primitive recursion:

\[
\begin{align*}
\text{num}(0) & = \# 0 \\
\text{num}(n + 1) & = \# (\# \text{num}(n) \#).
\end{align*}
\]

□

### 3.4 Coding Formulas
Proposition 3.7. The relation $\text{Atom}(x)$ which holds iff $x$ is the Gödel number of an atomic formula, is primitive recursive.

Proof. The number $x$ is the Gödel number of an atomic formula iff one of the following holds:

1. There are $n$, $j < x$, and $z < x$ such that for each $i < n$, $\text{Term}((z)_i)$ and $x = \#P^n_j(\# \text{ flatten}(z) \#)\#$.

2. There are $z_1, z_2 < x$ such that $\text{Term}(z_1)$, $\text{Term}(z_2)$, and $x = \#(\#z_1 \#z_2 \#)\#$.

3. $x = \#\bot\#$.

□

Proposition 3.8. The relation $\text{Frm}(x)$ which holds iff $x$ is the Gödel number of a formula is primitive recursive.

Proof. A sequence of symbols $s$ is a formula iff there is formation sequence $s_0, \ldots, s_{k-1} = s$ of formula which records how $s$ was formed from atomic formulas according to the formation rules. The code for each $s_i$ (and indeed of the code of the sequence $\langle s_0, \ldots, s_{k-1} \rangle$ is less than the code $x$ of $s$.

□

Proposition 3.9. The relation $\text{FreeOcc}(x, z, i)$, which holds iff the $i$-th symbol of the formula with Gödel number $x$ is a free occurrence of the variable with Gödel number $z$, is primitive recursive.

Proof. Exercise.

□
**Proposition 3.10.** The property $\text{Sent}(x)$ which holds iff $x$ is the Gödel number of a sentence is primitive recursive.

*Proof.* A sentence is a formula without free occurrences of variables. So $\text{Sent}(x)$ holds iff

$$(\forall i < \text{len}(x)) (\forall z < x) ((\exists j < z) z = ^{\#}v_j \rightarrow \neg \text{FreeOcc}(x, z, i)).$$

□

### 3.5 Substitution

**Proposition 3.11.** There is a primitive recursive function $\text{Subst}(x, y, z)$ with the property that

$$\text{Subst}(^{\#}A, ^{\#}t, ^{\#}u) = {^{\#}A[t/u]}.$$

*Proof.* We can then define a function $h\text{Subst}$ by primitive recursion as follows:

\[
\begin{align*}
h\text{Subst}(x, y, z, 0) &= \Lambda \\
h\text{Subst}(x, y, z, i + 1) &= \begin{cases}
h\text{Subst}(x, y, z, i) \sim y & \text{if } \text{FreeOcc}(x, z, i + 1) \\
\text{append}(h\text{Subst}(x, y, z, i), (x)_{i+1}) & \text{otherwise.}
\end{cases}
\end{align*}
\]

$\text{Subst}(x, y, z)$ can now be defined as $h\text{Subst}(x, y, z, \text{len}(x)).$ □

**Proposition 3.12.** The relation $\text{FreeFor}(x, y, z)$, which holds iff the term with Gödel number $y$ is free for the variable with Gödel number $z$ in the formula with Gödel number $x$, is primitive recursive.

*Proof.* Exercise. □
3.6 Derivations in Natural Deduction

In order to arithmetize derivations, we must represent derivations as numbers. Since derivations are trees of formulas where each inference carries one or two labels, a recursive representation is the most obvious approach: we represent a derivation as a tuple, the components of which are the end-formula, the labels, and the representations of the sub-derivations leading to the premises of the last inference.

Definition 3.13. If $\delta$ is a derivation in natural deduction, then $\#\delta\#$ is

1. $\langle 0, A, n \rangle$ if $\delta$ consists only of the assumption $A$. The number $n$ is 0 if it is an undischarged assumption, and the numerical label otherwise.

2. $\langle 1, A, k, \#\delta_1\# \rangle$ if $\delta$ ends in an inference with one premise, $k$ is given by the following table according to which rule was used in the last inference, and $\delta_1$ is the immediate subproof ending in the premise of the last inference. $n$ is the label of the inference, or 0 if the inference does not discharge any assumptions.

   Rule: $\land$Elim $\lor$Intro $\rightarrow$Intro $\neg$Intro $\perp$I
   $k$: 1 2 3 4 5

   Rule: $\perp$C $\forall$Intro $\forall$Elim $\exists$Intro $=$Intro
   $k$: 6 7 8 9 10

3. $\langle 2, A, n, k, \#\delta_1\#, \#\delta_2\# \rangle$ if $\delta$ ends in an inference with two premises, $k$ is given by the following table according to which rule was used in the last inference, and $\delta_1$, $\delta_2$ are the immediate subderivations ending in the left and right premise of the last inference, respectively. $n$ is the label of the inference, or 0 if the inference does not discharge any assumptions.
CHAPTER 3. ARITHMETIZATION OF SYNTAX

Rule: \( \land \text{Intro} \rightarrow \text{Elim} \quad \neg \text{Elim} \)

\( k: \quad 1 \quad 2 \quad 3 \)

4. \( \langle 3, \, \#A, \, n, \, \#\delta_1, \, \#\delta_2, \, \#\delta_3 \rangle \) if \( \delta \) ends in an \( \lor \text{Elim} \) inference. \( \delta_1, \delta_2, \delta_3 \) are the immediate subderivations ending in the left, middle, and right premise of the last inference, respectively, and \( n \) is the label of the inference.

Example 3.14. Consider the very simple derivation

\[
\frac{[(A \land B)]^1}{A} \quad \land \text{Elim} \\
\frac{1}{(A \rightarrow B)} \quad \rightarrow \text{Intro}
\]

The Gödel number of the assumption would be \( d_0 = \langle 0, \#(A \land B), 1 \rangle \). The Gödel number of the derivation ending in the conclusion of \( \land \text{Elim} \) would be \( d_1 = \langle 1, \#A, 0, 1, d_0 \rangle \) (1 since \( \land \text{Elim} \) has one premise, Gödel number of conclusion \( A \), 0 because no assumption is discharged, 1 is the number coding \( \land \text{Elim} \)). The Gödel number of the entire derivation then is \( \langle 1, \#(A \rightarrow B), 1, 3, d_1 \rangle \), i.e.,

\[
2^2 \cdot 3^{\#(A\rightarrow B)+1} \cdot 5^2 \cdot 7^4 \cdot 11^{(2^2 \cdot 3^{\#(A\rightarrow B)+1} \cdot 5^2 \cdot 7^5 \cdot 11^{(2^2 \cdot 3^{\#(A\land B)+1} \cdot 5^2)})}.
\]

Having settled on a representation of derivations, we must also show that we can manipulate such derivations primitive recursively, and express their essential properties and relations so. Some operations are simple: e.g., given a Gödel number \( d \) of a derivation, \( (d)_1 \) gives us the Gödel number of its end-formula. Some are much harder. We'll at least sketch how to do this. The goal is to show that the relation “\( \delta \) is a derivation of \( A \) from \( \Gamma \)” is primitive recursive on the Gödel numbers of \( \delta \) and \( A \).

Proposition 3.15. The following relations are primitive recursive:

1. \( A \) occurs as an assumption in \( \delta \) with label \( n \).

2. All assumption in \( \delta \) with label \( n \) are of the form \( A \) (i.e., we can discharge the assumption \( A \) using label \( n \) in \( \delta \)).
3. A is an undischarged assumption of \(\delta\).

4. An inference with conclusion \(A\), upper derivations \(\delta_1\) (and \(\delta_2, \delta_3\)), and discharge label \(n\) is correct.

5. \(\delta\) is a correct natural deduction derivation.

Proof. We have to show that the corresponding relations between Gödel numbers of formulas, sequences of Gödel numbers of formulas (which code sets of formulas), and Gödel numbers of derivations are primitive recursive.

1. We want to show that \(\text{Assum}(x, d, n)\), which holds if \(x\) is the Gödel number of an assumption of the derivation with Gödel number \(d\) labelled \(n\), is primitive recursive. For this we need a helper relation \(h\text{Assum}(x, d, n, i)\) which holds if the formula \(A\) with Gödel number \(x\) occurs as an initial formula with label \(n\) in the derivation with Gödel number \(d\) within \(i\) inferences up from the end-formula.

\[
\begin{align*}
h\text{Assum}(x, d, n, 0) & \iff \top \\
h\text{Assum}(x, d, n, i + 1) & \iff \\
\text{Sent}(x) & \land (d = \langle 0, x, n \rangle) \lor \\
((d)_0 = 1 & \land h\text{Assum}(x, (d)_4, n, i)) \lor \\
((d)_0 = 2 & \land (h\text{Assum}(x, (d)_4, n, i) \lor \\
& h\text{Assum}(x, (d)_5, n, i)) \lor \\
((d)_0 = 3 & \land (h\text{Assum}(x, (d)_3, n, i) \lor \\
& h\text{Assum}(x, (d)_2, n, i)) \lor h\text{Assum}(x, (d)_3, n, i))
\end{align*}
\]

If the number \(i\) is large enough, e.g., larger than the maximum number of inferences between an initial formula and the end-formula of \(\delta\), it holds of \(x, d, n\), and \(i\) iff \(A\) is an initial formula in \(\delta\) labelled \(n\). The number \(d\) itself is larger than that maximum number of inferences. So we can define

\[
\text{Assum}(x, d, n) = h\text{Assum}(x, d, n, d).
\]
2. We want to show that Discharge\((x, d, n)\), which holds if all assumptions with label \(n\) in the derivation with Gödel number \(d\) all are the formula with Gödel number \(x\). But this relation holds iff \((\forall y < d) (\text{Assum}(y, d, n) \rightarrow y = x)\).

3. An occurrence of an assumption is not open if it occurs with label \(n\) in a subderivation that ends in a rule with discharge label \(n\). Define the helper relation \(\text{hNotOpen}(x, d, n, i)\) as

\[
\text{hNotOpen}(x, d, n, 0) \iff \top
\]

\[
\text{hNotOpen}(x, d, n, i + 1) \iff
\]

\[
(d)_2 = n \lor
\]

\[
((d)_0 = 1 \land \text{hNotOpen}(x, (d)_4, n, i)) \lor
\]

\[
((d)_0 = 2 \land \text{hNotOpen}(x, (d)_4, n, i) \land
\]

\[
\text{hNotOpen}(x, (d)_5, n, i)) \lor
\]

\[
((d)_0 = 3 \land \text{hNotOpen}(x, (d)_3, n, i) \land
\]

\[
\text{hNotOpen}(x, (d)_4, n, i) \land \text{hNotOpen}(x, (d)_5, n, i))
\]

Note that all assumptions of the form \(A\) labelled \(n\) are discharged in \(\delta\) iff either the last inference of \(\delta\) discharges them (i.e., the last inference has label \(n\)), or if it is discharged in all of the immediate subderivations.

A formula \(A\) is an open assumption of \(\delta\) iff it is an initial formula of \(\delta\) (with label \(n\)) and is not discharged in \(\delta\) (by a rule with label \(n\)). We can then define \(\text{OpenAssum}(x, d)\) as

\[
(\exists n < d) (\text{Assum}(x, d, n, d) \land \neg \text{hNotOpen}(x, d, n, d)).
\]

4. Here we have to show that for each rule of inference \(R\) the relation \(\text{FollowsBy}_R(x, d_1, n)\) which holds if \(x\) is the Gödel number of the conclusion and \(d_1\) is the Gödel number of a derivation ending in the premise of a correct application of \(R\) with label \(n\) is primitive recursive, and similarly for rules with two or three premises.
The simplest case is that of the $=\text{Intro}$ rule. Here there is no premise, i.e., $d_1 = 0$. However, $A$ must be of the form $t = t$, for a closed term $t$. Here, a primitive recursive definition is

$$(\exists t < x) \ (\text{ClTerm}(t) \wedge x = (^{\#} = (^{\#} t \sim ^{\#} t \sim ^{\#} t \sim ^{\#} t \sim ^{\#})) \wedge d_1 = 0).$$

For a more complicated example, $\text{FollowsBy}_{\to \text{Intro}}(x, d_1, n)$ holds iff $A$ is of the form $(B \to C)$, the end-formula of $\delta$ is $C$, and any initial formula in $\delta$ labelled $n$ is of the form $B$. We can express this primitive recursively by

$$(\exists y < x) \ (\text{Sent}(y) \wedge \text{Discharge}(y, d_1) \wedge (\exists z < x) \ (\text{Sent}(y) \wedge (d_1 = z) \wedge x = (^{\#} (^{\#} y \sim ^{\#} z \sim ^{\#} z \sim ^{\#})))).$$

(Think of $y$ as the Gödel number of $B$ and $z$ as that of $C$.)

For another example, consider $\exists \text{Intro}$. Here, $A$ is the conclusion of a correct inference with one upper derivation iff there is a formula $B$, a closed term $t$ and a variable $x$ such that $B[t/x]$ is the end-formula of the upper derivation and $\exists x B$ is the conclusion $A$, i.e., the formula with Gödel number $x$. So $\text{FollowsBy}_{\exists \text{Intro}}(x, d_1, n)$ holds iff

$$\text{Sent}(x) \wedge (\exists y < x) (\exists v < x) (\exists t < d) (\text{Frm}(y) \wedge \text{Term}(t) \wedge \text{Var}(v) \wedge \text{FreeFor}(y, t, v) \wedge \text{Subst}(y, t, v) = (d_1)_1 \wedge x = (^{\#} \exists^y v \sim z)).$$

5. We first define a helper relation $\text{hDeriv}(d, i)$ which holds if $d$ codes a correct derivation at least to $i$ inferences up from the end sequent. $\text{hDeriv}(d, 0)$ holds always. Otherwise, $\text{hDeriv}(d, i + 1)$ iff either $d$ just codes an assumption or $d$ ends in a correct inference and the codes of the immediate sub-derivations satisfy $\text{hDeriv}(d', i)$.

$$\text{hDeriv}(d, 0) \leftrightarrow \top$$

$$\text{hDeriv}(d, i + 1) \leftrightarrow$$
\[\exists x < d \ (\exists n < d \ (\text{Sent}(x) \land d = \langle 0, x, n \rangle)) \lor ((d)_0 = 1 \land ((d)_3 = 1 \land \text{FollowsBy}_{\text{Elim}}((d)_1, (d)_4, (d)_2)) \lor \cdots \land \text{FollowsBy}_{\text{Intro}}((d)_1, (d)_4, (d)_2)) \land \text{nDeriv}((d)_4, i)) \lor ((d)_0 = 2 \land ((d)_3 = 1 \land \text{FollowsBy}_{\text{Intro}}((d)_1, (d)_4, (d)_5, (d)_2)) \lor \cdots \land \text{FollowsBy}_{\text{Intro}}((d)_1, (d)_4, (d)_2)) \land \text{hDeriv}((d)_4, i) \land \text{hDeriv}((d)_5, i)) \lor ((d)_0 = 3 \land \text{FollowsBy}_{\text{Elim}}((d)_1, (d)_3, (d)_4, (d)_5, (d)_2) \land \text{hDeriv}((d)_3, i) \land \text{hDeriv}((d)_4, i) \land \text{hDeriv}((d)_5, i))\]

This is a primitive recursive definition. Again we can define \(\text{Deriv}(d)\) as \(\text{hDeriv}(d, d)\).

\[\square\]

**Proposition 3.16.** Suppose \(\Gamma\) is a primitive recursive set of sentences. Then the relation \(\text{Prf}_\Gamma(x, y)\) expressing “\(x\) is the code of a derivation \(\delta\) of \(A\) from undischarged assumptions in \(\Gamma\) and \(y\) is the Gödel number of \(A\)” is primitive recursive.

**Proof.** Suppose “\(y \in \Gamma\)” is given by the primitive recursive predicate \(R_\Gamma(y)\). We have to show that \(\text{Prf}_\Gamma(x, y)\) which holds iff \(y\) is the Gödel number of a sentence \(A\) and \(x\) is the code of a natural deduction derivation with end formula \(A\) and all undischarged assumptions in \(\Gamma\) is primitive recursive.

By the previous proposition, the property \(\text{Deriv}(x)\) which holds iff \(x\) is the code of a correct derivation \(\delta\) in natural deduction is
primitive recursive. If \( x \) is such a code, then \((x)_1\) is the code of the end-formula of \( \delta \). Thus we can define \( \text{Prf}_\Gamma(x, y) \) by

\[
\text{Prf}_\Gamma(x, y) \iff \text{Deriv}(x) \land (x)_1 = y \land \\
(\forall z < x) (\text{OpenAssum}(z, x) \rightarrow R_\Gamma(z))
\]

\( \square \)

**Summary**

The proof of the incompleteness theorems requires that we have a way to talk about provability in a theory (such as \( \text{PA} \)) in the language of the theory itself, i.e., in the language of arithmetic. But the language of arithmetic only deals with numbers, not with formulas or derivations. The solution to this problem is to define a systematic mapping from formulas and derivations to numbers. The number associated with a formula or a derivation is called its **Gödel number**. If \( A \) is a formula, \( \#A \) is its Gödel number. By defining the Gödel numbers in a suitable way, we can show that a whole lot of important operations on and properties of formulas and derivations turn into primitive recursive functions and relations on the respective Gödel numbers. For instance, \( A[t/x] \), the operation of substituting a term \( t \) for every free occurrence of \( x \) in \( A \), corresponds to an arithmetical function \( \text{subst}(n, m, k) \) which, if applied to the Gödel numbers of \( A \), \( t \), and \( x \), yields the Gödel number of \( A[t/x] \). In other words, \( \text{subst}(\#A, \#t, \#x) = \#A[t/x] \). Likewise, the property \( \text{Deriv}(n) \) that holds of \( n \) if it is the Gödel number of a correct derivation in natural deduction, is primitive recursive. Showing that these are primitive recursive required a fair amount of work, and at times some ingenuity, and depended essentially on the fact that operating with sequences is primitive recursive. If a theory \( T \) is decidable, then we can use \( \text{Deriv} \) to define a decidable relation \( \text{Prf}_T(n, m) \) which holds if \( n \) is the Gödel number of a derivation of the sentence with Gödel number \( m \) from \( T \). This relation is primitive recursive if the set of axioms
of $T$ is, and merely general recursive if the axioms of $T$ are decidable but not primitive recursive.

**Problems**

**Problem 3.1.** Show that the function $\text{flatten}(z)$, which turns the sequence $\langle t_1^#, \ldots, t_n^# \rangle$ into $t_1, \ldots, t_n^#$, is primitive recursive.

**Problem 3.2.** Give a detailed proof of Proposition 3.8 along the lines of the first proof of Proposition 3.5

**Problem 3.3.** Give a detailed proof of Proposition 3.8 along the lines of the alternate proof of Proposition 3.5

**Problem 3.4.** Prove Proposition 3.9. You may make use of the fact that any substring of a formula which is a formula is a subformula of it.

**Problem 3.5.** Prove Proposition 3.12

**Problem 3.6.** Define the following relations as in Proposition 3.15:

1. $\text{FollowsBy}_{\rightarrow}\text{Elim}(x, d_1, d_2, n)$,
2. $\text{FollowsBy}_{=}\text{Elim}(x, d_1, d_2, n)$,
3. $\text{FollowsBy}_{\lor}\text{Elim}(x, d_1, d_2, d_3, n)$,
4. $\text{FollowsBy}_{\forall}\text{Intro}(x, d_1, n)$.

For the last one, you will have to also show that you can test primitive recursively if the formula with Gödel number $x$ and the derivation with Gödel number $d$ satisfy the eigenvariable condition, i.e., the eigenvariable $a$ of the $\forall\text{Intro}$ inference occurs neither in $x$ nor in an open assumption of $d$. 
4.1 Introduction

We will describe a very minimal such theory called “\( \mathbf{Q} \)” (or, sometimes, “Robinson’s \( \mathbf{Q} \),” after Raphael Robinson). We will say what it means for a function to be representable in \( \mathbf{Q} \), and then we will prove the following:

A function is representable in \( \mathbf{Q} \) if and only if it is computable.

For one thing, this provides us with another model of computability. But we will also use it to show that the set \( \{ A : \mathbf{Q} \vdash A \} \) is not decidable, by reducing the halting problem to it. By the time we are done, we will have proved much stronger things than this.

The language of \( \mathbf{Q} \) is the language of arithmetic; \( \mathbf{Q} \) consists of the following axioms (to be used in conjunction with the other axioms and rules of first-order logic with identity predicate):

\[
\begin{align*}
\forall x \forall y \left( x' = y' \rightarrow x = y \right) & \quad (Q_1) \\
\forall x \ 0 \neq x' & \quad (Q_2) \\
\forall x \left( x \neq 0 \rightarrow \exists y \ x = y' \right) & \quad (Q_3)
\end{align*}
\]
\[ \forall x (x + 0) = x \]  
\[ \forall x \forall y (x + y') = (x + y)' \]  
\[ \forall x (x \times 0) = 0 \]  
\[ \forall x \forall y (x \times y') = ((x \times y) + x) \]  
\[ \forall x \forall y (x < y \leftrightarrow \exists z (z' + x) = y) \]

For each natural number \( n \), define the numeral \( \overline{n} \) to be the term \( 0''...'' \) where there are \( n \) tick marks in all. So, \( \overline{0} \) is the constant symbol \( \overline{0} \) by itself, \( \overline{1} \) is \( \overline{0}' \), \( \overline{2} \) is \( \overline{0}'' \), etc.

As a theory of arithmetic, \( \mathbb{Q} \) is extremely weak; for example, you can’t even prove very simple facts like \( \forall x x \neq x' \) or \( \forall x \forall y (x + y) = (y + x) \). But we will see that much of the reason that \( \mathbb{Q} \) is so interesting is because it is so weak. In fact, it is just barely strong enough for the incompleteness theorem to hold. Another reason \( \mathbb{Q} \) is interesting is because it has a finite set of axioms.

A stronger theory than \( \mathbb{Q} \) (called Peano arithmetic \( \mathbb{PA} \)) is obtained by adding a schema of induction to \( \mathbb{Q} \):

\[ (A(\overline{0}) \land \forall x (A(x) \rightarrow A(x'))) \rightarrow \forall x A(x) \]

where \( A(x) \) is any formula. If \( A(x) \) contains free variables other than \( x \), we add universal quantifiers to the front to bind all of them (so that the corresponding instance of the induction schema is a sentence). For instance, if \( A(x,y) \) also contains the variable \( y \) free, the corresponding instance is

\[ \forall y ((A(\overline{0}) \land \forall x (A(x) \rightarrow A(x'))) \rightarrow \forall x A(x)) \]

Using instances of the induction schema, one can prove much more from the axioms of \( \mathbb{PA} \) than from those of \( \mathbb{Q} \). In fact, it takes a good deal of work to find “natural” statements about the natural numbers that can’t be proved in Peano arithmetic!
4.1. INTRODUCTION

**Definition 4.1.** A function $f(x_0, \ldots, x_k)$ from the natural numbers to the natural numbers is said to be *representable in $\mathbb{Q}$* if there is a formula $A_f(x_0, \ldots, x_k, y)$ such that whenever $f(n_0, \ldots, n_k) = m$, $\mathbb{Q}$ proves

1. $A_f(n_0, \ldots, n_k, m)$
2. $\forall y (A_f(n_0, \ldots, n_k, y) \rightarrow m = y)$.

There are other ways of stating the definition; for example, we could equivalently require that $\mathbb{Q}$ proves $\forall y (A_f(n_0, \ldots, n_k, y) \leftrightarrow y = m)$.

**Theorem 4.2.** A function is representable in $\mathbb{Q}$ if and only if it is computable.

There are two directions to proving the theorem. The left-to-right direction is fairly straightforward once arithmetization of syntax is in place. The other direction requires more work. Here is the basic idea: we pick “general recursive” as a way of making “computable” precise, and show that every general recursive function is representable in $\mathbb{Q}$. Recall that a function is general recursive if it can be defined from zero, the successor function $\text{succ}$, and the projection functions $P^n_i$, using composition, primitive recursion, and regular minimization. So one way of showing that every general recursive function is representable in $\mathbb{Q}$ is to show that the basic functions are representable, and whenever some functions are representable, then so are the functions defined from them using composition, primitive recursion, and regular minimization. In other words, we might show that the basic functions are representable, and that the representable functions are “closed under” composition, primitive recursion, and regular minimization. This guarantees that every general recursive function is representable.

It turns out that the step where we would show that representable functions are closed under primitive recursion is hard.
In order to avoid this step, we show first that in fact we can do without primitive recursion. That is, we show that every general recursive function can be defined from basic functions using composition and regular minimization alone. To do this, we show that primitive recursion can actually be done by a specific regular minimization. However, for this to work, we have to add some additional basic functions: addition, multiplication, and the characteristic function of the identity relation \( \chi_\approx \). Then, we can prove the theorem by showing that all of these basic functions are representable in \( \mathbb{Q} \), and the representable functions are closed under composition and regular minimization.

### 4.2 Functions Representable in \( \mathbb{Q} \) are Computable

**Lemma 4.3.** Every function that is representable in \( \mathbb{Q} \) is computable.

**Proof.** Let’s first give the intuitive idea for why this is true. If \( f(x_0, \ldots, x_k) \) is representable in \( \mathbb{Q} \), there is a formula \( A(x_0, \ldots, x_k, y) \) such that

\[
\mathbb{Q} \vdash A_f(\overline{n_0}, \ldots, \overline{n_k}, m) \iff m = f(n_0, \ldots, n_k).
\]

To compute \( f \), we do the following. List all the possible derivations \( \delta \) in the language of arithmetic. This is possible to do mechanically. For each one, check if it is a derivation of a formula of the form \( A_f(\overline{n_0}, \ldots, \overline{n_k}, \overline{m}) \). If it is, \( m \) must be \( f(n_0, \ldots, n_k) \) and we’ve found the value of \( f \). The search terminates because \( \mathbb{Q} \vdash A_f(\overline{n_0}, \ldots, \overline{n_k}, f(n_0, \ldots, n_k)) \), so eventually we find a \( \delta \) of the right sort.

This is not quite precise because our procedure operates on derivations and formulas instead of just on numbers, and we haven’t explained exactly why “listing all possible derivations” is mechanically possible. But as we’ve seen, it is possible to code terms, formulas, and derivations by Gödel numbers. We’ve also
introduced a precise model of computation, the general recursive functions. And we’ve seen that the relation \( \Prf_Q(d, y) \), which holds iff \( d \) is the Gödel number of a derivation of the formula with Gödel number \( x \) from the axioms of \( Q \), is (primitive) recursive. Other primitive recursive functions we’ll need are \( \text{num} \) and \( \text{Subst} \). From these, it is possible to define \( f \) by minimization; thus, \( f \) is recursive.

First, define

\[
A(n_0, \ldots, n_k, m) = \\
\text{Subst}(\text{Subst}(\ldots \text{Subst}(\#A_f\#, \text{num}(n_0), \#x_0\#), \\
\ldots, \text{num}(n_k), \#x_k\#), \text{num}(m), \#y\#)
\]

This looks complicated, but it’s just the function \( A(n_0, \ldots, n_k, m) = \#A_f(\overline{n_0}, \ldots, \overline{n_k}, \overline{m})\).

Now, consider the relation \( R(n_0, \ldots, n_k, s) \) which holds if \( (s)_0 \) is the Gödel number of a derivation from \( Q \) of \( A_f(\overline{n_0}, \ldots, \overline{n_k}, (s)_1) \):

\[
R(n_0, \ldots, n_k, s) \iff \Prf_Q((s)_0, A(n_0, \ldots, n_k, (s)_1)
\]

If we can find an \( s \) such that \( R(n_0, \ldots, n_k, s) \) hold, we have found a pair of numbers—\( (s)_0 \) and \( (s)_1 \)—such that \( (s)_0 \) is the Gödel number of a derivation of \( A_f(\overline{n_0}, \ldots, \overline{n_k}, (s)_1) \). So looking for \( s \) is like looking for the pair \( d \) and \( m \) in the informal proof. And a computable function that “looks for” such an \( s \) can be defined by regular minimization. Note that \( R \) is regular: for every \( n_0, \ldots, n_k \), there is a derivation \( \delta \) of \( Q \vdash A_f(\overline{n_0}, \ldots, \overline{n_k}, f(n_0, \ldots, n_k)) \), so \( R(n_0, \ldots, n_k, s) \) holds for \( s = (\#\delta\#, f(n_0, \ldots, n_k)) \). So, we can write \( f \) as

\[
f(n_0, \ldots, n_k) = (\mu s \ R(n_0, \ldots, n_k, s))_1.
\]

\[\square\]

### 4.3 The Beta Function Lemma

In order to show that we can carry out primitive recursion if addition, multiplication, and \( \chi_\theta \) are available, we need to develop
functions that handle sequences. (If we had exponentiation as well, our task would be easier.) When we had primitive recursion, we could define things like the “\(n\)-th prime,” and pick a fairly straightforward coding. But here we do not have primitive recursion—in fact we want to show that we can do primitive recursion using minimization—so we need to be more clever.

**Lemma 4.4.** There is a function \(\beta(d, i)\) such that for every sequence \(a_0, \ldots, a_n\) there is a number \(d\), such that for every \(i \leq n\), \(\beta(d, i) = a_i\). Moreover, \(\beta\) can be defined from the basic functions using just composition and regular minimization.

Think of \(d\) as coding the sequence \(\langle a_0, \ldots, a_n \rangle\), and \(\beta(d, i)\) returning the \(i\)-th element. (Note that this “coding” does not use the power-of-primes coding we’re already familiar with!). The lemma is fairly minimal; it doesn’t say we can concatenate sequences or append elements, or even that we can compute \(d\) from \(a_0, \ldots, a_n\) using functions definable by composition and regular minimization. All it says is that there is a “decoding” function such that every sequence is “coded.”

The use of the notation \(\beta\) is Gödel’s. To repeat, the hard part of proving the lemma is defining a suitable \(\beta\) using the seemingly restricted resources, i.e., using just composition and minimization—however, we’re allowed to use addition, multiplication, and \(\chi\). There are various ways to prove this lemma, but one of the cleanest is still Gödel’s original method, which used a number-theoretic fact called the Chinese Remainder theorem.

**Definition 4.5.** Two natural numbers \(a\) and \(b\) are relatively prime if their greatest common divisor is 1; in other words, they have no other divisors in common.
Definition 4.6. \( a \equiv b \mod c \) means \( c \mid (a - b) \), i.e., \( a \) and \( b \) have the same remainder when divided by \( c \).

Here is the Chinese Remainder theorem:

**Theorem 4.7.** Suppose \( x_0, \ldots, x_n \) are (pairwise) relatively prime. Let \( y_0, \ldots, y_n \) be any numbers. Then there is a number \( z \) such that

\[
\begin{align*}
  z &\equiv y_0 \mod x_0 \\
  z &\equiv y_1 \mod x_1 \\
  \vdots \\
  z &\equiv y_n \mod x_n.
\end{align*}
\]

Here is how we will use the Chinese Remainder theorem: if \( x_0, \ldots, x_n \) are bigger than \( y_0, \ldots, y_n \) respectively, then we can take \( z \) to code the sequence \( \langle y_0, \ldots, y_n \rangle \). To recover \( y_i \), we need only divide \( z \) by \( x_i \) and take the remainder. To use this coding, we will need to find suitable values for \( x_0, \ldots, x_n \).

A couple of observations will help us in this regard. Given \( y_0, \ldots, y_n \), let

\[
  j = \max(n, y_0, \ldots, y_n) + 1,
\]

and let

\[
\begin{align*}
  x_0 &= 1 + j! \\
  x_1 &= 1 + 2 \cdot j! \\
  x_2 &= 1 + 3 \cdot j! \\
  \vdots \\
  x_n &= 1 + (n + 1) \cdot j!
\end{align*}
\]

Then two things are true:

1. \( x_0, \ldots, x_n \) are relatively prime.

2. For each \( i \), \( y_i < x_i \).
To see that (1) is true, note that if \( p \) is a prime number and \( p \mid x_i \) and \( p \mid x_k \), then \( p \mid 1 + (i + 1)j! \) and \( p \mid 1 + (k + 1)j! \). But then \( p \) divides their difference,

\[
(1 + (i + 1)j!) - (1 + (k + 1)j!) = (i - k)j!.
\]

Since \( p \) divides \( 1 + (i + 1)j! \), it can’t divide \( j! \) as well (otherwise, the first division would leave a remainder of 1). So \( p \) divides \( i - k \), since \( p \) divides \( (i - k)j! \). But \( |i - k| \) is at most \( n \), and we have chosen \( j > n \), so this implies that \( p \mid j! \), again a contradiction. So there is no prime number dividing both \( x_i \) and \( x_k \). Clause (2) is easy: we have \( y_i < j < j! < x_i \).

Now let us prove the \( \beta \) function lemma. Remember that we can use 0, successor, plus, times, \( \chi = \), projections, and any function defined from them using composition and minimization applied to regular functions. We can also use a relation if its characteristic function is so definable. As before we can show that these relations are closed under boolean combinations and bounded quantification; for example:

1. \( \operatorname{not}(x) = \chi = (x, 0) \)

2. \( (\min x \leq z) R(x, y) = \mu x (R(x, y) \lor x = z) \)

3. \( (\exists x \leq z) R(x, y) \iff R((\min x \leq z) R(x, y), y) \)

We can then show that all of the following are also definable without primitive recursion:

1. The pairing function, \( J(x, y) = \frac{1}{2}[(x + y)(x + y + 1)] + x \)

2. Projections

\[
K(z) = (\min x \leq q) (\exists y \leq z [z = J(x, y)])
\]

and

\[
L(z) = (\min y \leq q) (\exists x \leq z [z = J(x, y)]).
\]

3. \( x < y \)
4. $x \mid y$

5. The function $\text{rem}(x, y)$ which returns the remainder when $y$ is divided by $x$

Now define

$$\beta^*(d_0, d_1, i) = \text{rem}(1 + (i + 1)d_1, d_0)$$

and

$$\beta(d, i) = \beta^*(K(d), L(d), i).$$

This is the function we need. Given $a_0, \ldots, a_n$, as above, let

$$j = \max(n, a_0, \ldots, a_n) + 1,$$

and let $d_1 = j!$. By the observations above, we know that $1 + d_1, 1 + 2d_1, \ldots, 1 + (n + 1)d_1$ are relatively prime and all are bigger than $a_0, \ldots, a_n$. By the Chinese Remainder theorem there is a value $d_0$ such that for each $i$,

$$d_0 \equiv a_i \mod (1 + (i + 1)d_1)$$

and so (because $d_1$ is greater than $a_i$),

$$a_i = \text{rem}(1 + (i + 1)d_1, d_0).$$

Let $d = J(d_0, d_1)$. Then for each $i \leq n$, we have

$$\beta(d, i) = \beta^*(d_0, d_1, i) = \text{rem}(1 + (i + 1)d_1, d_0) = a_i$$

which is what we need. This completes the proof of the $\beta$-function lemma.
4.4 Simulating Primitive Recursion

Now we can show that definition by primitive recursion can be “simulated” by regular minimization using the beta function. Suppose we have $f(z)$ and $g(u,v,z)$. Then the function $h(x,z)$ defined from $f$ and $g$ by primitive recursion is

$$
\begin{align*}
    h(0,z) &= f(z) \\
    h(x+1,z) &= g(x,h(x,z),z).
\end{align*}
$$

We need to show that $h$ can be defined from $f$ and $g$ using just composition and regular minimization, using the basic functions and functions defined from them using composition and regular minimization (such as $\beta$).

**Lemma 4.8.** If $h$ can be defined from $f$ and $g$ using primitive recursion, it can be defined from $f$, $g$, the functions zero, succ, $P^n$, add, mult, $\chi=,$ using composition and regular minimization.

**Proof.** First, define an auxiliary function $\hat{h}(x,z)$ which returns the least number $d$ such that $d$ codes a sequence which satisfies

1. $(d)_0 = f(z)$, and
2. for each $i < x$, $(d)_{i+1} = g(i,(d)_i,z)$,

where now $(d)_i$ is short for $\beta(d,i)$. In other words, $\hat{h}$ returns the sequence $\langle h(0,z), h(1,z), \ldots, h(x,z) \rangle$. We can write $\hat{h}$ as

$$
\hat{h}(x,z) = \mu d \ (\beta(d,0) = f(z) \land \forall i < x \beta(d,i+1) = g(i,\beta(d,i),z)).
$$

Note: no primitive recursion is needed here, just minimization. The function we minimize is regular because of the beta function lemma Lemma 4.4.

But now we have

$$
h(x,z) = \beta(\hat{h}(x,z),x),
$$

so $h$ can be defined from the basic functions using just composition and regular minimization. □
4.5 Basic Functions are Representable in \( \mathcal{Q} \)

First we have to show that all the basic functions are representable in \( \mathcal{Q} \). In the end, we need to show how to assign to each \( k \)-ary basic function \( f(x_0, \ldots, x_{k-1}) \) a formula \( A_f(x_0, \ldots, x_{k-1}, y) \) that represents it.

We will be able to represent zero, successor, plus, times, the characteristic function for equality, and projections. In each case, the appropriate representing function is entirely straightforward; for example, zero is represented by the formula \( y = 0 \), successor is represented by the formula \( x' = y \), and addition is represented by the formula \( (x_0 + x_1) = y \). The work involves showing that \( \mathcal{Q} \) can prove the relevant sentences; for example, saying that addition is represented by the formula above involves showing that for every pair of natural numbers \( m \) and \( n \), \( \mathcal{Q} \) proves

\[
\overline{n + m} = \overline{n + m} \text{ and } \\
\forall y ((\overline{n + m}) = y \rightarrow y = \overline{n + m}).
\]

**Proposition 4.9.** The zero function \( \text{zero}(x) = 0 \) is represented in \( \mathcal{Q} \) by \( y = 0 \).

**Proposition 4.10.** The successor function \( \text{succ}(x) = x + 1 \) is represented in \( \mathcal{Q} \) by \( y = x' \).

**Proposition 4.11.** The projection function \( P^n_i(x_0, \ldots, x_{n-1}) = x_i \) is represented in \( \mathcal{Q} \) by \( y = x_i \).

**Proposition 4.12.** The characteristic function of \( = \),

\[
\chi_{=}(x_0, x_1) = \begin{cases} 
1 & \text{if } x_0 = x_1 \\
0 & \text{otherwise}
\end{cases}
\]
is represented in $Q$ by

$$(x_0 = x_1 \land y = \overline{1}) \lor (x_0 \neq x_1 \land y = \overline{0}).$$

The proof requires the following lemma.

**Lemma 4.13.** Given natural numbers $n$ and $m$, if $n \neq m$, then $Q \vdash \overline{n} \neq \overline{m}$.

**Proof.** Use induction on $n$ to show that for every $m$, if $n \neq m$, then $Q \vdash \overline{n} \neq \overline{m}$.

In the base case, $n = 0$. If $m$ is not equal to 0, then $m = k + 1$ for some natural number $k$. We have an axiom that says $\forall x \ 0 \neq x'$. By a quantifier axiom, replacing $x$ by $k$, we can conclude $0 \neq \overline{k}'$. But $\overline{k}'$ is just $\overline{m}$.

In the induction step, we can assume the claim is true for $n$, and consider $n + 1$. Let $m$ be any natural number. There are two possibilities: either $m = 0$ or for some $k$ we have $m = k + 1$. The first case is handled as above. In the second case, suppose $n + 1 \neq k + 1$. Then $n \neq k$. By the induction hypothesis for $n$ we have $Q \vdash \overline{n} \neq \overline{k}$. We have an axiom that says $\forall x \forall y \ x' = y' \rightarrow x = y$. Using a quantifier axiom, we have $\overline{n}' = \overline{k}' \rightarrow \overline{n} = \overline{k}$. Using propositional logic, we can conclude, in $Q$, $\overline{n} \neq \overline{k} \rightarrow \overline{n}' \neq \overline{k}'$. Using modus ponens, we can conclude $\overline{n}' \neq \overline{k}'$, which is what we want, since $\overline{k}'$ is $\overline{m}$. □

Note that the lemma does not say much: in essence it says that $Q$ can prove that different numerals denote different objects. For example, $Q$ proves $0'' \neq 0'''$. But showing that this holds in general requires some care. Note also that although we are using induction, it is induction outside of $Q$.

**Proof of Proposition 4.12.** If $n = m$, then $\overline{n}$ and $\overline{m}$ are the same term, and $\chi_{=}(n, m) = 1$. But $Q \vdash (\overline{n} = \overline{m} \land \overline{1} = \overline{1})$, so it proves $A_{=}(\overline{n}, \overline{m}, \overline{1})$. If $n \neq m$, then $\chi_{=}(n, m) = 0$. By Lemma 4.13, $Q \vdash \overline{n} \neq \overline{m}$ and so also $(\overline{n} \neq \overline{m} \land 0 = 0)$. Thus $Q \vdash A_{=}(\overline{n}, \overline{m}, 0)$. 
4.5. BASIC FUNCTIONS ARE REPRESENTABLE IN $Q$

For the second part, we also have two cases. If $n = m$, we have to show that that $Q \vdash \forall(A_=(\bar{n}, \bar{m}, y) \rightarrow y = \bar{1})$. Arguing informally, suppose $A_=(\bar{n}, \bar{m}, y)$, i.e.,

$$(\bar{n} = \bar{n} \land y = \bar{1}) \lor (\bar{n} \neq \bar{n} \land y = \bar{0})$$

The left disjunct implies $y = \bar{1}$ by logic; the right contradicts $\bar{n} = \bar{n}$ which is provable by logic.

Suppose, on the other hand, that $n \neq m$. Then $A_=(\bar{n}, \bar{m}, y)$ is

$$(\bar{n} = \bar{m} \land y = \bar{1}) \lor (\bar{n} \neq \bar{m} \land y = \bar{0})$$

Here, the left disjunct contradicts $\bar{n} \neq \bar{m}$, which is provable in $Q$ by Lemma 4.13; the right disjunct entails $y = \bar{0}$. □

**Proposition 4.14.** The addition function $\text{add}(x_0, x_1) = x_0 + x_1$ is is represented in $Q$ by

$$y = (x_0 + x_1).$$

**Lemma 4.15.** $Q \vdash (\bar{n} + \bar{m}) = \bar{n + m}$

*Proof.* We prove this by induction on $m$. If $m = 0$, the claim is that $Q \vdash (\bar{n} + 0) = \bar{n}$. This follows by axiom $Q_4$. Now suppose the claim for $m$; let’s prove the claim for $m + 1$, i.e., prove that $Q \vdash (\bar{n} + m + 1) = n + m + 1$. Note that $m + 1$ is just $\bar{m}'$, and $\bar{n + m + 1}$ is just $\bar{n + m}'$. By axiom $Q_5$, $Q \vdash (\bar{n} + \bar{m}') = (\bar{n + m}')$. By induction hypothesis, $Q \vdash (\bar{n} + \bar{m}) = n + \bar{m}$. So $Q \vdash (\bar{n} + \bar{m}') = n + \bar{m}'$. □

*Proof of Proposition 4.14.* The formula $A_{\text{add}}(x_0, x_1, y)$ representing add is $y = (x_0 + x_1)$. First we show that if $\text{add}(n, m) = k$, then $Q \vdash A_{\text{add}}(\bar{n}, \bar{m}, k)$, i.e., $Q \vdash k = (\bar{n} + \bar{m})$. But since $k = n + m$, $\bar{k}$ just is $\bar{n + m}$, and we’ve shown in Lemma 4.15 that $Q \vdash (\bar{n} + \bar{m}) = \bar{n + m}$.

We also have to show that if $\text{add}(n, m) = k$, then

$$Q \vdash \forall y (A_{\text{add}}(\bar{n}, \bar{m}, y) \rightarrow y = \bar{k}).$$
Suppose we have $\bar{n} + \bar{m} = y$. Since
\[
\mathbb{Q} \vdash (\bar{n} + \bar{m}) = \bar{n} + \bar{m},
\]
we can replace the left side with $\bar{n} + \bar{m}$ and get $\bar{n} + \bar{m} = y$, for arbitrary $y$.

\begin{proposition}
The multiplication function $\text{mult}(x_0, x_1) = x_0 \cdot x_1$ is represented in $\mathbb{Q}$ by
\[
y = (x_0 \times x_1).
\]
\end{proposition}

\begin{proof}
Exercise.
\end{proof}

\begin{lemma}
$\mathbb{Q} \vdash (\bar{n} \times \bar{m}) = \bar{n} \cdot \bar{m}$
\end{lemma}

\begin{proof}
Exercise.
\end{proof}

\section{Composition is Representable in $\mathbb{Q}$}

Suppose $h$ is defined by
\[
h(x_0, \ldots, x_{l-1}) = f(g_0(x_0, \ldots, x_{l-1}), \ldots, g_{k-1}(x_0, \ldots, x_{l-1})).
\]
where we have already found formulas $A_f, A_{g_0}, \ldots, A_{g_{k-1}}$ representing the functions $f$, and $g_0, \ldots, g_{k-1}$, respectively. We have to find a formula $A_h$ representing $h$.

Let’s start with a simple case, where all functions are 1-place, i.e., consider $h(x) = f(g(x))$. If $A_f(y, z)$ represents $f$, and $A_g(x, y)$ represents $g$, we need a formula $A_h(x, z)$ that represents $h$. Note that $h(x) = z$ iff there is a $y$ such that both $z = f(y)$ and $y = g(x)$. (If $h(x) = z$, then $g(x)$ is such a $y$; if such a $y$ exists, then since $y = g(x)$ and $z = f(y)$, $z = f(g(x))$.) This suggests that $\exists y (A_g(x, y) \land A_f(y, z))$ is a good candidate for $A_h(x, z)$. We just have to verify that $\mathbb{Q}$ proves the relevant formulas.
Proposition 4.18. If \( h(n) = m \), then \( Q \vdash A_h(\overline{n}, \overline{m}) \).

Proof. Suppose \( h(n) = m \), i.e., \( f(g(n)) = m \). Let \( k = g(n) \). Then

\[
Q \vdash A_g(\overline{n}, \overline{k})
\]
since \( A_g \) represents \( g \), and

\[
Q \vdash A_f(\overline{k}, \overline{m})
\]
since \( A_f \) represents \( f \). Thus,

\[
Q \vdash A_g(\overline{n}, \overline{k}) \land A_f(\overline{k}, \overline{m})
\]
and consequently also

\[
Q \vdash \exists y (A_g(\overline{n}, y) \land A_f(y, \overline{m}))
\]
i.e., \( Q \vdash A_h(n, m) \). \( \square \)

Proposition 4.19. If \( h(n) = m \), then \( Q \vdash \forall z (A_h(\overline{n}, z) \rightarrow z = \overline{m}) \).

Proof. Suppose \( h(n) = m \), i.e., \( f(g(n)) = m \). Let \( k = g(n) \). Then

\[
Q \vdash \forall y (A_g(\overline{n}, y) \rightarrow y = \overline{k})
\]
since \( A_g \) represents \( g \), and

\[
Q \vdash \forall z (A_f(\overline{k}, z) \rightarrow z = \overline{m})
\]
since \( A_f \) represents \( f \). Using just a little bit of logic, we can show that also

\[
Q \vdash \forall z (\exists y (A_g(\overline{n}, y) \land A_f(y, z)) \rightarrow z = \overline{m})
\]
i.e., \( Q \vdash \forall y (A_h(\overline{n}, y) \rightarrow y = \overline{m}) \). \( \square \)

The same idea works in the more complex case where \( f \) and \( g \) have arity greater than 1.
Proposition 4.20. If \( A_f(y_0, \ldots, y_{k-1}, z) \) represents \( f(y_0, \ldots, y_{k-1}) \) in \( Q \), and \( A_g_i(x_0, \ldots, x_{l-1}, y) \) represents \( g_i(x_0, \ldots, x_{l-1}) \) in \( Q \), then

\[
\exists y_0, \ldots \exists y_{k-1} (A_{g_0}(x_0, \ldots, x_{l-1}, y_0) \land \cdots \land A_{g_{k-1}}(x_0, \ldots, x_{l-1}, y_{k-1}) \land A_f(y_0, \ldots, y_{k-1}, z))
\]

represents

\[
h(x_0, \ldots, x_{k-1}) = f(g_0(x_0, \ldots, x_{k-1}), \ldots, g_0(x_0, \ldots, x_{k-1})).
\]

Proof. Exercise. □

4.7 Regular Minimization is Representable in \( Q \)

Let’s consider unbounded search. Suppose \( g(x, z) \) is regular and representable in \( Q \), say by the formula \( A_g(x, z, y) \). Let \( f \) be defined by \( f(z) = \mu x \ [g(x, z) = 0] \). We would like to find a formula \( A_f(z, y) \) representing \( f \). The value of \( f(z) \) is that number \( x \) which (a) satisfies \( g(x, z) = 0 \) and (b) is the least such, i.e., for any \( w < x, g(w, z) \neq 0 \). So the following is a natural choice:

\[
A_f(z, y) \equiv A_g(y, z, 0) \land \forall w (w < y \rightarrow \neg A_g(w, z, 0)).
\]

In the general case, of course, we would have to replace \( z \) with \( z_0, \ldots, z_k \).

The proof, again, will involve some lemmas about things \( Q \) is strong enough to prove.
Lemma 4.21. For every variable $x$ and every natural number $n$,

$$Q \vdash (x' + \overline{n}) = (x + \overline{n})'.$$

Proof. The proof is, as usual, by induction on $n$. In the base case, $n = 0$, we need to show that $Q$ proves $(x' + 0) = (x + 0)'.$ But we have:

$$Q \vdash (x' + 0) = x' \quad \text{by axiom } Q_4 \tag{4.1}$$
$$Q \vdash (x + 0) = x \quad \text{by axiom } Q_4 \tag{4.2}$$
$$Q \vdash (x + 0)' = x' \quad \text{by eq. (4.2)} \tag{4.3}$$
$$Q \vdash (x' + 0) = (x + 0)' \quad \text{by eq. (4.1) and eq. (4.3)}$$

In the induction step, we can assume that we have shown that $Q \vdash (x' + \overline{n}) = (x + \overline{n})'$. Since $\overline{n + 1}$ is $\overline{n}'$, we need to show that $Q$ proves $(x' + \overline{n}') = (x + \overline{n}')'$. We have:

$$Q \vdash (x' + \overline{n}') = (x' + \overline{n})' \quad \text{by axiom } Q_5 \tag{4.4}$$
$$Q \vdash (x' + \overline{n}') = (x + \overline{n}')' \quad \text{inductive hypothesis} \tag{4.5}$$
$$Q \vdash (x' + \overline{n})' = (x + \overline{n}')' \quad \text{by eq. (4.4) and eq. (4.5)} \tag{4.6}.$$

\[\square\]

It is again worth mentioning that this is weaker than saying that $Q$ proves $\forall x \forall y (x' + y) = (x + y)'$. Although this sentence is true in $N$, $Q$ does not prove it.

Lemma 4.22.

1. $Q \vdash \forall x \neg x < 0.$

2. For every natural number $n$,

$$Q \vdash \forall x (x < \overline{n + 1} \rightarrow (x = 0 \lor \cdots \lor x = \overline{n})).$$

Proof. Let us do 1 and part of 2, informally (i.e., only giving hints as to how to construct the formal derivation).

For part 1, by the definition of $<$, we need to prove $\neg \exists y (y' + x) = 0$ in $Q$, which is equivalent (using the axioms and rules of
first-order logic) to $\forall y (y' + x) \neq 0$. Here is the idea: suppose $(y' + x) = 0$. If $x = 0$, we have $(y' + 0) = 0$. But by axiom $Q_4$ of $Q$, we have $(y' + 0) = y'$, and by axiom $Q_2$ we have $y' \neq 0$, a contradiction. So $\forall y (y' + x) \neq 0$. If $x \neq 0$, by axiom $Q_3$, there is a $z$ such that $x = z'$. But then we have $(y' + z') = 0$. By axiom $Q_5$, we have $(y' + z)' = 0$, again contradicting axiom $Q_2$.

For part 2, use induction on $n$. Let us consider the base case, when $n = 0$. In that case, we need to show $x < 1 \rightarrow x = 0$. Suppose $x < 1$. Then by the defining axiom for $<$, we have $\exists y (y' + x) = o'$. Suppose $y$ has that property; so we have $y' + x = o'$.

We need to show $x = 0$. By axiom $Q_3$, if $x \neq 0$, we get $x = z'$ for some $z$. Then we have $(y' + z') = o'$. By axiom $Q_5$ of $Q$, we have $(y' + z)' = o'$. By axiom $Q_1$, we have $(y' + z) = 0$. But this means, by definition, $z < 0$, contradicting part 1. □

**Lemma 4.23.** For every $m \in \mathbb{N}$,

$$Q \vdash \forall y ((y < \overline{m} \lor \overline{m} < y) \lor y = \overline{m}).$$

**Proof.** By induction on $m$. First, consider the case $m = 0$. $Q \vdash \forall y (y \neq 0 \rightarrow \exists z y = z')$ by $Q_3$. But if $y = z'$, then $(z' + 0) = (y + 0)$ by the logic of $\Rightarrow$. By $Q_4$, $(y + 0) = y$, so we have $(z' + 0) = y$, and hence $\exists z (z' + 0) = y$. By the definition of $<$ in $Q_8$, $0 < y$. If $0 < y$, then also $0 < y \lor y < 0$. We obtain: $y \neq 0 \rightarrow (0 < y \lor y < 0)$, which is equivalent to $(0 < y \lor y < 0) \lor y = 0$.

Now suppose we have

$$Q \vdash \forall y ((y < \overline{m} \lor \overline{m} < y) \lor y = \overline{m})$$

and we want to show

$$Q \vdash \forall y ((y < \overline{m + 1} \lor \overline{m + 1} < y) \lor y = \overline{m + 1})$$

The first disjunct $y < \overline{m}$ is equivalent (by $Q_8$) to $\exists z (z' + y) = \overline{m}$. If $(z' + y) = \overline{m}$, then also $(z' + y)' = \overline{m'}$. By $Q_4$, $(z' + y)' = (z'' + y)$. Hence, $(z'' + y) = \overline{m'}$. We get $\exists u (u' + y) = m + 1$ by existentially
generalizing on \( z' \) and keeping in mind that \( m' \) is \( m + 1 \). Hence, if \( y < m \) then \( y < m + 1 \).

Now suppose \( m < y \), i.e., \( \exists z (z' + m) = y \). By \( Q_3 \) and some logic, we have \( z = 0 \lor \exists u \ z = u' \). If \( z = 0 \), we have \( (o' + m) = y \). Since \( Q_4 \vdash (o' + m) = m + 1 \), we have \( y = m + 1 \). Now suppose \( \exists u \ z = u' \). Then:

\[
\begin{align*}
y &= (z' + m) & \text{by assumption} \\
(z' + m) &= (u'' + m)' & \text{from } z = u' \\
(u'' + m) &= (u' + m)' & \text{by Lemma 4.21} \\
(u' + m)' &= (u' + m) & \text{by } Q_5, \text{ so} \\
y &= (u' + m + 1)
\end{align*}
\]

By existential generalization, \( \exists u (u' + m + 1) = y \), i.e., \( m + 1 < y \). So, if \( m < y \), then \( m + 1 < y \lor y = m + 1 \).

Finally, assume \( y = m \). Then, since \( Q_4 \vdash (o' + m) = m + 1 \), \( (o' + y) = m + 1 \). From this we get \( \exists z (z' + y) = m + 1 \), or \( y < m + 1 \).

Hence, from each disjunct of the case for \( m \), we can obtain the case for \( m + 1 \). \( \square \)

**Proposition 4.24.** If \( A_g(x, z, y) \) represents \( g(x, y) \) in \( Q \), then

\[ A_f(z, y) \equiv A_g(y, z, o) \land \forall w (w < y \rightarrow \neg A_g(w, z, o)). \]

represents \( f(z) = \mu x [g(x, z) = 0] \).

**Proof.** First we show that if \( f(n) = m \), then \( Q_4 \vdash A_f(n, m) \), i.e.,

\[ Q_4 \vdash A_g(m, n, o) \land \forall w (w < m \rightarrow \neg A_g(w, n, o)). \]

Since \( A_g(x, z, y) \) represents \( g(x, z) \) and \( g(m, n) = 0 \) if \( f(n) = m \), we have

\[ Q_4 \vdash A_g(m, n, o). \]
If \( f(n) = m \), then for every \( k < m \), \( g(k, n) \neq 0 \). So

\[ \mathbb{Q} \vdash \neg A_g(k, n) \, 0. \]

We get that

\[ \mathbb{Q} \vdash \forall w (w < m \rightarrow \neg A_g(w, n, 0)). \quad (4.6) \]

by Lemma 4.22 (by (1) in case \( m = 0 \) and by (2) otherwise).

Now let’s show that if \( f(n) = m \), then \( \mathbb{Q} \vdash \forall y (A_f(n, y) \rightarrow y = m) \). We again sketch the argument informally, leaving the formalization to the reader.

Suppose \( A_f(n, y) \). From this we get (a) \( A_g(y, n, 0) \) and (b) \( \forall w (w < y \rightarrow \neg A_g(w, n, 0)) \). By Lemma 4.23, \( (y < m \vee m < y) \vee y = m \). We’ll show that both \( y < m \) and \( m < y \) leads to a contradiction.

If \( m < y \), then \( \neg A_g(m, n, 0) \) from (b). But \( m = f(n) \), so \( g(m, n) = 0 \), and so \( \mathbb{Q} \vdash A_g(m, n, 0) \) since \( A_g \) represents \( g \). So we have a contradiction.

Now suppose \( y < m \). Then since \( \mathbb{Q} \vdash \forall w (w < m \rightarrow \neg A_g(w, n, 0)) \) by eq. (4.6), we get \( \neg A_g(y, n, 0) \). This again contradicts (a). \( \square \)

### 4.8 Computable Functions are Representable in \( \mathbb{Q} \)

**Theorem 4.25.** Every computable function is representable in \( \mathbb{Q} \).

**Proof.** For definiteness, and using the Church-Turing Thesis, let’s say that a function is computable iff it is general recursive. The general recursive functions are those which can be defined from the zero function zero, the successor function succ, and the projection function \( P^n_i \) using composition, primitive recursion, and regular minimization. By Lemma 4.8, any function \( h \) that can be defined from \( f \) and \( g \) can also be defined using composition and regular minimization from \( f \), \( g \), and zero, succ, \( P^n_i \), add, mult,
χ=. Consequently, a function is general recursive iff it can be defined from zero, succ, \( P^n \), add, mult, \( \chi \) using composition and regular minimization.

We’ve furthermore shown that the basic functions in question are representable in \( \mathbb{Q} \) (Propositions 4.9 to 4.12, 4.14 and 4.16), and that any function defined from representable functions by composition or regular minimization (Proposition 4.20, Proposition 4.24) is also representable. Thus every general recursive function is representable in \( \mathbb{Q} \). □

We have shown that the set of computable functions can be characterized as the set of functions representable in \( \mathbb{Q} \). In fact, the proof is more general. From the definition of representability, it is not hard to see that any theory extending \( \mathbb{Q} \) (or in which one can interpret \( \mathbb{Q} \)) can represent the computable functions. But, conversely, in any proof system in which the notion of proof is computable, every representable function is computable. So, for example, the set of computable functions can be characterized as the set of functions representable in Peano arithmetic, or even Zermelo-Fraenkel set theory. As Gödel noted, this is somewhat surprising. We will see that when it comes to provability, questions are very sensitive to which theory you consider; roughly, the stronger the axioms, the more you can prove. But across a wide range of axiomatic theories, the representable functions are exactly the computable ones; stronger theories do not represent more functions as long as they are axiomatizable.

## 4.9 Representing Relations

Let us say what it means for a relation to be representable.
**Definition 4.26.** A relation $R(x_0, \ldots, x_k)$ on the natural numbers is *representable in $\mathbb{Q}$* if there is a formula $A_R(x_0, \ldots, x_k)$ such that whenever $R(n_0, \ldots, n_k)$ is true, $\mathbb{Q}$ proves $A_R(\overline{n_0}, \ldots, \overline{n_k})$, and whenever $R(n_0, \ldots, n_k)$ is false, $\mathbb{Q}$ proves $\neg A_R(\overline{n_0}, \ldots, \overline{n_k})$.

**Theorem 4.27.** A relation is representable in $\mathbb{Q}$ if and only if it is computable.

**Proof.** For the forwards direction, suppose $R(x_0, \ldots, x_k)$ is represented by the formula $A_R(x_0, \ldots, x_k)$. Here is an algorithm for computing $R$: on input $n_0, \ldots, n_k$, simultaneously search for a proof of $A_R(\overline{n_0}, \ldots, \overline{n_k})$ and a proof of $\neg A_R(\overline{n_0}, \ldots, \overline{n_k})$. By our hypothesis, the search is bound to find one or the other; if it is the first, report “yes,” and otherwise, report “no.”

In the other direction, suppose $R(x_0, \ldots, x_k)$ is computable. By definition, this means that the function $\chi_R(x_0, \ldots, x_k)$ is computable. By Theorem 4.2, $\chi_R$ is represented by a formula, say $A_{\chi_R}(x_0, \ldots, x_k, y)$. Let $A_R(x_0, \ldots, x_k)$ be the formula $A_{\chi_R}(x_0, \ldots, x_k, \overline{1})$. Then for any $n_0, \ldots, n_k$, if $R(n_0, \ldots, n_k)$ is true, then $\chi_R(n_0, \ldots, n_k) = 1$, in which case $\mathbb{Q}$ proves $A_{\chi_R}(\overline{n_0}, \ldots, \overline{n_k}, \overline{1})$, and so $\mathbb{Q}$ proves $A_R(\overline{n_0}, \ldots, \overline{n_k})$. On the other hand, if $R(n_0, \ldots, n_k)$ is false, then $\chi_R(n_0, \ldots, n_k) = 0$. This means that $\mathbb{Q}$ proves

$$\forall y (A_{\chi_R}(\overline{n_0}, \ldots, \overline{n_k}, y) \rightarrow y = \overline{0}).$$

Since $\mathbb{Q}$ proves $\overline{0} \neq \overline{1}$, $\mathbb{Q}$ proves $\neg A_{\chi_R}(\overline{n_0}, \ldots, \overline{n_k}, \overline{1})$, and so it proves $\neg A_R(\overline{n_0}, \ldots, \overline{n_k})$. □

### 4.10 Undecidability

We call a theory $T$ *undecidable* if there is no computational procedure which, after finitely many steps and unfailingly, provides a correct answer to the question “does $T$ prove $A$?” for any sentence $A$ in the language of $T$. So $\mathbb{Q}$ would be decidable iff
there were a computational procedure which decides, given a sentence $A$ in the language of arithmetic, whether $Q \vdash A$ or not. We can make this more precise by asking: Is the relation $\text{Prov}_Q(y)$, which holds of $y$ iff $y$ is the Gödel number of a sentence provable in $Q$, recursive? The answer is: no.

**Theorem 4.28.** $Q$ is undecidable, i.e., the relation

\[ \text{Prov}_Q(y) \Leftrightarrow \text{Sent}(y) \land \exists x \, \text{Prf}_Q(x, y) \]

is not recursive.

**Proof.** Suppose it were. Then we could solve the halting problem as follows: Given $e$ and $n$, we know that $\varphi_e(n) \downarrow$ iff there is an $s$ such that $T(e, n, s)$, where $T$ is Kleene’s predicate from Theorem 2.8. Since $T$ is primitive recursive it is representable in $Q$ by a formula $B_T$, that is, $Q \vdash B_T(\bar{e}, \bar{n}, \bar{s})$ iff $T(e, n, s)$. If $Q \vdash B_T(\bar{e}, \bar{n}, \bar{s})$ then also $Q \vdash \exists y \, B_T(\bar{e}, \bar{n}, y)$. If no such $s$ exists, then $Q \vdash \neg B_T(\bar{e}, \bar{n}, \bar{s})$ for every $s$. But $Q$ is $\omega$-consistent, i.e., if $Q \vdash \neg A(\bar{n})$ for every $n \in \mathbb{N}$, then $Q \not\vdash \exists y \, A(y)$. We know this because the axioms of $Q$ are true in the standard model $\mathbb{N}$. So, $Q \not\vdash \exists y \, B_T(\bar{e}, \bar{n}, y)$. In other words, $Q \vdash \exists y \, B_T(\bar{e}, \bar{n}, y)$ iff there is an $s$ such that $T(e, n, s)$, i.e., iff $\varphi_e(n) \downarrow$. From $e$ and $n$ we can compute \"\exists y \, B_T(\bar{e}, \bar{n}, y)\", let $g(e, n)$ be the primitive recursive function which does that. So

\[ h(e, n) = \begin{cases} 1 & \text{if } \text{Prov}_Q(g(e, n)) \\ 0 & \text{otherwise}. \end{cases} \]

This would show that $h$ is recursive if $\text{Prov}_Q$ is. But $h$ is not recursive, by Theorem 2.9, so $\text{Prov}_Q$ cannot be either. \qed

**Corollary 4.29.** First-order logic is undecidable.

**Proof.** If first-order logic were decidable, provability in $Q$ would be as well, since $Q \vdash A$ iff $\vdash O \rightarrow A$, where $O$ is the conjunction of the axioms of $Q$. \qed
Summary

In order to show how theories like $\mathcal{Q}$ can “talk” about computable functions—and especially about provability (via Gödel numbers)—we established that $\mathcal{Q}$ represents all computable functions. By $\mathcal{Q}$ represents $f(n)$" we mean that there is a formula $A_f(x,y)$ in $\mathcal{L}_A$ which expresses that $f(x) = y$, and $\mathcal{Q}$ can prove that it does. This means that whenever $f(n) = m$, then $\mathcal{T} \vdash A_f(\overline{n}, \overline{m})$ and $\mathcal{T} \vdash \forall y (A_f(numn, y) \rightarrow y = \overline{m})$. (Here, $\overline{n}$ is the standard numeral for $n$, i.e., the term $0'\ldots'$ with $n$ $0$s. The term $\overline{n}$ picks out $n$ in the standard model $\mathbb{N}$, so it’s a convenient way of representing the number $n$ in $\mathcal{L}_A$.) To prove that $\mathcal{Q}$ represents all computable functions we go back to the characterization of computable functions as those that can be defined from zero, succ, and the projections, by composition, primitive recursion, and regular minimization. While it is relatively easy to prove that the basic functions are representable and that functions defined by composition and regular minimization from representable functions are also representable, primitive recursion is harder. To avoid this problem, we showed that we can actually avoid definition by primitive recursion, if we allow a few additional basic functions (namely, addition, multiplication, and the characteristic function of $=\)$. This required a beta function which allows us to deal with sequences of numbers in a rudimentary way, and which can be defined without using primitive recursion.

Problems

Problem 4.1. Prove that $y = 0$, $y = x'$, and $y = x_i$ represent zero, succ, and $P^n_i$, respectively.

Problem 4.2. Prove Lemma 4.17.

Problem 4.3. Use Lemma 4.17 to prove Proposition 4.16.
Problem 4.4. Using the proofs of Proposition 4.19 and Proposition 4.19 as a guide, carry out the proof of Proposition 4.20 in detail.

Problem 4.5. Show that if $R$ is representable in $\mathbb{Q}$, so is $\chi_R$. 
CHAPTER 5

Incompleteness and Provability

5.1 Introduction

Hilbert thought that a system of axioms for a mathematical structure, such as the natural numbers, is inadequate unless it allows one to derive all true statements about the structure. Combined with his later interest in formal systems of deduction, this suggests that he thought that we should guarantee that, say, the formal systems we are using to reason about the natural numbers is not only consistent, but also complete, i.e., every statement in its language is either provable or its negation is. Gödel’s first incompleteness theorem shows that no such system of axioms exists: there is no complete, consistent, axiomatizable formal system for arithmetic. In fact, no “sufficiently strong,” consistent, axiomatizable mathematical theory is complete.

A more important goal of Hilbert’s, the centerpiece of his program for the justification of modern (“classical”) mathemat-
ics, was to find finitary consistency proofs for formal systems representing classical reasoning. With regard to Hilbert’s program, then, Gödel’s second incompleteness theorem was a much bigger blow. The second incompleteness theorem can be stated in vague terms, like the first incompleteness theorem. Roughly speaking, it says that no sufficiently strong theory of arithmetic can prove its own consistency. We will have to take “sufficiently strong” to include a little bit more than $\mathbf{Q}$.

The idea behind Gödel’s original proof of the incompleteness theorem can be found in the Epimenides paradox. Epimenides, a Cretan, asserted that all Cretans are liars; a more direct form of the paradox is the assertion “this sentence is false.” Essentially, by replacing truth with provability, Gödel was able to formalize a sentence which, in a roundabout way, asserts that it itself is not provable. If that sentence were provable, the theory would then be inconsistent. Assuming $\omega$-consistency—a property stronger than consistency—Gödel was able to show that this sentence is also not refutable from the system of axioms he was considering.

The first challenge is to understand how one can construct a sentence that refers to itself. For every formula $A$ in the language of $\mathbf{Q}$, let $\langle A \rangle$ denote the numeral corresponding to $\#A\#$. Think about what this means: $A$ is a formula in the language of $\mathbf{Q}$, $\#A\#$ is a natural number, and $\langle A \rangle$ is a term in the language of $\mathbf{Q}$. So every formula $A$ in the language of $\mathbf{Q}$ has a name, $\langle A \rangle$, which is a term in the language of $\mathbf{Q}$; this provides us with a conceptual framework in which formulas in the language of $\mathbf{Q}$ can “say” things about other formulas. The following lemma is known as the fixed-point lemma.
Lemma 5.1. Let $T$ be any theory extending $Q$, and let $B(x)$ be any formula with only the variable $x$ free. Then there is a sentence $A$ such that $T \vdash A \leftrightarrow B(⌜A⌝)$.

The lemma asserts that given any property $B(x)$, there is a sentence $A$ that asserts “$B(x)$ is true of me.”

How can we construct such a sentence? Consider the following version of the Epimenides paradox, due to Quine:

“Yields falsehood when preceded by its quotation”
yields falsehood when preceded by its quotation.

This sentence is not directly self-referential. It simply makes an assertion about the syntactic objects between quotes, and, in doing so, it is on par with sentences like

1. “Robert” is a nice name.
2. “I ran.” is a short sentence.
3. “Has three words” has three words.

But what happens when one takes the phrase “yields falsehood when preceded by its quotation,” and precedes it with a quoted version of itself? Then one has the original sentence! In short, the sentence asserts that it is false.

5.2 The Fixed-Point Lemma

The fixed-point lemma says that for any formula $B(x)$, there is a sentence $A$ such that $T \vdash A \leftrightarrow B(⌜A⌝)$, provided $T$ extends $Q$. In the case of the liar sentence, we’d want $A$ to be equivalent (provably in $T$) to “$⌜A⌝$ is false,” i.e., the statement that $⌜A⌝$ is the Gödel number of a false sentence. To understand the idea of the proof, it will be useful to compare it with Quine’s informal gloss of $A$ as, “‘yields a falsehood when preceded by its own
5.2. THE FIXED-POINT LEMMA

quotation’ yields a falsehood when preceded by its own quotation.” The operation of taking an expression, and then forming a sentence by preceding this expression by its own quotation may be called diagonalizing the expression, and the result its diagonalization. So, the diagonalization of ‘yields a falsehood when preceded by its own quotation’ is “‘yields a falsehood when preceded by its own quotation’ yields a falsehood when preceded by its own quotation.” Now note that Quine’s liar sentence is not the diagonalization of ‘yields a falsehood’ but of ‘yields a falsehood when preceded by its own quotation.’ So the property being diagonalized to yield the liar sentence itself involves diagonalization!

In the language of arithmetic, we form quotations of a formula with one free variable by computing its Gödel numbers and then substituting the standard numeral for that Gödel number into the free variable. The diagonalization of $E(x)$ is $E(\bar{n})$, where $n = \#E(x)^\#$. (From now on, let’s abbreviate $\#E(x)^\#$ as $\bar{\gamma}E(x)$.) So if $B(x)$ is “is a falsehood,” then “yields a falsehood if preceded by its own quotation,” would be “yields a falsehood when applied to the Gödel number of its diagonalization.” If we had a symbol $\text{diag}$ for the function $\text{diag}(n)$ which computes the Gödel number of the diagonalization of the formula with Gödel number $n$, we could write $E(x)$ as $B(\text{diag}(x))$. And Quine’s version of the liar sentence would then be the diagonalization of it, i.e., $E(\bar{\gamma}E)$ or $B(\text{diag}(\bar{\gamma}B(\text{diag}(x))))$. Of course, $B(x)$ could now be any other property, and the same construction would work. For the incompleteness theorem, we’ll take $B(x)$ to be “$x$ is unprovable in $T$.” Then $E(x)$ would be “yields a sentence unprovable in $T$ when applied to the Gödel number of its diagonalization.”

To formalize this in $T$, we have to find a way to formalize diag. The function $\text{diag}(n)$ is computable, in fact, it is primitive recursive: if $n$ is the Gödel number of a formula $E(x)$, $\text{diag}(n)$ returns the Gödel number of $E(\bar{\gamma}E(x)^\gamma)$. (Recall, $\bar{\gamma}E(x)^\gamma$ is the standard numeral of the Gödel number of $E(x)$, i.e., $\#E(x)^\#$). If $\text{diag}$ were a function symbol in $T$ representing the function diag, we could
take $A$ to be the formula $B(diag(\lceil B(diag(x) \rceil))$. Notice that
\[
\text{diag}(\# B(diag(x))\#) = \# B(diag(\lceil B(diag(x) \rceil))\# = \# A\#.
\]

Assuming $T$ can prove
\[
diag(\lceil B(diag(x) \rceil) = \lceil A \rceil,
\]
it can prove $B(diag(\lceil B(diag(x) \rceil)) \leftrightarrow B(\lceil A \rceil)$. But the left hand side is, by definition, $A$.

Of course, $diag$ will in general not be a function symbol of $T$, and certainly is not one of $Q$. But, since $diag$ is computable, it is representable in $Q$ by some formula $D_{diag}(x,y)$. So instead of writing $B(diag(x))$ we can write $\exists y (D_{diag}(x,y) \land B(y))$. Otherwise, the proof sketched above goes through, and in fact, it goes through already in $Q$.

**Lemma 5.2.** Let $B(x)$ be any formula with one free variable $x$. Then there is a sentence $A$ such that $Q \vdash A \leftrightarrow B(\lceil A \rceil)$.

**Proof.** Given $B(x)$, let $E(x)$ be the formula $\exists y (D_{diag}(x,y) \land B(y))$ and let $A$ be its diagonalization, i.e., the formula $E(\lceil E(x) \rceil)$.

Since $D_{diag}$ represents $diag$, and $\text{diag}(\# E(x)\#) = \# A\#$, $Q$ can prove
\[
D_{diag}(\lceil E(x) \rceil, \lceil A \rceil) \quad (5.1) \\
\forall y (D_{diag}(\lceil E(x) \rceil, y) \rightarrow y = \lceil A \rceil). \quad (5.2)
\]

Now we show that $Q \vdash A \leftrightarrow B(\lceil A \rceil)$. We argue informally, using just logic and facts provable in $Q$.

First, suppose $A$, i.e., $E(\lceil E(x) \rceil)$. Going back to the definition of $E(x)$, we see that $E(\lceil E(x) \rceil)$ just is
\[
\exists y (D_{diag}(\lceil E(x) \rceil, y) \land B(y)).
\]

Consider such a $y$. Since $D_{diag}(\lceil E(x) \rceil, y)$, by eq. (5.2), $y = \lceil A \rceil$. So, from $B(y)$ we have $B(\lceil A \rceil)$. 
Now suppose $B(\vdash A)$. By eq. (5.1), we have $D_{\text{diag}}(\vdash E(x) \vdash, \vdash A) \land B(\vdash A)$. It follows that $\exists y (D_{\text{diag}}(\vdash E(x) \vdash, y) \land B(y))$. But that’s just $E(\vdash E)$, i.e., $A$. □

You should compare this to the proof of the fixed-point lemma in computability theory. The difference is that here we want to define a statement in terms of itself, whereas there we wanted to define a function in terms of itself; this difference aside, it is really the same idea.

### 5.3 The First Incompleteness Theorem

We can now describe Gödel’s original proof of the first incompleteness theorem. Let $T$ be any computably axiomatized theory in a language extending the language of arithmetic, such that $T$ includes the axioms of $Q$. This means that, in particular, $T$ represents computable functions and relations.

We have argued that, given a reasonable coding of formulas and proofs as numbers, the relation $\text{Prf}_T(x, y)$ is computable, where $\text{Prf}_T(x, y)$ holds if and only if $x$ is the Gödel number of a derivation of the formula with Gödel number $y$ in $T$. In fact, for the particular theory that Gödel had in mind, Gödel was able to show that this relation is primitive recursive, using the list of 45 functions and relations in his paper. The 45th relation, $xBy$, is just $\text{Prf}_T(x, y)$ for his particular choice of $T$. Remember that where Gödel uses the word “recursive” in his paper, we would now use the phrase “primitive recursive.”

Since $\text{Prf}_T(x, y)$ is computable, it is representable in $T$. We will use $\text{Prf}_T(x, y)$ to refer to the formula that represents it. Let $\text{Prov}_T(y)$ be the formula $\exists x \text{Prf}_T(x, y)$. This describes the 46th relation, $\text{Bew}(y)$, on Gödel’s list. As Gödel notes, this is the only relation that “cannot be asserted to be recursive.” What he probably meant is this: from the definition, it is not clear that it is computable; and later developments, in fact, show that it isn’t.
**Definition 5.3.** A theory \( T \) is \( \omega \)-consistent if the following holds: if \( \exists x A(x) \) is any sentence and \( T \) proves \( \neg A(0), \neg A(1), \neg A(2), \ldots \) then \( T \) does not prove \( \exists x A(x) \).

We can now prove the following.

**Theorem 5.4.** Let \( T \) be any \( \omega \)-consistent, axiomatizable theory extending \( Q \). Then \( T \) is not complete.

**Proof.** Let \( T \) be an axiomatizable theory containing \( Q \). Then \( \text{Prf}_T(x, y) \) is decidable, hence representable in \( Q \) by a formula \( \text{Prf}_T(x, y) \). Let \( \text{Prov}_T(y) \) be the formula we described above. By the fixed-point lemma, there is a formula \( G_T \) such that \( Q \) (and hence \( T \)) proves

\[
G_T \leftrightarrow \neg \text{Prov}_T(\text{⌜}G_T\text{⌝})).
\]

(5.3)

Note that \( A \) says, in essence, “\( A \) is not provable.”

We claim that

1. If \( T \) is consistent, \( T \) doesn’t prove \( G_T \)

2. If \( T \) is \( \omega \)-consistent, \( T \) doesn’t prove \( \neg G_T \).

This means that if \( T \) is \( \omega \)-consistent, it is incomplete, since it proves neither \( G_T \) nor \( \neg G_T \). Let us take each claim in turn.

Suppose \( T \) proves \( G_T \). Then there is a derivation, and so, for some number \( m \), the relation \( \text{Prf}_T(m, \text{⌜}G_T\text{⌝}) \) holds. But then \( Q \) proves the sentence \( \text{Prf}_T(m, \text{⌜}G_T\text{⌝}) \). So \( Q \) proves \( \exists x \text{Prf}_T(x, \text{⌜}G_T\text{⌝}) \), which is, by definition, \( \text{Prov}_T(\text{⌜}G_T\text{⌝}) \). By eq. (5.3), \( Q \) proves \( \neg G_T \), and since \( T \) extends \( Q \), so does \( T \). We have shown that if \( T \) proves \( G_T \), then it also proves \( \neg G_T \), and hence it would be inconsistent.

For the second claim, let us show that if \( T \) proves \( \neg G_T \), then it is \( \omega \)-inconsistent. Suppose \( T \) proves \( \neg G_T \). If \( T \) is inconsistent, it is \( \omega \)-inconsistent, and we are done. Otherwise, \( T \) is consistent,
5.4 Rosser’s Theorem

Can we modify Gödel’s proof to get a stronger result, replacing “ω-consistent” with simply “consistent”? The answer is “yes,” using a trick discovered by Rosser. Rosser’s trick is to use a “modified” provability predicate \( \text{RProv}_T(y) \) instead of \( \text{Prov}_T(y) \).

**Theorem 5.5.** Let \( T \) be any consistent, axiomatizable theory extending \( Q \). Then \( T \) is not complete.

**Proof.** Recall that \( \text{Prov}_T(y) \) is defined as \( \exists x \, \text{Prf}_T(x, y) \), where \( \text{Prf}_T(x, y) \) represents the decidable relation which holds iff \( x \) is the Gödel number of a derivation of the sentence with Gödel number \( y \). The relation that holds between \( x \) and \( y \) if \( x \) is the Gödel number of a refutation of the sentence with Gödel number \( y \) is also decidable. Let \( \text{not}(x) \) be the primitive recursive function which does the following: if \( x \) is the code of a formula \( A \), \( \text{not}(x) \) is a code of \( \neg A \). Then \( \text{Ref}_T(x, y) \) holds iff \( \text{Prf}_T(x, \text{not}(y)) \). Let \( \text{Ref}_T(x, y) \) represent it. Then, if \( T \vdash \neg A \) and \( \delta \) is a corresponding derivation, \( Q \vdash \text{Ref}_T(\neg \delta, \neg A) \). We define \( \text{RProv}_T(y) \) as

\[
\exists x \, (\text{Prf}_T(x, y) \land \forall z \, (z < x \rightarrow \neg \text{Ref}_T(z, y))).
\]

Roughly, \( \text{RProv}_T(y) \) says “there is a proof of \( y \) in \( T \), and there is no shorter refutation of \( y \).” (You might find it convenient to read \( \text{RProv}_T(y) \) as “\( y \) is shmovable.”) Assuming \( T \) is consistent, \( \text{RProv}_T(y) \) is true of the same numbers as \( \text{Prov}_T(y) \); but from the point of view of provability in \( T \) (and we now know that there is a
difference between truth and provability!) the two have different properties. (If $T$ is inconsistent, then the two do not hold of the same numbers!)

By the fixed-point lemma, there is a formula $R_T$ such that

$$Q \vdash R_T \leftrightarrow \neg \text{Prov}_T(\overline{R_T}).$$

In contrast to the proof of Theorem 5.4, here we claim that if $T$ is consistent, $T$ doesn’t prove $R_T$, and $T$ also doesn’t prove $\neg R_T$. (In other words, we don’t need the assumption of $\omega$-consistency.)

First, let’s show that $T \nvdash R_T$. Suppose it did, so there is a derivation of $R_T$ from $T$; let $n$ be its Gödel number. Then $Q \vdash \text{Prf}_T(n, \overline{R_T})$, since $\text{Prf}_T$ represents $\text{Prf}_T$ in $Q$. Also, for each $k < n$, $k$ is not the Gödel number of $\neg R_T$, since $T$ is consistent.

So for each $k < n$, $Q \vdash \neg \text{Ref}_T(k, \overline{R_T})$. By Lemma 4.22(2), $Q \vdash \forall z (z < n \to \neg \text{Ref}_T(z, \overline{R_T}))$. Thus,

$$Q \vdash \exists x (\text{Prf}_T(x, \overline{R_T}) \land \forall z (z < x \to \neg \text{Ref}_T(z, \overline{R_T}))),$$

but that’s just $\text{Prov}_T(\overline{R_T})$. By eq. (5.4), $Q \vdash \neg R_T$. Since $T$ extends $Q$, also $T \vdash \neg R_T$. We’ve assumed that $T \vdash R_T$, so $T$ would be inconsistent, contrary to the assumption of the theorem.

Now, let’s show that $T \nvdash \neg R_T$. Again, suppose it did, and suppose $n$ is the Gödel number of a derivation of $\neg R_T$. Then $\text{Ref}_T(n, \overline{\neg R_T})$ holds, and since $\text{Ref}_T$ represents $\text{Ref}_T$ in $Q$, $Q \vdash \text{Ref}_T(n, \overline{\neg R_T})$. We’ll again show that $T$ would then be inconsistent because it would also prove $R_T$. Since $Q \vdash R_T \leftrightarrow \neg \text{Prov}_T(\overline{R_T})$, and since $T$ extends $Q$, it suffices to show that $Q \vdash \neg \text{Prov}_T(\overline{R_T})$.

The sentence $\neg \text{Prov}_T(\overline{R_T})$, i.e.,

$$\neg \exists x (\text{Prf}_T(x, \overline{R_T}) \land \forall z (z < x \to \neg \text{Ref}_T(z, \overline{R_T})))$$

is logically equivalent to

$$\forall x (\text{Prf}_T(x, \overline{R_T}) \to \exists z (z < x \land \text{Ref}_T(z, \overline{R_T})))$$

We argue informally using logic, making use of facts about what $Q$ proves. Suppose $x$ is arbitrary and $\text{Prf}_T(x, \overline{R_T})$. We already
know that $T \not\models R_{\bar{T}}$, and so for every $k$, $Q \vdash \neg \text{Prf}_T(\bar{k}, \neg R_{\bar{T}})$. Thus, for every $k$ it follows that $x \neq \bar{n}$. We also have $\neg(x = 0 \lor x = 1 \lor \cdots \lor x = n - 1)$ and so by Lemma 4.22(2), (b) $\neg(x < \bar{n})$. By Lemma 4.23, $\bar{n} < x$. Since $Q \vdash \text{Ref}_T(\bar{n}, \neg R_{\bar{T}})$, we have $\bar{n} < x \land \text{Ref}_T(\bar{n}, \neg R_{\bar{T}})$, and from that $\exists z (z < x \land \text{Ref}_T(z, \neg R_{\bar{T}}))$. Since $x$ was arbitrary we get

$$\forall x (\text{Prf}_T(x, \neg R_{\bar{T}}) \rightarrow \exists z (z < x \land \text{Ref}_T(z, \neg R_{\bar{T}})))$$

as required. □

5.5 Comparison with Gödel’s Original Paper

It is worthwhile to spend some time with Gödel’s 1931 paper. The introduction sketches the ideas we have just discussed. Even if you just skim through the paper, it is easy to see what is going on at each stage: first Gödel describes the formal system $P$ (syntax, axioms, proof rules); then he defines the primitive recursive functions and relations; then he shows that $xBx$ is primitive recursive, and argues that the primitive recursive functions and relations are represented in $P$. He then goes on to prove the incompleteness theorem, as above. In section 3, he shows that one can take the unprovable assertion to be a sentence in the language of arithmetic. This is the origin of the $\beta$-lemma, which is what we also used to handle sequences in showing that the recursive functions are representable in $Q$. Gödel doesn’t go so far to isolate a minimal set of axioms that suffice, but we now know that $Q$ will do the trick. Finally, in Section 4, he sketches a proof of the second incompleteness theorem.

5.6 The Provability Conditions for PA

Peano arithmetic, or $\text{PA}$, is the theory extending $Q$ with induction axioms for all formulas. In other words, one adds to $Q$
axioms of the form

\[(A(0) \land \forall x (A(x) \rightarrow A(x'))) \rightarrow \forall x A(x)\]

for every formula \(A\). Notice that this is really a schema, which is to say, infinitely many axioms (and it turns out that \(\text{PA}\) is not finitely axiomatizable). But since one can effectively determine whether or not a string of symbols is an instance of an induction axiom, the set of axioms for \(\text{PA}\) is computable. \(\text{PA}\) is a much more robust theory than \(\text{Q}\). For example, one can easily prove that addition and multiplication are commutative, using induction in the usual way. In fact, most finitary number-theoretic and combinatorial arguments can be carried out in \(\text{PA}\).

Since \(\text{PA}\) is computably axiomatized, the provability predicate \(\text{Prf}_{\text{PA}}(x, y)\) is computable and hence represented in \(\text{Q}\) (and so, in \(\text{PA}\)). As before, I will take \(\text{Prf}_{\text{PA}}(x, y)\) to denote the formula representing the relation. Let \(\text{Prov}_{\text{PA}}(y)\) be the formula \(\exists x \text{Prf}_{\text{PA}}(x, y)\), which, intuitively says, “\(y\) is provable from the axioms of \(\text{PA}\).” The reason we need a little bit more than the axioms of \(\text{Q}\) is we need to know that the theory we are using is strong enough to prove a few basic facts about this provability predicate. In fact, what we need are the following facts:

P1. If \(\text{PA} \vdash A\), then \(\text{PA} \vdash \text{Prov}_{\text{PA}}(\ulcorner A \urcorner)\)

P2. For all formulas \(A\) and \(B\),

\[\text{PA} \vdash \text{Prov}_{\text{PA}}(\ulcorner A \rightarrow B \urcorner) \rightarrow (\text{Prov}_{\text{PA}}(\ulcorner A \urcorner) \rightarrow \text{Prov}_{\text{PA}}(\ulcorner B \urcorner))\]

P3. For every formula \(A\),

\[\text{PA} \vdash \text{Prov}_{\text{PA}}(\ulcorner A \urcorner) \rightarrow \text{Prov}_{\text{PA}}(\ulcorner \text{Prov}_{\text{PA}}(\ulcorner A \urcorner) \urcorner)\).

The only way to verify that these three properties hold is to describe the formula \(\text{Prov}_{\text{PA}}(y)\) carefully and use the axioms of \(\text{PA}\) to describe the relevant formal proofs. Conditions (1) and (2) are easy; it is really condition (3) that requires work. (Think about
what kind of work it entails...) Carrying out the details would be tedious and uninteresting, so here we will ask you to take it on faith that PA has the three properties listed above. A reasonable choice of $\text{Prov}_{PA}(y)$ will also satisfy

$$P_4. \text{ If } PA \vdash \text{Prov}_{PA}(\neg A), \text{ then } PA \vdash A.$$ 

But we will not need this fact.

Incidentally, Gödel was lazy in the same way we are being now. At the end of the 1931 paper, he sketches the proof of the second incompleteness theorem, and promises the details in a later paper. He never got around to it; since everyone who understood the argument believed that it could be carried out (he did not need to fill in the details.)

### 5.7 The Second Incompleteness Theorem

How can we express the assertion that PA doesn’t prove its own consistency? Saying PA is inconsistent amounts to saying that PA proves $0 = 1$. So we can take $\text{Con}_{PA}$ to be the formula $\neg \text{Prov}_{PA}(\neg 0 = 1)$, and then the following theorem does the job:

**Theorem 5.6.** Assuming PA is consistent, then PA does not prove $\text{Con}_{PA}$.

It is important to note that the theorem depends on the particular representation of $\text{Con}_{PA}$ (i.e., the particular representation of $\text{Prov}_{PA}(y)$). All we will use is that the representation of $\text{Prov}_{PA}(y)$ has the three properties above, so the theorem generalizes to any theory with a provability predicate having these properties.

It is informative to read Gödel’s sketch of an argument, since the theorem follows like a good punch line. It goes like this. Let $G_{PA}$ be the Gödel sentence that we constructed in the proof of Theorem 5.4. We have shown “If PA is consistent, then PA does not prove $G_{PA}$.” If we formalize this in PA, we have a proof of

$$\text{Con}_{PA} \rightarrow \neg \text{Prov}_{PA}(\neg G_{PA}).$$
Now suppose \( \text{PA} \) proves \( \text{Con}_{\text{PA}} \). Then it proves \( \neg \text{Prov}_{\text{PA}}(\neg \text{G}_{\text{PA}}) \). But since \( \text{G}_{\text{PA}} \) is a Gödel sentence, this is equivalent to \( \text{G}_{\text{PA}} \). So \( \text{PA} \) proves \( \text{G}_{\text{PA}} \).

But: we know that if \( \text{PA} \) is consistent, it doesn’t prove \( \text{G}_{\text{PA}} \)! So if \( \text{PA} \) is consistent, it can’t prove \( \text{Con}_{\text{PA}} \).

To make the argument more precise, we will let \( \text{G}_{\text{PA}} \) be the Gödel sentence for \( \text{PA} \) and use the provability conditions (1)–(3) above to show that \( \text{PA} \) proves \( \text{Con}_{\text{PA}} \to \text{G}_{\text{PA}} \). This will show that \( \text{PA} \) doesn’t prove \( \text{Con}_{\text{PA}} \). Here is a sketch of the proof, in \( \text{PA} \).

(For simplicity, we drop the \( \text{PA} \) subscripts.)

\[
G \leftrightarrow \neg \text{Prov}(\neg G) \tag{5.5}
\]

\( G \) is a Gödel sentence

\[
G \to \neg \text{Prov}(\neg G) \tag{5.6}
\]

from eq. (5.5)

\[
G \to (\text{Prov}(\neg G) \to \bot) \tag{5.7}
\]

from eq. (5.6) by logic

\[
\text{Prov}(G \to (\text{Prov}(\neg G) \to \bot)) \tag{5.8}
\]

by from eq. (5.7) by condition P1

\[
\text{Prov}(\neg G) \to \text{Prov}(G \to (\text{Prov}(\neg G) \to \bot)) \tag{5.9}
\]

from eq. (5.8) by condition P2

\[
\text{Prov}(\neg G) \to (\text{Prov}(\text{Prov}(\neg G) \to \bot) \to \text{Prov}(\bot)) \tag{5.10}
\]

from eq. (5.9) by condition P2 and logic

\[
\text{Prov}(\neg G) \to \text{Prov}(\text{Prov}(\neg G)) \tag{5.11}
\]

by P3

\[
\text{Prov}(\neg G) \to \text{Prov}(\bot) \tag{5.12}
\]

from eq. (5.10) and eq. (5.11) by logic

\[
\text{Con} \to \neg \text{Prov}(\neg G) \tag{5.13}
\]

contraposition of eq. (5.12) and \( \text{Con} \equiv \neg \text{Prov}(\bot) \)

\[
\text{Con} \to G \tag{5.13}
\]

from eq. (5.5) and eq. (5.13) by logic
The use of logic in the above just elementary facts from propositional logic, e.g., eq. (5.7) uses $\vdash \neg A \leftrightarrow (A \to \bot)$ and eq. (5.12) uses $A \to (B \to C), A \to B \vdash A \to C$. The use of condition $P_2$ in eq. (5.9) and eq. (5.10) relies on instances of $P_2$, $\text{Prov}(\Gamma A \to B') \rightarrow (\text{Prov}(\Gamma A') \rightarrow \text{Prov}(\Gamma B'))$. In the first one, $A \equiv G$ and $B \equiv \text{Prov}(\Gamma G') \rightarrow \bot$; in the second, $A \equiv \text{Prov}(\Gamma G')$ and $B \equiv \bot$.

The more abstract version of the incompleteness theorem is as follows:

**Theorem 5.7.** Let $T$ be any axiomatized theory extending $Q$, and let $\text{Prov}_T(y)$ be any formula satisfying provability conditions $P_1$–$P_3$ for $T$. Then if $T$ is consistent, then $T$ does not prove $\text{Con}_T$.

The moral of the story is that no “reasonable” consistent theory for mathematics can prove its own consistency. Suppose $T$ is a theory of mathematics that includes $Q$ and Hilbert’s “finitary” reasoning (whatever that may be). Then, the whole of $T$ cannot prove the consistency of $T$, and so, a fortiori, the finitary fragment can’t prove the consistency of $T$ either. In that sense, there cannot be a finitary consistency proof for “all of mathematics.”

There is some leeway in interpreting the term “finitary,” and Gödel, in the 1931 paper, grants the possibility that something we may consider “finitary” may lie outside the kinds of mathematics Hilbert wanted to formalize. But Gödel was being charitable; today, it is hard to see how we might find something that can reasonably be called finitary but is not formalizable in, say, $\text{ZFC}$.

### 5.8 Löb’s Theorem

The Gödel sentence for a theory $T$ is a fixed point of $\neg \text{Prov}_T(x)$, i.e., a sentence $G$ such that

$$T \vdash \neg \text{Prov}_T(\Gamma G) \leftrightarrow G.$$

It is not provable, because if $T \vdash G$, (a) by provability condition (1), $T \vdash \text{Prov}_T(\Gamma G')$, and (b) $T \vdash G$ together with $T \vdash$
\[ \neg \text{Prov}_T(\neg G) \iff G \text{ gives } T \vdash \neg \text{Prov}_T(\neg G), \] and so \( T \) would be inconsistent. Now it is natural to ask about the status of a fixed point of \( \text{Prov}_T(x) \), i.e., a sentence \( H \) such that

\[ T \vdash \text{Prov}_T(\neg H) \iff H. \]

If it were provable, \( T \vdash \text{Prov}_T(\neg H) \) by condition (1), but the same conclusion follows if we apply modus ponens to the equivalence above. Hence, we don’t get that \( T \) is inconsistent, at least not by the same argument as in the case of the Gödel sentence. This of course does not show that \( T \) does prove \( H \).

We can make headway on this question if we generalize it a bit. The left-to-right direction of the fixed point equivalence, \( \text{Prov}_T(\neg H) \to H \), is an instance of a general schema called a reflection principle: \( \text{Prov}_T(\neg A) \to A \). It is called that because it expresses, in a sense, that \( T \) can “reflect” about what it can prove; basically it says, “If \( T \) can prove \( A \), then \( A \) is true,” for any \( A \). This is true for sound theories only, of course, and this suggests that theories will in general not prove every instance of it. So which instances can a theory (strong enough, and satisfying the provability conditions) prove? Certainly all those where \( A \) itself is provable. And that’s it, as the next result shows.

**Theorem 5.8.** Let \( T \) be an axiomatizable theory extending \( Q \), and suppose \( \text{Prov}_T(y) \) is a formula satisfying conditions P1–P3 from section 5.7. If \( T \) proves \( \text{Prov}_T(\neg A) \to A \), then in fact \( T \) proves \( A \).

Put differently, if \( T \nvdash A \), then \( T \nvdash \text{Prov}_T(\neg A) \to A \). This result is known as Löb’s theorem.

The heuristic for the proof of Löb’s theorem is a clever proof that Santa Claus exists. (If you don’t like that conclusion, you are free to substitute any other conclusion you would like.) Here it is:

1. Let \( X \) be the sentence, “If \( X \) is true, then Santa Claus exists.”
2. Suppose $X$ is true.

3. Then what it says holds; i.e., we have: if $X$ is true, then Santa Claus exists.

4. Since we are assuming $X$ is true, we can conclude that Santa Claus exists, by modus ponens from (2) and (3).

5. We have succeeded in deriving (4), “Santa Claus exists,” from the assumption (2), “$X$ is true.” By conditional proof, we have shown: “If $X$ is true, then Santa Claus exists.”

6. But this is just the sentence $X$. So we have shown that $X$ is true.

7. But then, by the argument (2)–(4) above, Santa Claus exists.

A formalization of this idea, replacing “is true” with “is provable,” and “Santa Claus exists” with $A$, yields the proof of Löb’s theorem. The trick is to apply the fixed-point lemma to the formula $\text{Prov}_T(y) \rightarrow A$. The fixed point of that corresponds to the sentence $X$ in the preceding sketch.

**Proof.** Suppose $A$ is a sentence such that $T$ proves $\text{Prov}_T(\neg A) \rightarrow A$. Let $B(y)$ be the formula $\text{Prov}_T(y) \rightarrow A$, and use the fixed-point lemma to find a sentence $D$ such that $T$ proves $D \leftrightarrow B(\neg D)$. Then each of the following is provable in $T$:

\begin{align*}
D & \leftrightarrow (\text{Prov}_T(\neg D) \rightarrow A) & (5.14) \\
D & \rightarrow \text{Prov}_T(\neg D) & (5.15) \\
& \text{from eq. (5.14)} \\
\text{Prov}_T(\neg D \rightarrow (\text{Prov}_T(\neg D) \rightarrow A)) & \text{from eq. (5.15) by condition P1} \\
& \text{Prov}_T(\neg D) & \text{from eq. (5.16) using condition P2} \\
& \text{Prov}_T(\neg A) & \text{from eq. (5.17) using condition P2} \\
& A & \text{from condition P2}
\end{align*}
\[ \text{Prov}_T(\neg \bot) \rightarrow (\text{Prov}_T(\neg \text{Prov}_T(\neg \bot)) \rightarrow \text{Prov}_T(\neg A)) \]  
(5.18)

from eq. (5.17) using P2 again

\[ \text{Prov}_T(\neg \bot) \rightarrow \text{Prov}_T(\neg \text{Prov}_T(\neg \bot)) \]  
(5.19)

by provability condition P3

\[ \text{Prov}_T(\neg \bot) \rightarrow \text{Prov}_T(\neg A) \]  
(5.20)

from eq. (5.18) and eq. (5.19)

\[ \text{Prov}_T(\neg A) \rightarrow A \]  
(5.21)

by assumption of the theorem

\[ \text{Prov}_T(\neg \bot) \rightarrow A \]  
(5.22)

from eq. (5.20) and eq. (5.21)

\[ (\text{Prov}_T(\neg \bot) \rightarrow A) \rightarrow D \]  
(5.23)

from eq. (5.14)

\[ D \]  
(5.24)

from eq. (5.22) and eq. (5.23)

\[ \text{Prov}_T(\neg \bot) \]  
(5.25)

from eq. (5.24) by condition P1

\[ A \]  
from eq. (5.21) and eq. (5.25)

\[ \square \]

With Löb’s theorem in hand, there is a short proof of the first incompleteness theorem (for theories having a provability predicate satisfying conditions P1–P3: if \( T \vdash \text{Prov}_T(\neg \bot) \rightarrow \bot \), then \( T \vdash \bot \). If \( T \) is consistent, \( T \not\vdash \bot \). So, \( T \not\vdash \text{Prov}_T(\neg \bot) \rightarrow \bot \), i.e., \( T \not\vdash \text{Con}_T \). We can also apply it to show that \( H \), the fixed point of \( \text{Prov}_T(x) \), is provable. For since

\[ T \vdash \text{Prov}_T(\neg H) \leftrightarrow H \]

in particular

\[ T \vdash \text{Prov}_T(\neg H) \rightarrow H \]

and so by Löb’s theorem, \( T \vdash H \).
5.9 The Undefinability of Truth

The notion of definability depends on having a formal semantics for the language of arithmetic. We have described a set of formulas and sentences in the language of arithmetic. The “intended interpretation” is to read such sentences as making assertions about the natural numbers, and such an assertion can be true or false. Let \( N \) be the structure with domain \( \mathbb{N} \) and the standard interpretation for the symbols in the language of arithmetic. Then \( N \models A \) means “\( A \) is true in the standard interpretation.”

**Definition 5.9.** A relation \( R(x_1, \ldots, x_k) \) of natural numbers is **definable** in \( N \) if and only if there is a formula \( A(x_1, \ldots, x_k) \) in the language of arithmetic such that for every \( n_1, \ldots, n_k \), \( R(n_1, \ldots, n_k) \) if and only if \( N \models A(\bar{n}_1, \ldots, \bar{n}_k) \).

Put differently, a relation is definable in \( N \) if and only if it is representable in the theory \( \mathsf{TA} \), where \( \mathsf{TA} = \{ A : N \models A \} \) is the set of true sentences of arithmetic. (If this is not immediately clear to you, you should go back and check the definitions and convince yourself that this is the case.)

**Lemma 5.10.** Every computable relation is definable in \( N \).

**Proof.** It is easy to check that the formula representing a relation in \( \mathbb{Q} \) defines the same relation in \( N \). □

Now one can ask, is the converse also true? That is, is every relation definable in \( N \) computable? The answer is no. For example:
Lemma 5.11. The halting relation is definable in $\mathbb{N}$.

Proof. Let $H$ be the halting relation, i.e.,

$$H = \{\langle e, x \rangle : \exists s T(e, x, s)\}.$$ 

Let $D_T$ define $T$ in $\mathbb{N}$. Then

$$H = \{\langle e, x \rangle : \mathbb{N} \models \exists s D_T(e, x, s)\},$$

so $\exists s D_T(z, x, s)$ defines $H$ in $\mathbb{N}$. □

What about $\text{TA}$ itself? Is it definable in arithmetic? That is: is the set $\{^\# A^\# : \mathbb{N} \models A\}$ definable in arithmetic? Tarski’s theorem answers this in the negative.

Theorem 5.12. The set of true statements of arithmetic is not definable in arithmetic.

Proof. Suppose $D(x)$ defined it. By the fixed-point lemma, there is a formula $A$ such that $Q$ proves $A \leftrightarrow \neg D(\rceil A \rceil)$, and hence $\mathbb{N} \models A \leftrightarrow \neg D(\rceil A \rceil)$. But then $\mathbb{N} \models A$ if and only if $\mathbb{N} \models \neg D(\rceil A \rceil)$, which contradicts the fact that $D(y)$ is supposed to define the set of true statements of arithmetic. □

Tarski applied this analysis to a more general philosophical notion of truth. Given any language $L$, Tarski argued that an adequate notion of truth for $L$ would have to satisfy, for each sentence $X$,

‘$X$’ is true if and only if $X$.

Tarski’s oft-quoted example, for English, is the sentence

‘Snow is white’ is true if and only if snow is white.
However, for any language strong enough to represent the diagonal function, and any linguistic predicate $T(x)$, we can construct a sentence $X$ satisfying “$X$ if and only if not $T(\text{‘}X\text{’})$.” Given that we do not want a truth predicate to declare some sentences to be both true and false, Tarski concluded that one cannot specify a truth predicate for all sentences in a language without, somehow, stepping outside the bounds of the language. In other words, a the truth predicate for a language cannot be defined in the language itself.

Summary

The first incompleteness theorem states that for any consistent, axiomatizable theory $T$ that extends $Q$, there is a sentence $G_T$ such that $T \not\Rightarrow G_T$. $G_T$ is constructed in such a way that $G_T$, in a roundabout way, says “$T$ does not prove $G_T$.” Since $T$ does not, in fact, prove it, what it says is true. If $N \models T$, then $T$ does not prove any false claims, so $T \not\Rightarrow \neg G_T$. Such a sentence is independent or undecidable in $T$. Gödel’s original proof established that $G_T$ is independent on the assumption that $T$ is $\omega$-consistent. Rosser improved the result by finding a different sentence $R_T$ with is neither provable nor refutable in $T$ as long as $ThT$ is simply consistent.

The construction of $G_T$ is effective: given an axiomatization of $T$ we could, in principle, write down $G_T$. The “roundabout way” in which $G_T$ states its own unprovability, is a special case of a general result, the fixed-point lemma. It states that for any formula $B(y)$ in $L_A$, there is a sentence $A$ such that $Q \models A \leftrightarrow B(\ulcorner A\urcorner)$. (Here, $\ulcorner A\urcorner$ is the standard numeral for the Gödel number of $A$, i.e., $\#A\). To obtain $G_T$, we use the formula $\neg \text{Prov}_T(y)$ as $B(y)$. We get $\text{Prov}_T$ as the culmination of our previous efforts: We know that $\text{Prf}_T(n, m)$, which holds if $n$ is the Gödel number of a derivation of the sentence with Gödel number $m$ from $T$, is primitive recursive. We also know that $Q$ represents all primitive recursive relations, and so there is some formula $\text{Prf}_T(x, y)$ that repre-
sents Prf\_T in Q. The provability predicate for T is \( \text{Prov}_T(y) \) is \( \exists x \text{Prf}_{T,T}(x,y) \) then expresses provability in T. (It doesn’t represent it though: if T ⊢ A, then Q ⊢ Prov\_T(⌜A⌝); but if T \not\models A, then Q does not in general prove \( \neg\text{Prov}_T(⌜A⌝) \).)

The second incompleteness theorem establishes that the sentence Con\_T that expresses that T is consistent, i.e., T also does not prove \( \neg\text{Prov}_T(⌜\bot⌝) \). The proof of the second incompleteness theorem requires some additional conditions on T, the provability conditions. PA satisfies them, although Q does not. Theories that satisfy the provability conditions also satisfy Löb’s theorem: \( T \vdash \text{Prov}_T(⌜A⌝) \rightarrow A \) iff \( T \vdash A \).

The fixed-point theorem also has another important consequence. We say a property \( Rn \) is definable in \( \mathcal{L}_A \) if there is a formula \( A_R(x) \) such that \( N \models A_R(n) \) iff \( Rn \) holds. For instance, \( \text{Prov}_T \) is definable, since \( \text{Prov}_T \) defines it. The property \( n \) has iff it is the Gödel number of a sentence true in N, however, is not definable. This is Tarski’s theorem about the undefinability of truth.

**Problems**

**Problem 5.1.** Show that PA proves \( G_{PA} \rightarrow \text{Con}_{PA} \).

**Problem 5.2.** Let T be a computably axiomatized theory, and let \( \text{Prov}_T \) be a provability predicate for T. Consider the following four statements:

1. If \( T \vdash A \), then \( T \vdash \text{Prov}_T(⌜A⌝) \).
2. \( T \vdash A \rightarrow \text{Prov}_T(⌜A⌝) \).
3. If \( T \vdash \text{Prov}_T(⌜A⌝) \), then \( T \vdash A \).
4. \( T \vdash \text{Prov}_T(⌜A⌝) \rightarrow A \)

Under what conditions are each of these statements true?
Problem 5.3. Show that $Q(n) \iff n \in \{^*A^# : Q \vdash A\}$ is definable in arithmetic.
CHAPTER 6

Models of Arithmetic

6.1 Introduction

The standard model of arithmetic is the structure $N$ with $|N| = \mathbb{N}$ in which $0$, $\cdot$, $+$, $\times$, and $<$ are interpreted as you would expect. That is, $0$ is $0$, $\cdot$ is the successor function, $+$ is interpreted as addition and $\times$ as multiplication of the numbers in $\mathbb{N}$. Specifically,

$$0^N = 0$$
$$\cdot^N(n) = n + 1$$
$$+^N(n, m) = n + m$$
$$\times^N(n, m) = nm$$

Of course, there are structures for $\mathcal{L}_A$ that have domains other than $\mathbb{N}$. For instance, we can take $M$ with domain $|M| = \{a\}^*$ (the finite sequences of the single symbol $a$, i.e., $\emptyset$, $a$, $aa$, $aaa$, $\ldots$), and interpretations

$$0^M = \emptyset$$
$$\cdot^M(s) = s \cdot a$$
$$+^M(n, m) = a^{n+m}$$
These two structures are “essentially the same” in the sense that the only difference is the elements of the domains but not how the elements of the domains are related among each other by the interpretation functions. We say that the two structures are isomorphic.

It is an easy consequence of the compactness theorem that any theory true in \( N \) also has models that are not isomorphic to \( N \). Such structures are called non-standard. The interesting thing about them is that while the elements of a standard model (i.e., \( N \), but also all structures isomorphic to it) are exhausted by the values of the standard numerals \( \bar{n} \), i.e.,

\[
|N| = \{ \text{Val}^N(\bar{n}) : n \in \mathbb{N} \}
\]

that isn’t the case in non-standard models: if \( M \) is non-standard, then there is at least one \( x \in |M| \) such that \( x \neq \text{Val}^M(\bar{n}) \) for all \( n \).

These non-standard elements are pretty neat: they are “infinite natural numbers.” But their existence also explains, in a sense, the incompleteness phenomena. Consider an example, e.g., the consistency statement for Peano arithmetic, \( \text{Con}_{\text{PA}} \), i.e., \( \neg \exists x \Prf_{\text{PA}}(x, \lnot \bot) \). Since \( \text{PA} \) neither proves \( \text{Con}_{\text{PA}} \) nor \( \neg \text{Con}_{\text{PA}} \), either can be consistently added to \( \text{PA} \). Since \( \text{PA} \) is consistent, \( N \models \text{Con}_{\text{PA}} \), and consequently \( N \not\models \neg \text{Con}_{\text{PA}} \). So \( N \) is not a model of \( \text{PA} \cup \{ \neg \text{Con}_{\text{PA}} \} \), and all its models must be nonstandard. Models of \( \text{PA} \cup \{ \neg \text{Con}_{\text{PA}} \} \) must contain some element that serves as the witness that makes \( \exists x \Prf_{\text{PA}}(\bar{n}, \lnot \bot) \) true, i.e., a Gödel number of a derivation of a contradiction from \( \text{PA} \). Such an element can’t be standard—since \( \text{PA} \vdash \neg \Prf_{\text{PA}}(\bar{n}, \lnot \bot) \) for every \( n \).

## 6.2 Reducts and Expansions

Often it is useful or necessary to compare languages which have symbols in common, as well as structures for these languages. The most common case is when all the symbols in a language \( \mathcal{L} \)
are also part of a language $\mathcal{L}'$, i.e., $\mathcal{L} \subseteq \mathcal{L}'$. An $\mathcal{L}$-structure $M$ can then always be expanded to an $\mathcal{L}'$-structure by adding interpretations of the additional symbols while leaving the interpretations of the common symbols the same. On the other hand, from an $\mathcal{L}'$-structure $M'$ we can obtain an $\mathcal{L}$-structure simply by “forgetting” the interpretations of the symbols that do not occur in $\mathcal{L}$.

**Definition 6.1.** Suppose $\mathcal{L} \subseteq \mathcal{L}'$, $M$ is an $\mathcal{L}$-structure and $M'$ is an $\mathcal{L}'$-structure. $M$ is the reduct of $M'$ to $\mathcal{L}$, and $M'$ is an expansion of $M$ to $\mathcal{L}'$ iff

1. $|M| = |M'|$
2. For every constant symbol $c \in \mathcal{L}$, $c^M = c^{M'}$.
3. For every function symbol $f \in \mathcal{L}$, $f^M = f^{M'}$.
4. For every predicate symbol $P \in \mathcal{L}$, $P^M = P^{M'}$.

**Proposition 6.2.** If an $\mathcal{L}$-structure $M$ is a reduct of an $\mathcal{L}'$-structure $M'$, then for all $\mathcal{L}$-sentences $A$,

$$M \models A \iff M' \models A.$$  

*Proof.* Exercise. □

**Definition 6.3.** When we have an $\mathcal{L}$-structure $M$, and $\mathcal{L}' = \mathcal{L} \cup \{P\}$ is the expansion of $\mathcal{L}$ obtained by adding a single $n$-place predicate symbol $P$, and $R \subseteq |M|^n$ is an $n$-place relation, then we write $(M, R)$ for the expansion $M'$ of $M$ with $P^M = R$.

### 6.3 Isomorphic Structures

First-order structures can be alike in one of two ways. One way in which the can be alike is that they make the same sentences
true. We call such structures \textit{elementarily equivalent}. But structures can be very different and still make the same sentences true—for instance, one can be countable and the other not. This is because there are lots of features of a structure that cannot be expressed in first-order languages, either because the language is not rich enough, or because of fundamental limitations of first-order logic such as the Löwenheim-Skolem theorem. So another, stricter, aspect in which structures can be alike is if they are fundamentally the same, in the sense that they only differ in the objects that make them up, but not in their structural features. A way of making this precise is by the notion of an \textit{isomorphism}.

\textbf{Definition 6.4.} Given two structures \( M \) and \( M' \) for the same language \( \mathcal{L} \), we say that \( M \) is \textit{elementarily equivalent to} \( M' \), written \( M \equiv M' \), if and only if for every sentence \( A \) of \( \mathcal{L} \), \( M \models A \) iff \( M' \models A \).

\textbf{Definition 6.5.} Given two structures \( M \) and \( M' \) for the same language \( \mathcal{L} \), we say that \( M \) is \textit{isomorphic to} \( M' \), written \( M \simeq M' \), if and only if there is a function \( h : |M| \rightarrow |M'| \) such that:

1. \( h \) is injective: if \( h(x) = h(y) \) then \( x = y \);
2. \( h \) is surjective: for every \( y \in |M'| \) there is \( x \in |M| \) such that \( h(x) = y \);
3. for every constant symbol \( c \): \( h(c^M) = c^{M'} \);
4. for every \( n \)-place predicate symbol \( P \):
   \[ \langle a_1, \ldots, a_n \rangle \in P^M \iff \langle h(a_1), \ldots, h(a_n) \rangle \in P^{M'} ; \]
5. for every \( n \)-place function symbol \( f \):
   \[ h(f^M(a_1, \ldots, a_n)) = f^{M'}(h(a_1), \ldots, h(a_n)) . \]
**Theorem 6.6.** If $M \approx M'$ then $M \equiv M'$.

*Proof.* Let $h$ be an isomorphism of $M$ onto $M'$. For any assignment $s$, $h \circ s$ is the composition of $h$ and $s$, i.e., the assignment in $M'$ such that $(h \circ s)(x) = h(s(x))$. By induction on $t$ and $A$ one can prove the stronger claims:

a. $h(\text{Val}_s^M(t)) = \text{Val}_{h \circ s}^{M'}(t)$.

b. $M, s \models A$ iff $M', h \circ s \models A$.

The first is proved by induction on the complexity of $t$.

1. If $t \equiv c$, then $\text{Val}_s^M(c) = c^M$ and $\text{Val}_{h \circ s}^{M'}(c) = c^{M'}$. Thus, $h(\text{Val}_s^M(t)) = h(c^M) = c^{M'}$ (by (3) of Definition 6.5) = $\text{Val}_{h \circ s}^{M'}(t)$.

2. If $t \equiv x$, then $\text{Val}_s^M(x) = s(x)$ and $\text{Val}_{h \circ s}^{M'}(x) = h(s(x))$. Thus, $h(\text{Val}_s^M(x)) = h(s(x)) = \text{Val}_{h \circ s}^{M'}(x)$.

3. If $t \equiv f(t_1, \ldots, t_n)$, then

\[
\text{Val}_s^M(t) = f^M(\text{Val}_s^M(t_1), \ldots, \text{Val}_s^M(t_n)) \quad \text{and} \quad \text{Val}_{h \circ s}^{M'}(t) = f^{M'}(\text{Val}_{h \circ s}^{M'}(t_1), \ldots, \text{Val}_{h \circ s}^{M'}(t_n)).
\]

The induction hypothesis is that for each $i$, $h(\text{Val}_s^M(t_i)) = \text{Val}_{h \circ s}^{M'}(t_i)$. So,

\[
h(\text{Val}_s^M(t)) = h(f^M(\text{Val}_s^M(t_1), \ldots, \text{Val}_s^M(t_n))
\]
\[
= h(f^M(\text{Val}_{h \circ s}^{M'}(t_1), \ldots, \text{Val}_{h \circ s}^{M'}(t_n))) \quad (6.1)
\]
\[
= f^{M'}(\text{Val}_{h \circ s}^{M'}(t_1), \ldots, \text{Val}_{h \circ s}^{M'}(t_n)) \quad (6.2)
\]
\[
= \text{Val}_{h \circ s}^{M'}(t)
\]

Here, eq. (6.1) follows by induction hypothesis and eq. (6.2) by (5) of Definition 6.5.
Part (2) is left as an exercise.

If $A$ is a sentence, the assignments $s$ and $h \circ s$ are irrelevant, and we have $M \models A$ iff $M' \models A$. □

**Definition 6.7.** An *automorphism* of a structure $\mathcal{M}$ is an isomorphism of $\mathcal{M}$ onto itself.

### 6.4 The Theory of a Structure

Every structure $M$ makes some sentences true, and some false. The set of all the sentences it makes true is called its *theory*. That set is in fact a theory, since anything it entails must be true in all its models, including $M$.

**Definition 6.8.** Given a structure $M$, the *theory* of $M$ is the set $\text{Th}(M)$ of sentences that are true in $M$, i.e., $\text{Th}(M) = \{ A : M \models A \}$.

We also use the term “theory” informally to refer to sets of sentences having an intended interpretation, whether deductively closed or not.

**Proposition 6.9.** For any $M$, $\text{Th}(M)$ is complete.

*Proof.* For any sentence $A$ either $M \models A$ or $M \models \neg A$, so either $A \in \text{Th}(M)$ or $\neg A \in \text{Th}(M)$. □
**Proposition 6.10.** If $N \models A$ for every $A \in \text{Th}(M)$, then $M \equiv N$.

*Proof.* Since $N \models A$ for all $A \in \text{Th}(M)$, $\text{Th}(M) \subseteq \text{Th}(N)$. If $N \models A$, then $N \not\models \neg A$, so $\neg A \notin \text{Th}(M)$. Since $\text{Th}(M)$ is complete, $A \in \text{Th}(M)$. So, $\text{Th}(N) \subseteq \text{Th}(M)$, and we have $M \equiv N$. □

**Remark 1.** Consider $R = \langle \mathbb{R}, < \rangle$, the structure whose domain is the set $\mathbb{R}$ of the real numbers, in the language comprising only a 2-place predicate symbol interpreted as the $<$ relation over the reals. Clearly $R$ is uncountable; however, since $\text{Th}(R)$ is obviously consistent, by the Löwenheim-Skolem theorem it has a countable model, say $S$, and by Proposition 6.10, $R \equiv S$. Moreover, since $R$ and $S$ are not isomorphic, this shows that the converse of Theorem 6.6 fails in general.

### 6.5 Standard Models of Arithmetic

The language of arithmetic $\mathcal{L}_A$ is obviously intended to be about numbers, specifically, about natural numbers. So, “the” standard model $N$ is special: it is the model we want to talk about. But in logic, we are often just interested in structural properties, and any two structures that are isomorphic share those. So we can be a bit more liberal, and consider any structure that is isomorphic to $N$ “standard.”

**Definition 6.11.** A structure for $\mathcal{L}_A$ is **standard** if it is isomorphic to $N$. 
Proposition 6.12. If a structure $M$ standard, its domain is the set of values of the standard numerals, i.e.,

$$|M| = \{ \text{Val}^M(n) : n \in \mathbb{N} \}$$

\textbf{Proof.} Clearly, every $\text{Val}^M(n) \in |M|$. We just have to show that every $x \in |M|$ is equal to $\text{Val}^M(n)$ for some $n$. Since $M$ is standard, it is isomorphic to $\mathbb{N}$. Suppose $g : \mathbb{N} \to |M|$ is an isomorphism. Then $g(n) = g(\text{Val}^\mathbb{N}(n)) = \text{Val}^M(n)$. But for every $x \in |M|$, there is an $n \in \mathbb{N}$ such that $g(n) = x$, since $g$ is surjective. \hfill \Box

If a structure $M$ for $\mathcal{L}_A$ is standard, the elements of its domain can all be named by the standard numerals $0, 1, 2, \ldots$, i.e., the terms $0, 0', 0'', \ldots$, etc. Of course, this does not mean that the elements of $|M|$ \emph{are} the numbers, just that we can pick them out the same way we can pick out the numbers in $|\mathbb{N}|$.

Proposition 6.13. If $M \models Q$, and $|M| = \{ \text{Val}^M(n) : n \in \mathbb{N} \}$, then $M$ is standard.

\textbf{Proof.} We have to show that $M$ is isomorphic to $\mathbb{N}$. Consider the function $g : \mathbb{N} \to |M|$ defined by $g(n) = \text{Val}^M(n)$. By the hypothesis, $g$ is surjective. It is also injective: $Q \models \overline{n} \neq \overline{m}$ whenever $n \neq m$. Thus, since $M \models Q$, $M \models \overline{n} \neq \overline{m}$, whenever $n \neq m$. Thus, if $n \neq m$, then $\text{Val}^M(n) \neq \text{Val}^M(m)$, i.e., $g(n) \neq g(m)$.

We also have to verify that $g$ is an isomorphism.

1. We have $g(0^N) = g(0)$ since, $0^N = 0$. By definition of $g$, $g(0) = \text{Val}^M(0)$. But $0$ is just $0$, and the value of a term which happens to be a constant symbol is given by what the structure assigns to that constant symbol, i.e., $\text{Val}^M(0) = 0^M$. So we have $g(0^N) = 0^M$ as required.

2. $g(\text{n}(n)) = g(n + 1)$, since $\text{n}$ in $\mathbb{N}$ is the successor function on $\mathbb{N}$. Then, $g(n + 1) = \text{Val}^M(n + 1)$ by definition of $g$. But $n + 1$ is the same term as $n'$, so $\text{Val}^M(n + 1) = \text{Val}^M(n')$. By
the definition of the value function, this is \( r^M(\text{Val}^M(\bar{n})) \). Since \( \text{Val}^M(\bar{n}) = g(n) \) we get \( g(r^N(n)) = r^M(g(n)) \).

3. \( g(\bar{+}^N(n, m)) = g(n+m) \), since \( + \) in \( N \) is the addition function on \( \mathbb{N} \). Then, \( g(n+m) = \text{Val}^M(\bar{n} + \bar{m}) \) by definition of \( g \). But \( \mathbb{Q} \vdash n + m = (\bar{n} + \bar{m}) \), so \( \text{Val}^M(n + m) = \text{Val}^M(\bar{n} + \bar{m}) \). By the definition of the value function, this is \( = +^M(\text{Val}^M(\bar{n}), \text{Val}^M(\bar{m})) \). Since \( \text{Val}^M(\bar{n}) = g(n) \) and \( \text{Val}^M(\bar{m}) = g(m) \), we get \( g(\bar{+}^N(n, m)) = +^M(g(n), g(m)) \).

4. \( g(\bar{\times}^N(n, m)) = \times^M(g(n), g(m)) \): Exercise.

5. \( \langle n, m \rangle \in \langle^N \) iff \( n < m \). If \( n < m \), then \( \mathbb{Q} \vdash \bar{n} < \bar{m} \), and also \( M \models \bar{n} < \bar{m} \). Thus \( \langle \text{Val}^M(\bar{n}), \text{Val}^M(\bar{m}) \rangle \in \langle^M \), i.e., \( \langle g(n), g(m) \rangle \in \langle^M \). If \( n \not< m \), then \( \mathbb{Q} \vdash \neg \bar{n} < \bar{m} \), and consequently \( M \not\models \bar{n} < \bar{m} \). Thus, as before, \( \langle g(n), g(m) \rangle \not\in \langle^M \). Together, we get: \( \langle n, m \rangle \in \langle^N \) iff \( \langle g(n), g(m) \rangle \in \langle^M \).

\( \square \)

The function \( g \) is the most obvious way of defining a mapping from \( \mathbb{N} \) to the domain of any other structure \( M \) for \( \mathcal{L}_A \), since every such \( M \) contains elements named by \( 0, \bar{1}, \bar{2}, \) etc. So it isn’t surprising that if \( M \) makes at least some basic statements about the \( \bar{n} \)'s true in the same way that \( N \) does, and \( g \) is also bijective, then \( g \) will turn into an isomorphism. In fact, if \( |M| \) contains no elements other than what the \( \bar{n} \)'s name, it’s the only one.

**Proposition 6.14.** If \( M \) is standard, then \( g \) from the proof of Proposition 6.13 is the only isomorphism from \( N \) to \( M \).

**Proof.** Suppose \( h \colon \mathbb{N} \to |M| \) is an isomorphism between \( N \) and \( M \). We show that \( g = h \) by induction on \( n \). If \( n = 0 \), then \( g(0) = \bar{0}^M \) by definition of \( g \). But since \( h \) is an isomorphism, \( h(0) = h(\bar{0}^N) = \bar{0}^M \), so \( g(0) = h(0) \).

Now consider the case for \( n + 1 \). We have

\[ g(n + 1) = \text{Val}^M(\bar{n} + \bar{1}) \] by definition of \( g \).
For any countably infinite set $X$, there’s a bijection between $\mathbb{N}$ and $X$, so every such set $X$ is potentially the domain of a standard model. In fact, once you pick an object $z \in X$ and a suitable function $s : X \to X$ as $o^X$ and $r^X$, the interpretation of $+$, $\times$, and $<$ is already fixed. Only functions $s = r^X$ that are both injective and surjective are suitable in a standard model. It has to be injective since the successor function in $\mathbb{N}$ is, and that $r$ is injective is expressed by a sentence true in $\mathbb{N}$ which $X$ thus also has to make true. It has to be surjective because otherwise there would be some $x \in X$ not in the domain of $s$, i.e., the sentence $\forall x \exists y y' = x$ would be false—but it is true in $\mathbb{N}$.

### 6.6 Non-Standard Models

We call a structure for $\mathcal{L}_A$ standard if it is isomorphic to $\mathbb{N}$. If a structure isn’t isomorphic to $\mathbb{N}$, it is called non-standard.

**Definition 6.15.** A structure $M$ for $\mathcal{L}_A$ is non-standard if it is not isomorphic to $\mathbb{N}$. The elements $x \in |M|$ which are equal to $\text{Val}^M(\bar{n})$ for some $n \in \mathbb{N}$ are called standard numbers (of $M$), and those not, non-standard numbers.

By Proposition 6.12, any standard structure for $\mathcal{L}_A$ contains only standard elements. Consequently, a non-standard structure
must contain at least one non-standard element. In fact, the existence of a non-standard element guarantees that the structure is non-standard.

**Proposition 6.16.** If a structure $M$ for $\mathcal{L}_A$ contains a non-standard number, $M$ is non-standard.

*Proof.* Suppose not, i.e., suppose $M$ standard but contains a non-standard number $x$. Let $g : \mathbb{N} \to |M|$ be an isomorphism. It is easy to see (by induction on $n$) that $g(\text{Val}^N(n)) = \text{Val}^M(n)$. In other words, $g$ maps standard numbers of $N$ to standard numbers of $M$. If $M$ contains a non-standard number, $g$ cannot be surjective, contrary to hypothesis. \[\square\]

It is easy enough to specify non-standard structures for $\mathcal{L}_A$. For instance, take the structure with domain $\mathbb{Z}$ and interpret all non-logical symbols as usual. Since negative numbers are not values of $\bar{n}$ for any $n$, this structure is non-standard. Of course, it will not be a model of arithmetic in the sense that it makes the same sentences true as $\mathbb{N}$. For instance, $\forall x \, x' \neq 0$ is false. However, we can prove that non-standard models of arithmetic exist easily enough, using the compactness theorem.

**Proposition 6.17.** Let $\mathcal{T}A = \{ A : \mathbb{N} \models A \}$ be the theory of $\mathbb{N}$. $\mathcal{T}A$ has a countable non-standard model.

*Proof.* Expand $\mathcal{L}_A$ by a new constant symbol $c$ and consider the set of sentences

$$\Gamma = \mathcal{T}A \cup \{ c \neq \bar{0}, c \neq \bar{1}, c \neq \bar{2}, \ldots \}$$

Any model $M^c$ of $\Gamma$ would contain an element $x = c^M$ which is non-standard, since $x \neq \text{Val}^M(\bar{n})$ for all $n \in \mathbb{N}$. Also, obviously, $M^c \models \mathcal{T}A$, since $\mathcal{T}A \subseteq \Gamma$. If we turn $M^c$ into a structure $M$ for $\mathcal{L}_A$ simply by forgetting about $c$, its domain still contains the non-standard $x$, and also $M \models \mathcal{T}A$. The latter is guaranteed since $c$ does not occur in $\mathcal{T}A$. So, it suffices to show that $\Gamma$ has a model.
We use the compactness theorem to show that \( \Gamma \) has a model. If every finite subset of \( \Gamma \) is satisfiable, so is \( \Gamma \). Consider any finite subset \( \Gamma_0 \subseteq \Gamma \). \( \Gamma_0 \) includes some sentences of \( \text{TA} \) and some of the form \( c \neq \bar{n} \), but only finitely many. Suppose \( k \) is the largest number so that \( c \neq \bar{k} \in \Gamma_0 \). Define \( N_k \) by expanding \( N \) to include the interpretation \( c^{N_k} = k + 1 \). \( N_k \models \Gamma_0 \): if \( A \in \text{TA} \), \( N_k \models A \) since \( N_k \) is just like \( N \) in all respects except \( c \), and \( c \) does not occur in \( A \). And \( N_k \models c \neq \bar{n} \), since \( n \leq k \), and \( \text{Val}^{N_k}(c) = k + 1 \). Thus, every finite subset of \( \Gamma \) is satisfiable. \( \square \)

### 6.7 Models of \( \mathcal{Q} \)

We know that there are non-standard structures that make the same sentences true as \( N \) does, i.e., is a model of \( \text{TA} \). Since \( N \models \mathcal{Q} \), any model of \( \text{TA} \) is also a model of \( \mathcal{Q} \). \( \mathcal{Q} \) is much weaker than \( \text{TA} \), e.g., \( \mathcal{Q} \not\models \forall x \forall y (x + y) = (y + x) \). Weaker theories are easier to satisfy: they have more models. E.g., \( \mathcal{Q} \) has models which make \( \forall x \forall y (x + y) = (y + x) \) false, but those cannot also be models of \( \text{TA} \), or \( \text{PA} \) for that matter. Models of \( \mathcal{Q} \) are also relatively simple: we can specify them explicitly.

**Example 6.18.** Consider the structure \( K \) with domain \( |K| = \mathbb{N} \cup \{a\} \) and interpretations

\[
\begin{align*}
\circ^K &= 0 \\
\land^K(x) &= \begin{cases} 
  x + 1 & \text{if } x \in \mathbb{N} \\
  a & \text{if } x = a
\end{cases} \\
\lor^K(x, y) &= \begin{cases} 
  x + y & \text{if } x, y \in \mathbb{N} \\
  a & \text{otherwise}
\end{cases} \\
\times^K(x, y) &= \begin{cases} 
  xy & \text{if } x, y \in \mathbb{N} \\
  a & \text{otherwise}
\end{cases} \\
<^K &= \{(x, y) : x, y \in \mathbb{N} \text{ and } x < y\} \cup \{(x, a) : x \in |K|\}
\end{align*}
\]
To show that $K \models Q$ we have to verify that all axioms of $Q$ are true in $K$. For convenience, let’s write $x^*$ for $K(x)$ (the “successor” of $x$ in $K$), $x \oplus y$ for $+K(x, y)$ (the “sum” of $x$ and $y$ in $K$, $x \otimes y$ for $\times K(x, y)$ (the “product” of $x$ and $y$ in $K$), and $x \otimes y$ for $\langle x, y \rangle \in < K$. With these abbreviations, we can give the operations in $K$ more perspicuously as

<table>
<thead>
<tr>
<th>$x$</th>
<th>$x^*$</th>
<th>$x \oplus y$</th>
<th>$m$</th>
<th>$a$</th>
<th>$x \otimes y$</th>
<th>$m$</th>
<th>$a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>$n + 1$</td>
<td>$n + m$</td>
<td>$a$</td>
<td>$a$</td>
<td>$n$</td>
<td>$nm$</td>
<td>$a$</td>
</tr>
<tr>
<td>$a$</td>
<td>$a$</td>
<td>$a$</td>
<td>$a$</td>
<td>$a$</td>
<td>$a$</td>
<td>$a$</td>
<td>$a$</td>
</tr>
</tbody>
</table>

We have $n \otimes m$ iff $n < m$ for $n, m \in \mathbb{N}$ and $x \otimes a$ for all $x \in |K|$.

$K \models \forall x \forall y (x' = y' \rightarrow x = y)$ since $*$ is injective. $K \models \forall x \circ \neq x'$ since 0 is not a $*$-successor in $K$. $N \models \forall x (x \neq 0 \rightarrow \exists y x = y')$ since for every $n > 0$, $n = (n - 1)^*$, and $a = a^*$.

$K \models \forall x (x + 0) = x$ since $n \oplus 0 = n + 0 = n$, and $a \oplus 0 = a$ by definition of $\oplus$. $K \models \forall x \forall y (x + y') = (x + y)'$ is a bit trickier. If $n, m$ are both standard, we have:

$$(n \oplus m^*) = (n + (m + 1)) = (n + m) + 1 = (n \oplus m)^*$$

since $\oplus$ and $^*$ agree with $+$ and $'$ on standard numbers. Now suppose $x \in |K|$. Then

$$(x \oplus a^*) = (x \oplus a) = a = a^* = (x \oplus a)^*$$

The remaining case is if $y \in |K|$ but $x = a$. Here we also have to distinguish cases according to whether $y = n$ is standard or $y = b$:

$$(a \oplus n^*) = (a \oplus (n + 1)) = a = a^* = (x \oplus n)^*$$

$$(a \oplus a^*) = (a \oplus a) = a = a^* = (x \oplus a)^*$$

This is of course a bit more detailed than needed. For instance, since $a \oplus z = a$ whatever $z$ is, we can immediately conclude $a \oplus a^* = a$. The remaining axioms can be verified the same way.

$K$ is thus a model of $Q$. Its “addition” $\oplus$ is also commutative. But there are other sentences true in $N$ but false in $K$, and vice versa. For instance, $a \otimes a$, so $K \models \exists x x < x$ and $K \not\models \forall x x < x$. This shows that $Q \not\models \forall x \neg x < x$. 


Example 6.19. Consider the structure $L$ with domain $|L| = \mathbb{N} \cup \{a, b\}$ and interpretations $\mathcal{I}^L = \cdot, +^L = \oplus$ given by

<table>
<thead>
<tr>
<th>$x$</th>
<th>$x^*$</th>
<th>$x \oplus y$</th>
<th>$m$</th>
<th>$a$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>$n + 1$</td>
<td>$n$</td>
<td>$n + m$</td>
<td>$b$</td>
<td>$a$</td>
</tr>
<tr>
<td>$a$</td>
<td>$a$</td>
<td>$a$</td>
<td>$a$</td>
<td>$b$</td>
<td>$a$</td>
</tr>
<tr>
<td>$b$</td>
<td>$b$</td>
<td>$b$</td>
<td>$b$</td>
<td>$b$</td>
<td>$a$</td>
</tr>
</tbody>
</table>

Since $\cdot$ is injective, 0 is not in its range, and every $x \in |L|$ other than 0 is, axioms $Q_1$–$Q_3$ are true in $L$. For any $x$, $x \oplus 0 = x$, so $Q_4$ is true as well. For $Q_5$, consider $x \oplus y^*$ and $(x \oplus y)^*$. They are equal if $x$ and $y$ are both standard, since then $\cdot$ and $\oplus$ agree with $\cdot$ and $\circ$. If $x$ is non-standard, and $y$ is standard, we have $x \oplus y^* = x = x^* = (x \oplus y)^*$. If $x$ and $y$ are both non-standard, we have four cases:

\[
\begin{align*}
    a \oplus a^* &= b = b^* = (a \oplus a)^* \\
    b \oplus b^* &= a = a^* = (b \oplus b)^* \\
    b \oplus a^* &= b = b^* = (b \oplus y)^* \\
    a \oplus b^* &= a = a^* = (a \oplus b)^*
\end{align*}
\]

If $x$ is standard, but $y$ is non-standard, we have

\[
\begin{align*}
    n \oplus a^* &= n \oplus a = b = b^* = (n \oplus a)^* \\
    n \oplus b^* &= n \oplus b = a = a^* = (n \oplus b)^*
\end{align*}
\]

So, $L \models Q_5$. However, $a \oplus 0 \neq 0 \oplus a$, so $L \not\models \forall x \forall y (x + y) = (y + x)$.

We’ve explicitly constructed models of $\mathbb{Q}$ in which the non-standard elements live “beyond” the standard elements. In fact, that much is required by the axioms. A non-standard element $x$ cannot be $\oplus 0$. Otherwise, for some $z$, $x \oplus z^* = 0$ by $Q_8$. But then $0 = x \oplus z^* = (x \oplus z)^*$ by $Q_5$, contradicting $Q_2$. Also, for every $n$, $\mathcal{Q} \vdash \forall x (x < \overline{n} \to (x = 0 \lor x = 1 \lor \cdots \lor x = \overline{n}))$, so we can’t have $a \oplus n$ for any $n > 0$. 

6.8 Models of PA

Any non-standard model of TA is also one of PA. We know that non-standard models of TA and hence of PA exist. We also know that such non-standard models contain non-standard “numbers,” i.e., elements of the domain that are “beyond” all the standard “numbers.” But how are they arranged? How many are there? We’ve seen that models of the weaker theory Q can contain as few as a single non-standard number. But these simple structures are not models of PA or TA.

The key to understanding the structure of models of PA or TA is to see what facts are derivable in these theories. For instance, already PA proves that ∀x x ≠ x’ and ∀x ∀y (x+y) = (y+x), so this rules out simple structures (in which these sentences are false) as models of PA.

Suppose M is a model of PA. Then if PA ⊢ A, M ⊨ A. Let’s again use z for 0M, * for 1M, ⊕ for +M, ⊗ for ×M, and ⊥ for ⊥M. Any sentence A then states some condition about z, *, ⊕, ⊗, and ⊥, and if M ⊨ A that condition must be satisfied. For instance, if M ⊨ Q1, i.e., M ⊨ ∀x ∀y (x’ = y’ → x = y), then * must be injective.

**Proposition 6.20.** In M, ⊗ is a linear strict order, i.e., it satisfies:

1. Not x ⊗ x for any x ∈ |M|.
2. If x ⊗ y and y ⊗ z then x ⊗ z.
3. For any x ≠ y, x ⊗ y or y ⊗ x.

**Proof.** PA proves:

1. ∀x ¬x < x
2. ∀x ∀y ∀z ((x < y ∧ y < z) → x < z)
3. ∀x ∀y ((x < y ∨ y < x) ∨ x = y)
Proposition 6.21. \( z \) is the least element of \(|M|\) in the \( \odot \)-ordering. For any \( x, x \odot x^*, \) and \( x^* \) is the \( \odot \)-least element with that property. For any \( x \), there is a unique \( y \) such that \( y^* = x \). (We call \( y \) the “predecessor” of \( x \) in \( M \), and denote it by \( \ast x \).)

Proof. Exercise. \( \square \)

Proposition 6.22. All standard elements of \( M \) are less than (according to \( \odot \)) than all non-standard elements.

Proof. We’ll use \( n \) as short for \( \text{Val}^M(n) \), a standard element of \( M \). Already \( Q \) proves that, for any \( n \in \mathbb{N}, \forall x (x < \overline{n} \rightarrow (x = \overline{0} \lor x = \overline{1} \lor \cdots \lor x = \overline{n})) \). There are no elements that are \( \odot z \). So if \( n \) is standard and \( x \) is non-standard, we cannot have \( x \odot n \). By definition, a non-standard element is one that isn’t \( \text{Val}^M(n) \) for any \( n \in \mathbb{N} \), so \( x \neq n \) as well. Since \( \odot \) is a linear order, we must have \( n \odot x \). \( \square \)

Proposition 6.23. Every nonstandard element \( x \) of \(|M|\) is an element of the subset

\[ \ldots \ast x \odot \ast x \odot x \odot x \odot x \odot x \odot x \odot x \odot x \odot x \odot \ldots \]

We call this subset the block of \( x \) and write it as \( [x] \). It has no least and no greatest element. It can be characterized as the set of those \( y \in |M| \) such that, for some standard \( n, x \odot n = y \) or \( y \odot n = x \).

Proof. Clearly, such a set \([x]\) always exists since every element \( y \) of \(|M|\) has a unique successor \( y^* \) and unique predecessor \( \ast y \). For successive elements \( y, y^* \) we have \( y \odot y^* \) and \( y^* \) is the \( \odot \)-least element of \(|M|\) such that \( y \) is \( \odot \)-less than it. Since always \( \ast y \odot y \) and \( y \odot y^* \), \([x]\) has no least or greatest element. If \( y \in [x] \) then \( x \in [y] \), for then either \( y^{*\ast \ldots \ast} = x \) or \( x^{*\ast \ldots \ast} = y \). If \( y^{*\ast \ldots \ast} = x \) (with \( n \) \( \ast \)'s), then \( y \odot n = x \) and conversely, since \( \text{PA} \vdash \forall x x^{*\ast \ldots \ast} = (x + \overline{n}) \) (if \( n \) is the number of \( \ast \)'s). \( \square \)
**Proposition 6.24.** If \([x] \neq [y]\) and \(x \ominus y\), then for any \(u \in [x]\) and any \(v \in [y]\), \(u \ominus v\).

**Proof.** Note that \(\text{PA} \vdash \forall x \forall y (x < y \rightarrow (x < y \lor x' = y))\). Thus, if \(u \ominus v\), we also have \(u \oplus n^* \ominus v\) for any \(n\) if \([u] \neq [v]\).

Any \(u \in [x]\) is \(\ominus y\) by assumption. If \(u \ominus x\), \(u \ominus y\) by transitivity. And if \(x \ominus u\) but \(u \in [x]\), we have \(u = x \oplus n^*\) for some \(n\), and so \(u \ominus y\) by the fact just proved.

Now suppose that \(v \in [y]\) is \(\ominus y\), i.e., \(v \oplus m^* = y\) for some standard \(m\). This rules out \(v \ominus x\), otherwise \(y = v \oplus m^* \ominus x\). Clearly also, \(x \neq v\), otherwise \(x \ominus m^* = v \ominus m^* = y\) and we would have \([x] = [y]\). So, \(x \ominus v\). But then also \(x \ominus n^* \ominus v\) for any \(n\). Hence, if \(x \ominus u\) and \(u \in [x]\), we have \(u \ominus v\). If \(u \ominus x\) then \(u \ominus v\) by transitivity.

Lastly, if \(y \ominus v\), \(u \ominus v\) since, as we’ve shown, \(u \ominus y\) and \(y \ominus v\). \(\square\)

**Corollary 6.25.** If \([x] \neq [y]\), \([x] \cap [y] = \emptyset\).

**Proof.** Suppose \(z \in [x]\) and \(x \ominus y\). Then \(z \ominus u\) for all \(u \in [y]\). If \(z \in [y]\), we would have \(z \ominus z\). Similarly if \(y \ominus x\). \(\square\)

This means that the blocks themselves can be ordered in a way that respects \(\ominus\): \([x] \ominus [y]\) iff \(x \ominus y\), or, equivalently, if \(u \ominus v\) for any \(u \in [x]\) and \(v \in [y]\). Clearly, the standard block \([0]\) is the least block. It intersects with no non-standard block, and no two non-standard blocks intersect either. Specifically, you cannot “reach” a different block by taking repeated successors or predecessors.

**Proposition 6.26.** If \(x\) and \(y\) are non-standard, then \(x \ominus x \ominus y\) and \(x \ominus y \notin [x]\).

**Proof.** If \(y\) is nonstandard, then \(y \neq z\). \(\text{PA} \vdash \forall x (y \neq 0 \rightarrow x < (x + y))\). Now suppose \(x \ominus y \in [x]\). Since \(x \ominus x \ominus y\), we would have \(x \ominus n^* = x \ominus y\). But \(\text{PA} \vdash \forall x \forall y \forall z ((x + y) = (x + z) \rightarrow y = z)\) (the cancellation law for addition). This would mean \(y = n^*\) for some standard \(n\); but \(y\) is assumed to be non-standard. \(\square\)
6.8. MODELS OF PA

Proposition 6.27. There is no least non-standard block.

Proof. PA ⊢ ∀x ∃y ((y + y) = x ∨ (y + y)' = x), i.e., that every x is divisible by 2 (possibly with remainder 1. If x is non-standard, so is y. By the preceding proposition, y ⊕ y ⊗ y and y ⊕ y ∉ [y]. Then also y ⊕ (y ⊗ y)* and (y ⊗ y)* ∉ [y]. But x = y ⊕ y or x = (y ⊗ y)*, so y ⊗ x and y ∉ [x]. □

Proposition 6.28. There is no largest block.

Proof. Exercise. □

Proposition 6.29. The ordering of the blocks is dense. That is, if x ⊗ y and [x] ≠ [y], then there is a block [z] distinct from both that is between them.

Proof. Suppose x ⊗ y. As before, x ⊗ y is divisible by two (possibly with remainder): there is a z ∈ |M| such that either x ⊗ y = z ⊗ z or x ⊗ y = (z ⊗ z)*. The element z is the “average” of x and y, and x ⊗ z and z ⊗ y. □

The non-standard blocks are therefore ordered like the rationals: they form a countably infinite linear ordering without endpoints. One can show that any two such countably infinite orderings are isomorphic. It follows that for any two countable non-standard models M₁ and M₂ of true arithmetic, their reducts to the language containing < and = only are isomorphic. Indeed, an isomorphism h can be defined as follows: the standard parts of M₁ and M₂ are isomorphic to the standard model N and hence to each other. The blocks making up the non-standard part are themselves ordered like the rationals and therefore isomorphic; an isomorphism of the blocks can be extended to an isomorphism within the blocks by matching up arbitrary elements in each, and then taking the image of the successor of x in M₁ to be the successor of the image of x in M₂. Note that it does not follow that
$\mathcal{M}_1$ and $\mathcal{M}_2$ are isomorphic in the full language of arithmetic (indeed, isomorphism is always relative to a language), as there are non-isomorphic ways to define addition and multiplication over $|M_1|$ and $|M_2|$. (This also follows from a famous theorem due to Vaught that the number of countable models of a complete theory cannot be 2.)

6.9 Computable Models of Arithmetic

The standard model $N$ has two nice features. Its domain is the natural numbers $\mathbb{N}$, i.e., its elements are just the kinds of things we want to talk about using the language of arithmetic, and the standard numeral $\bar{n}$ actually picks out $n$. The other nice feature is that the interpretations of the non-logical symbols of $\mathcal{L}_A$ are all computable. The successor, addition, and multiplication functions which serve as $\bar{N}, +^N$, and $\times^N$ are computable functions of numbers. (Computable by Turing machines, or definable by primitive recursion, say.) And the less-than relation on $N$, i.e., $<^N$, is decidable.

Non-standard models of arithmetical theories such as $\mathcal{Q}$ and $\mathcal{PA}$ must contain non-standard elements. Thus their domains typically include elements in addition to $\mathbb{N}$. However, any countable structure can be built on any countably infinite set, including $\mathbb{N}$. So there are also non-standard models with domain $\mathbb{N}$. In such models $M$, of course, at least some numbers cannot play the roles they usually play, since some $k$ must be different from $\text{Val}^M(\bar{n})$ for all $n \in \mathbb{N}$.

**Definition 6.30.** A structure $M$ for $\mathcal{L}_A$ is **computable** iff $|M| = \mathbb{N}$ and $\bar{N}, +^M, \times^M$ are computable functions and $<^M$ is a decidable relation.

**Example 6.31.** Recall the structure $K$ from Example 6.18. Its domain was $|K| = \mathbb{N} \cup \{a\}$ and interpretations

$$0^K = 0$$
\[ r_K(x) = \begin{cases} x + 1 & \text{if } x \in \mathbb{N} \\ a & \text{if } x = a \end{cases} \]

\[ +_K(x, y) = \begin{cases} x + y & \text{if } x, y \in \mathbb{N} \\ a & \text{otherwise} \end{cases} \]

\[ \times_K(x, y) = \begin{cases} xy & \text{if } x, y \in \mathbb{N} \\ a & \text{otherwise} \end{cases} \]

\[ <^K = \{ (x, y) : x, y \in \mathbb{N} \text{ and } x < y \} \cup \{ (x, a) : n \in |K| \} \]

But \(|K|\) is countably infinite and so is equinumerous with \(\mathbb{N}\). For instance, \(g : \mathbb{N} \to |K|\) with \(g(0) = a\) and \(g(n) = n + 1\) for \(n > 0\) is a bijection. We can turn it into an isomorphism between a new model \(K'\) of \(\mathbb{Q}\) and \(K\). In \(K'\), we have to assign different functions and relations to the symbols of \(\mathcal{L}_A\), since different elements of \(\mathbb{N}\) play the roles of standard and non-standard numbers.

Specifically, 0 now plays the role of \(a\), not of the smallest standard number. The smallest standard number is now 1. So we assign \(\sigma^{K'} = 1\). The successor function is also different now: given a standard number, i.e., an \(n > 0\), it still returns \(n + 1\). But 0 now plays the role of \(a\), which is its own successor. So \(r^{K'}(0) = 0\).

For addition and multiplication we likewise have

\[ +^{K'}(x, y) = \begin{cases} x + y & \text{if } x, y > 0 \\ 0 & \text{otherwise} \end{cases} \]

\[ \times^{K'}(x, y) = \begin{cases} xy & \text{if } x, y > 0 \\ 0 & \text{otherwise} \end{cases} \]

And we have \((x, y) \in <^{K'}\) iff \(x < y\) and \(x > 0\) and \(y > 0\), or if \(y = 0\).

All of these functions are computable functions of natural numbers and \(<^{K'}\) is a decidable relation on \(\mathbb{N}\)—but they are not the same functions as successor, addition, and multiplication on \(\mathbb{N}\), and \(<^{K'}\) is not the same relation as < on \(\mathbb{N}\).

This example shows that \(\mathbb{Q}\) has computable non-standard models with domain \(\mathbb{N}\). However, the following result shows
that this is not true for models of PA (and thus also for models of TA).

**Theorem 6.32** (Tennenbaum’s Theorem). \( N \) is the only computable model of PA.

**Summary**

A **model of arithmetic** is a structure for the language \( \mathcal{L}_A \) of arithmetic. There is one distinguished such model, the **standard model** \( N \), with \( |N| = \mathbb{N} \) and interpretations of 0, \( \prime \), +, \( \times \), and < given by 0, the successor, addition, and multiplication functions on \( \mathbb{N} \), and the less-than relation. \( N \) is a model of the theories \( Q \) and PA.

More generally, a structure for \( \mathcal{L}_A \) is called **standard** iff it is isomorphic to \( N \). Two structures are isomorphic if there is an **isomorphism** between them, i.e., a bijective function which preserves the interpretations of constant symbols, function symbols, and predicate symbols. By the **isomorphism theorem**, isomorphic structures are **elementarily equivalent**, i.e., they make the same sentences true. In standard models, the domain is just the set of values of all the numerals \( \bar{n} \).

Models of \( Q \) and \( PA \) that are not isomorphic to \( N \) are called **non-standard**. In non-standard models, the domain is not exhausted by the values of the numerals. An element \( x \in |M| \) where \( x \neq \text{Val}^M(\bar{n}) \) for all \( n \in \mathbb{N} \) is called a **non-standard element** of \( M \). If \( M \models Q \), non-standard elements must obey the axioms of \( Q \), e.g., they have unique successors, they can be added and multiplied, and compared using <. The standard elements of \( M \) are all \( <^M \) all the non-standard elements. Non-standard models exist because of the compactness theorem, and for \( Q \), they can relatively easily be given explicitly. Such models can be used to show that, e.g., \( Q \) is not strong enough to prove certain sentences, e.g., \( Q \not\models \forall x \forall y (x+y) = (y+x) \). This is done by defining a
non-standard \( M \) in which non-standard elements don’t obey the law of commutativity.

Non-standard models of \( PA \) cannot be so easily specified explicitly. By showing that \( PA \) proves certain sentences, we can investigate the structure of the non-standard part of a non-standard model of \( PA \). If a non-standard model \( M \) of \( PA \) is countable, every non-standard element is part of a “block” of non-standard elements which are ordered like \( \mathbb{Z} \) by \( <^M \). These blocks themselves are arranged like \( \mathbb{Q} \), i.e., there is no smallest or largest block, and there is always a block in between any two blocks.

Any countable model is isomorphic to one with domain \( \mathbb{N} \). If the interpretations of \( \cdot, +, \times, \text{ and } < \) in such a model are computable functions, we say it is a \textit{computable model}. The standard model \( \mathbb{N} \) is computable, since the successor, addition, and multiplication functions and the less-than relation on \( \mathbb{N} \) are computable. It is possible to define computable non-standard models of \( \mathbb{Q} \), but \( \mathbb{N} \) is the only computable model of \( PA \). This is \textit{Tanenbaum’s Theorem}.

**Problems**

**Problem 6.1.** Prove Proposition 6.2.

**Problem 6.2.** Carry out the proof of (b) of Theorem 6.6 in detail. Make sure to note where each of the five properties characterizing isomorphisms of Definition 6.5 is used.

**Problem 6.3.** Show that for any structure \( M \), if \( X \) is a definable subset of \( M \), and \( h \) is an automorphism of \( M \), then \( X = \{ h(x) : x \in X \} \) (i.e., \( X \) is fixed under \( h \)).

**Problem 6.4.** Show that the converse of Proposition 6.12 is false, i.e., give an example of a structure \( M \) with \( |M| = \{ \text{Val}^M(\overline{n}) : n \in \mathbb{N} \} \) that is not isomorphic to \( \mathbb{N} \).

**Problem 6.5.** Recall that \( \mathbb{Q} \) contains the axioms

\[
\forall x \forall y (x' = y' \to x = y) \quad (Q_1)
\]
∀x (x ≠ 0 → ∃y x = y') \quad (Q3)

Give structures $M_1, M_2, M_3$ such that

1. $M_1 \models Q_1, M_1 \models Q_2, M_1 \not\models Q_3$;
2. $M_2 \models Q_1, M_2 \not\models Q_2, M_2 \models Q_3$; and
3. $M_3 \not\models Q_1, M_3 \models Q_2, M_3 \models Q_3$;

Obviously, you just have to specify $o^{M_i}$ and $r^{M_i}$ for each.

**Problem 6.6.** Prove that $K$ from Example 6.18 satisfies the remaining axioms of $Q$,

\[
\forall x (x \times 0) = 0 \quad (Q6)
\]
\[
\forall x \forall y (x \times y') = ((x \times y) + x) \quad (Q7)
\]
\[
\forall x \forall y (x < y \leftrightarrow \exists z (x + z' = y)) \quad (Q8)
\]

Find a sentence only involving $r$ true in $N$ but false in $K$.

**Problem 6.7.** Expand $L$ of Example 6.19 to include $\otimes$ and $\ominus$ that interpret $\times$ and $<$.

Show that your structure satisfies the remaining axioms of $Q$,

\[
\forall x (x \times 0) = 0 \quad (Q6)
\]
\[
\forall x \forall y (x \times y') = ((x \times y) + x) \quad (Q7)
\]
\[
\forall x \forall y (x < y \leftrightarrow \exists z (x + z' = y)) \quad (Q8)
\]

**Problem 6.8.** In $L$ of Example 6.19, $a^* = a$ and $b^* = b$. Is there a model of $Q$ in which $a^* = b$ and $b^* = a$?

**Problem 6.9.** Find sentences in $L_A$ derivable in $PA$ (and hence true in $N$) which guarantee the properties of $z$, $\ast$, and $\ominus$ in Proposition 6.21.

**Problem 6.10.** Show that in a non-standard model of $PA$, there is no largest block.
Problem 6.11. Write out a detailed proof of Proposition 6.29. Which sentence must PA derive in order to guarantee the existence of $z$? Why is $x \otimes z$ and $z \otimes y$, and why is $[x] \neq [z]$ and $[z] \neq [y]$?

CHAPTER 7

Second-Order Logic

7.1 Introduction

In first-order logic, we combine the non-logical symbols of a given language, i.e., its constant symbols, function symbols, and predicate symbols, with the logical symbols to express things about first-order structures. This is done using the notion of satisfaction, which relates a structure \( M \), together with a variable assignment \( s \), and a formula \( A \): \( M, s \models A \) holds iff what \( A \) expresses when its constant symbols, function symbols, and predicate symbols are interpreted as \( M \) says, and its free variables are interpreted as \( s \) says, is true. The interpretation of the identity predicate \( = \) is built into the definition of \( M, s \models A \), as is the interpretation of \( \forall \) and \( \exists \). The former is always interpreted as the identity relation on the domain \(|M|\) of the structure, and the quantifiers are always interpreted as ranging over the entire domain. But, crucially, quantification is only allowed over elements of the domain, and so only object variables are allowed to follow a quantifier.

In second-order logic, both the language and the definition of satisfaction are extended to include free and bound function and predicate variables, and quantification over them. These vari-
ables are related to function symbols and predicate symbols the same way that object variables are related to constant symbols. They play the same role in the formation of terms and formulas of second-order logic, and quantification over them is handled in a similar way. In the standard semantics, the second-order quantifiers range over all possible objects of the right type ($n$-place functions from $|M|$ to $|M|$ for function variables, $n$-place relations for predicate variables). For instance, while $\forall v_0 (P^1_0(v_0) \lor \neg P^1_0(v_0))$ is a formula in both first- and second-order logic, in the latter we can also consider $\forall V^1_0 \forall v_0 (V^1_0(v_0) \lor \neg V^1_0(v_0))$ and $\exists V^1_0 \forall v_0 (V^1_0(v_0) \lor \neg V^1_0(v_0))$. Since these contain no free variables, they are sentences of second-order logic. Here, $V^1_0$ is a second-order $1$-place predicate variable. The allowable interpretations of $V^1_0$ are the same that we can assign to a $1$-place predicate symbol like $P^1_0$, i.e., subsets of $|M|$. Quantification over them then amounts to saying that $\forall v_0 (V^1_0(v_0) \lor \neg V^1_0(v_0))$ holds for all ways of assigning a subset of $|M|$ as the value of $V^1_0$, or for at least one. Since every set either contains or fails to contain a given object, both are true in any structure.

Since second-order logic can quantify over subsets of the domain as well as functions, it is to be expected that some amount, at least, of set theory can be carried out in second-order logic. By “carry out,” we mean that it is possible to express set theoretic properties and statements in second-order logic, and is possible without any special, non-logical vocabulary for sets (e.g., the membership predicate symbol of set theory). For instance, we can define unions and intersections of sets and the subset relationship, but also compare the sizes of sets, and state results such as Cantor’s Theorem.

### 7.2 Terms and Formulas

Like in first-order logic, expressions of second-order logic are built up from a basic vocabulary containing variables, constant symbols, predicate symbols and sometimes function symbols. From
them, together with logical connectives, quantifiers, and punctuation symbols such as parentheses and commas, terms and formulas are formed. The difference is that in addition to variables for objects, second-order logic also contains variables for relations and functions, and allows quantification over them. So the logical symbols of second-order logic are those of first-order logic, plus:

1. A countably infinite set of second-order relation variables of every arity $n$: $V_0^n, V_1^n, V_2^n, \ldots$

2. A countably infinite set of second-order function variables: $u_0^n, u_1^n, u_2^n, \ldots$

Just as we use $x, y, z$ as meta-variables for first-order variables $v_i$, we’ll use $X, Y, Z$, etc., as metavariables for $V_i^n$ and $u, v$, etc., as meta-variables for $u_i^n$.

The non-logical symbols of a second-order language are specified the same way a first-order language is: by listing its constant symbols, function symbols, and predicate symbols.

In first-order logic, the identity predicate $=$ is usually included. In first-order logic, the non-logical symbols of a language $\mathcal{L}$ are crucial to allow us to express anything interesting. There are of course sentences that use no non-logical symbols, but with only $=$ it is hard to say anything interesting. In second-order logic, since we have an unlimited supply of relation and function variables, we can say anything we can say in a first-order language even without a special supply of non-logical symbols.

**Definition 7.1** (Second-order Terms). The set of second-order terms of $\mathcal{L}$, $\text{Trm}^2(\mathcal{L})$, is defined by adding to Definition A.4 the clause

1. If $u$ is an $n$-place function variable and $t_1, \ldots, t_n$ are terms, then $u(t_1, \ldots, t_n)$ is a term.

So, a second-order term looks just like a first-order term, except that where a first-order term contains a function symbol $f_i^n$,
a second-order term may contain a function variable $u^n_i$ in its place.

**Definition 7.2** (Second-order formula). The set of second-order formulas $\text{Frm}^2(\mathcal{L})$ of the language $\mathcal{L}$ is defined by adding to Definition A.4 the clauses

1. If $X$ is an $n$-place predicate variable and $t_1, \ldots, t_n$ are second-order terms of $\mathcal{L}$, then $X(t_1, \ldots, t_n)$ is an atomic formula.

2. If $A$ is a formula and $u$ is a function variable, then $\forall u \, A$ is a formula.

3. If $A$ is a formula and $X$ is a predicate variable, then $\forall X \, A$ is a formula.

4. If $A$ is a formula and $u$ is a function variable, then $\exists u \, A$ is a formula.

5. If $A$ is a formula and $X$ is a predicate variable, then $\exists X \, A$ is a formula.

### 7.3 Satisfaction

To define the satisfaction relation $M, s \models A$ for second-order formulas, we have to extend the definitions to cover second-order variables.

**Definition 7.3** (Variable Assignment). A variable assignment $s$ for a structure $M$ is a function which maps each

1. object variable $v_i$ to an element of $|M|$, i.e., $s(v_i) \in |M|$.

2. $n$-place relation variable $V^n_i$ to an $n$-place relation on $|M|$, i.e., $s(V^n_i) \subseteq |M|^n$.
3. \( n \)-place function variable \( u^n_i \) to an \( n \)-place function from \(|M|\) to \(|M|\), i.e., \( s(u^n_i) : |M|^n \rightarrow |M|\);

A structure assigns a value to each constant symbol and function symbol, and a second-order variable assigns objects and functions to each object and function variable. Together, they let us assign a value to very term.

**Definition 7.4** (Value of a Term). If \( t \) is a term of the language \( \mathcal{L} \), \( M \) is a structure for \( \mathcal{L} \), and \( s \) is a variable assignment for \( M \), the **value** \( \text{Val}^M_s(t) \) is defined as for first-order terms, plus the following clause:

\[
t \equiv u(t_1, \ldots, t_n):
\]

\[
\text{Val}^M_s(t) = s(u)(\text{Val}^M_s(t_1), \ldots, \text{Val}^M_s(t_n)).
\]

**Definition 7.5** (Satisfaction). For second-order formulas \( A \), the definition of satisfaction is like **Definition A.23** with the addition of:

1. \( A \equiv X^n t_1, \ldots, t_n : M, s \models A \) iff \( \langle \text{Val}^M_s(t_1), \ldots, \text{Val}^M_s(t_n) \rangle \in s(X^n) \).
2. \( A \equiv \forall X B : M, s \models A \) iff for every \( X \)-variant \( s' \) of \( s \), \( M, s' \models B \).
3. \( A \equiv \exists X B : M, s \models A \) iff there is an \( X \)-variant \( s' \) of \( s \) so that \( M, s' \models B \).
4. \( A \equiv \forall u B : M, s \models A \) iff for every \( u \)-variant \( s' \) of \( s \), \( M, s' \models B \).
5. \( A \equiv \exists u B : M, s \models A \) iff there is an \( u \)-variant \( s' \) of \( s \) so that \( M, s' \models B \).

**Example 7.6.** \( M, s \models \forall z (X z \leftrightarrow Y z) \) whenever \( s(Y) = |M| \setminus s(X) \). So for instance, let \( |M| = \{1, 2, 3\} \), \( s(X) = \{1, 2\} \) and \( s(Y) = \{3\} \).
If there is an \( s' \sim_\mathcal{Y} s \) such that \( M, s \models (\exists Y \mathcal{Y} Y \mathcal{y} \land \forall z (X z \leftrightarrow \neg Y z)) \). And that is the case iff \( s'(Y) \neq \emptyset \) (so that \( M, s' \models \exists y Y \mathcal{y} \mathcal{y} \)) and, as before, \( s'(Y) = |M| \setminus s'(X) \). In other words, \( M, s \models (\exists Y (\exists y Y \mathcal{y} \mathcal{y} \land \forall z (X z \leftrightarrow \neg Y z)) \) iff \( |M| \setminus s(X) \) is non-empty, or, \( s(X) \neq |M| \). So, the formula is satisfied, e.g., if \( s(X) = \{1, 2\} \) but not if \( s(X) = \{1, 2, 3\} \).

### 7.4 Semantic Notions

The central logical notions of validity, entailment, and satisfiability are defined the same way for second-order logic as they are for first-order logic, except that the underlying satisfaction relation is now that for second-order formulas. A second-order sentence, of course, is a formula in which all variables, including predicate and function variables, are bound.

**Definition 7.7** (Validity). A sentence \( A \) is valid, \( \models A \), iff \( M \models A \) for every structure \( M \).

**Definition 7.8** (Entailment). A set of sentences \( \Gamma \) entails a sentence \( A \), \( \Gamma \models A \), iff for every structure \( M \) with \( M \models \Gamma \), \( M \models A \).

**Definition 7.9** (Satisfiability). A set of sentences \( \Gamma \) is satisfiable if \( M \models \Gamma \) for some structure \( M \). If \( \Gamma \) is not satisfiable it is called unsatisfiable.

### 7.5 Expressive Power

Quantification over second-order variables is responsible for an immense increase in the expressive power of the language over that of first-order logic. Second-order existential quantification lets us say that functions or relations with certain properties exist. In first-order logic, the only way to do that is to specify
non-logical symbol (i.e., a function symbol or predicate symbol) for this purpose. Second-order universal quantification lets us say that all subsets of, relations on, or functions from the domain to the domain have a property. In first-order logic, we can only say that the subsets, relations, or functions assigned to one of the non-logical symbols of the language have a property. And when we say that subsets, relations, functions exist that have a property, or that all of them have it, we can use second-order quantification in specifying this property as well. This lets us define relations not definable in first-order logic, and express properties of the domain not expressible in first-order logic.

Example 7.10. If $M$ is a structure for a language $\mathcal{L}$, a relation $R \subseteq |M|^2$ is definable in $\mathcal{L}$ if there is some formula $A_R(v_0, v_1)$ with only the variables $v_0$ and $v_1$ free, such that $R(x, y)$ holds (i.e., $\langle x, y \rangle \in R$) iff $M, s \models A_R(v_0, v_1)$ for $s(v_0) = x$ and $s(v_1) = y$. For instance, in first-order logic we can define the identity relation $\text{Id}_{|M|}$ (i.e., $\{\langle x, x \rangle : x \in |M|\}$) by the formula $v_0 = v_1$. In second-order logic, we can define this relation without $=$. For if $x$ and $y$ are the same element of $|M|$, then they are elements of the same subsets of $|M|$ (since sets are determined by their elements). Conversely, if $x$ and $y$ are different, then they are not elements of the same subsets: e.g., $x \in \{x\}$ but $y \notin \{x\}$ if $x \neq y$. So “being elements of the same subsets of $|M|$” is a relation that holds of $x$ and $y$ iff $x = y$. It is a relation that can be expressed in second-order logic, since we can quantify over all subsets of $|M|$. Hence, the following formula defines $\text{Id}_{|M|}$:

$$\forall X \ (X(v_0) \leftrightarrow X(v_1))$$

Example 7.11. If $R$ is a two-place predicate symbol, $R^M$ is a two-place relation on $|M|$. Its transitive closure $R^*$ is the relation that holds between $x$ and $y$ if for some $z_1, \ldots, z_k$, $R(x, z_1)$, $R(z_1, z_2)$, \ldots, $R(z_k, y)$ holds. This includes the case if $k = 0$, i.e., if $R(x, y)$ holds. This means that $R \subseteq R^*$. In fact, $R^*$ is the smallest relation that includes $R$ and that is transitive. We can say in second-order logic that $X$ is a transitive relation that includes $R$:
\[ B_R(X) \equiv \forall x \forall y (R(x, y) \rightarrow X(x, y)) \land \\
\quad \forall x \forall y \forall z ((X(x, y) \land X(y, z)) \rightarrow X(x, z)) \]

Here, somewhat confusingly, we use \( R \) as the predicate symbol for \( R \). The first conjunct says that \( R \subseteq X \) and the second that \( X \) is transitive.

To say that \( X \) is the smallest such relation is to say that it is itself included in every relation that includes \( R \) and is transitive. So we can define the transitive closure of \( R \) by the formula

\[
R^*(X) \equiv B_R(X) \land \forall Y (B_R(Y) \rightarrow \forall x \forall y (X(x, y) \rightarrow Y(x, y)))
\]

\( M, s \models R^*(X) \) iff \( s(X) = R^* \). The transitive closure of \( R \) cannot be expressed in first-order logic.

### 7.6 Describing Infinite and Countable Domains

A set \( M \) is (Dedekind) infinite iff there is an injective function \( f : M \rightarrow M \) which is not surjective, i.e., with \( \text{dom}(f) \neq M \). In first-order logic, we can consider a one-place function symbol \( f \) and say that the function \( f^M \) assigned to it in a structure \( M \) is injective and \( \text{ran}(f) \neq |M| \):

\[
\forall x \forall y (f(x) = f(y) \rightarrow x = y) \land \exists y \forall x y \neq f(x)
\]

If \( M \) satisfies this sentence, \( f^M : |M| \rightarrow |M| \) is injective, and so \( |M| \) must be infinite. If \( |M| \) is infinite, and hence such a function exists, we can let \( f^M \) be that function and \( M \) will satisfy the sentence. However, this requires that our language contains the non-logical symbol \( f \) we use for this purpose. In second-order logic, we can simply say that such a function exists. This no-longer requires \( f \), and we have the sentence in pure second-order logic

\[
\text{Inf} \equiv \exists u (\forall x \forall y (u(x) = u(y) \rightarrow x = y) \land \exists y \forall x y \neq u(x))
\]

\( M \models \text{Inf} \) iff \( |M| \) is infinite. We can then define \( \text{Fin} \equiv \neg \text{Inf} \); \( M \models \text{Fin} \) iff \( |M| \) is finite. No single sentence of pure first-order logic can
express that the domain is infinite although an infinite set of them can. There is no set of sentences of pure first-order logic that is satisfied in a structure iff its domain is finite.

**Proposition 7.12.** $M \models \text{Inf} \iff |M|$ is infinite.

*Proof.* $M \models \text{Inf} \iff M, s \models \forall x \forall y (u(x) = u(y) \rightarrow x = y) \land \exists y \forall x \ y \neq u(x)$ for some $s$. If it does, $s(u)$ is an injective function, and some $y \in |M|$ is not in the domain of $s(u)$. Conversely, if there is an injective $f : |M| \rightarrow |M|$ with $\text{dom}(f) \neq |M|$, then $s(u) = f$ is such a variable assignment. 

A set $M$ is countable if there is an enumeration

$$m_0, m_1, m_2, \ldots$$

of its elements (without repetitions). Such an enumeration exists iff there is an element $z \in M$ and a function $f : M \rightarrow M$ such that $z, f(z), f(f(z))$ are all the elements of $M$. For if the enumeration exists, $z = m_0$ and $f(m_k) = m_{k+1}$ (or $f(m_k) = m_k$ if $m_k$ is the last element of the enumeration) are the requisite element and function. On the other hand, if such a $z$ and $f$ exist, then $z, f(z), f(f(z)), \ldots$, is an enumeration of $M$, and $M$ is countable. We can express the existence of $z$ and $f$ in second-order logic to produce a sentence true in a structure iff the structure is countable:

$$\text{Count} \equiv \exists z \exists u \forall X ((X(z) \land \forall x (X(x) \rightarrow X(u(x)))) \rightarrow \forall x X(x))$$

**Proposition 7.13.** $M \models \text{Count} \iff |M|$ is countable.

*Proof.* Suppose $|M|$ is enumerable, and let $m_0, m_1, \ldots$, be an enumeration. By removing repetitions we can guarantee that no $m_k$ appears twice. Define $f(m_k) = m_{k+1}$ and let $s(z) = m_0$ and $s(u) = f$. We show that

$$M, s \models \forall X ((X(z) \land \forall x (X(x) \rightarrow X(u(x)))) \rightarrow \forall x X(x))$$
Suppose \( s' \sim_X s \), and \( M = s'(X) \). Suppose further that \( M, s' \models (X(z) \land \forall x (X(x) \to X(u(x)))) \). Then \( s'(z) \in M \) and whenever \( x \in M \), also \( s'(u)(x) \in M \). In other words, since \( s' \sim_X s \), \( m_0 \in M \) and if \( x \in M \) then \( f(x) \in M \), so \( m_0 = f(m_0) \in M \), \( m_2 = f(f(m_0)) \in M \), etc. Thus, \( M = |M| \), and \( M \models \forall x X(x)s' \). Since \( s' \) was an arbitrary \( X \)-variant of \( s \), we are done.

Now assume that \( M, s' \models \forall x ((X(z) \land \forall x (X(x) \to X(u(x)))) \to \forall x X(x)) \) for some \( s \). Let \( m = s(z) \) and \( f = s(u) \) and consider \( M = \{m, f(m), f(f(m)), \ldots \} \). Let \( s' \) be the \( X \)-variant of \( s \) with \( s(X) = M \). Then

\[
M, s' \models ((X(z) \land \forall x (X(x) \to X(u(x)))) \to \forall x X(x))
\]

by assumption. Also, \( M, s' \models X(z) \) since \( s'(X) = M \ni m = s'(z) \), and also \( M, s' \models \forall x (X(x) \to X(u(x))) \) since whenever \( x \in M \) also \( f(x) \in M \). So, since both antecedent and conditional are satisfied, the consequent must also be: \( M, s' \models \forall x X(x) \). But that means that \( M = |M| \), and so \( |M| \) is countable.

\[\square\]

### 7.7 Second-order Arithmetic

Recall that the theory \( \text{PA} \) of Peano arithmetic includes the eight axioms of \( \mathcal{Q} \),

\begin{align*}
\forall x \ x' & \neq 0 \\
\forall x \ \forall y \ (x' = y' \to x = y) \\
\forall x \ \forall y \ (x < y \iff \exists z \ (x + z') = y) \\
\forall x \ (x + 0) & = x \\
\forall x \ \forall y \ (x + y') & = (x + y)' \\
\forall x \ (x \times 0) & = 0 \\
\forall x \ \forall y \ (x \times y') & = ((x \times y) + x)
\end{align*}
plus all sentences of the form

$$(A(0) \land \forall x \ (A(x) \rightarrow A(x'))) \rightarrow \forall x \ A(x)$$

The latter is a “schema,” i.e., a pattern that generates infinitely many sentences of the language of arithmetic, one for each formula $A(x)$. We call this schema the (first-order) axiom schema of induction. In second-order Peano arithmetic $\mathbf{PA}^2$, induction can be stated as a single sentence. $\mathbf{PA}^2$ consists of the first eight axioms above plus the (second-order) induction axiom:

$$\forall X \ (X(0) \land \forall x \ (X(x) \rightarrow X(x'))) \rightarrow \forall x \ X(x)$$

It says that if a subset $X$ of the domain contains $0^M$ and with any $x \in |M|$ also contains $\cdot^M(x)$ (i.e., it is “closed under successor”) it contains everything in the domain (i.e., $X = |M|$).

The induction axiom guarantees that any structure satisfying it contains only those elements of $|M|$ the axioms require to be there, i.e., the values of $\overline{n}$ for $n \in \mathbb{N}$. A model of $\mathbf{PA}^2$ contains no non-standard numbers.

**Theorem 7.14.** If $M \models \mathbf{PA}^2$ then $|M| = \{\text{Val}^n(M) : n \in \mathbb{N}\}$.

**Proof.** Let $N = \{\text{Val}^M(\overline{n}) : n \in \mathbb{N}\}$, and suppose $M \models \mathbf{PA}^2$. Of course, for any $n \in \mathbb{N}$, $\text{Val}^M(\overline{n}) \in |M|$, so $N \subseteq |M|$.

Now for inclusion in the other direction. Consider a variable assignment $s$ with $s(X) = N$. By assumption,

$$M \models \forall X \ (X(0) \land \forall x \ (X(x) \rightarrow X(x'))) \rightarrow \forall x \ X(x), \text{ thus}$$

$$M, s \models (X(0) \land \forall x \ (X(x) \rightarrow X(x'))) \rightarrow \forall x \ X(x).$$

Consider the antecedent of this conditional. $\text{Val}^M(0) \in N$, and so $M, s \models X(0)$. The second conjunct, $\forall x \ (X(x) \rightarrow X(x'))$ is also satisfied. For suppose $x \in N$. By definition of $N$, $x = \text{Val}^M(\overline{n})$ for some $n$. That gives $\cdot^M(x) = \text{Val}^M(\overline{n + 1}) \in N$. So, $\cdot^M(x) \in N$.

We have that $M, s \models X(0) \land \forall x \ (X(x) \rightarrow X(x'))$. Consequently, $M, s \models \forall x \ X(x)$. But that means that for every $x \in |M|$ we have $x \in s(X) = N$. So, $|M| \subseteq N$. \qed
Corollary 7.15. Any two models of $\text{PA}^2$ are isomorphic.

Proof. By Theorem 7.14, the domain of any model of $\text{PA}^2$ is exhausted by $\text{Val}^M(\bar{n})$. Any such model is also a model of $\text{Q}$. By Proposition 6.13, any such model is standard, i.e., isomorphic to $\mathbb{N}$.

Above we defined $\text{PA}^2$ as the theory that contains the first eight arithmetical axioms plus the second-order induction axiom. In fact, thanks to the expressive power of second-order logic, only the first two of the arithmetical axioms plus induction are needed for second-order Peano arithmetic.

Proposition 7.16. Let $\text{PA}^{2\dagger}$ be the second-order theory containing the first two arithmetical axioms (the successor axioms) and the second-order induction axiom. $\succ, +$, and $\times$ are definable in $\text{PA}^{2\dagger}$.

Proof. Exercise.

Corollary 7.17. $M \models \text{PA}^2$ iff $M \models \text{PA}^{2\dagger}$.

Proof. Immediate from Proposition 7.16.

7.8 Second-order Logic is not Axiomatizable

Theorem 7.18. Second-order logic is undecidable.

Proof. A first-order sentence is valid in first-order logic iff it is valid in second-order logic, and first-order logic is undecidable.
Theorem 7.19. There is no sound and complete proof system for second-order logic.

Proof. Let $A$ be a sentence in the language of arithmetic, $N \models A$ iff $\text{PA}^2 \models A$. Let $P$ be the conjunction of the nine axioms of $\text{PA}^2$. $\text{PA}^2 \models A$ iff $\models P \rightarrow A$, i.e., $M \models P \rightarrow A$. Now consider the sentence $\forall z \forall u \forall u' \forall u'' \forall L (P' \rightarrow A')$ resulting by replacing $o$ by $z$, $r$ by the one-place function variable $u$, $+$ and $\times$ by the two-place function-variables $u'$ and $u''$, respectively, and $<$ by the two-place relation variable $L$ and universally quantifying. It is a valid sentence of pure second-order logic iff the original sentence was valid iff $\text{PA}^2 \models A$ iff $N \models A$. Thus if there were a sound and complete proof system for second-order logic, we could use it to define a computable enumeration $f : \mathbb{N} \rightarrow \text{Sent}(L_A)$ of the sentences true in $N$. This function would be representable in $\text{Q}$ by some first-order formula $B_f(x, y)$. Then the formula $\exists x B_f(x, y)$ would define the set of true first-order sentences of $N$, contradicting Tarski’s Theorem. □

7.9 Second-order Logic is not Compact

Call a set of sentences $\Gamma$ finitely satisfiable if every one of its finite subsets is satisfiable. First-order logic has the property that if a set of sentences $\Gamma$ is finitely satisfiable, it is satisfiable. This property is called compactness. It has an equivalent version involving entailment: if $\Gamma \models A$, then already $\Gamma_0 \models A$ for some finite subset $\Gamma_0 \subseteq \Gamma$. In this version it is an immediate corollary of the completeness theorem: for if $\Gamma \models A$, by completeness $\Gamma \vdash A$. But a derivation can only make use of finitely many sentences of $\Gamma$.

Compactness is not true for second-order logic. There are sets of second-order sentences that are finitely satisfiable but not satisfiable, and that entail some $A$ without a finite subset entailing $A$. 
Theorem 7.20. Second-order logic is not compact.

Proof. Recall that

$$\text{Inf} \equiv \exists u \forall x \forall y (u(x) = u(y) \rightarrow x = y)$$

is satisfied in a structure iff its domain is infinite. Let $A^{\geq n}$ be a sentence that asserts that the domain has at least $n$ elements, e.g.,

$$A^{\geq n} \equiv \exists x_1 \ldots \exists x_n (x_1 \neq x_2 \land x_1 \neq x_3 \land \cdots \land x_{n-1} \neq x_n)$$

Consider

$$\Gamma = \{ \neg\text{Inf}, A^{\geq 1}, A^{\geq 2}, A^{\geq 3}, \ldots \}$$

It is finitely satisfiable, since for any finite subset $\Gamma_0$ there is some $k$ so that $A^{\geq k} \in \Gamma$ but no $A^{\geq n} \in \Gamma$ for $n > k$. If $|M|$ has $k$ elements, $M \models \Gamma_0$. But, $\Gamma$ is not satisfiable: if $M \models \neg\text{Inf}$, $|M|$ must be finite, say, of size $k$. Then $M \not\models A^{\geq k+1}$.

7.10 The Löwenheim-Skolem Theorem Fails for Second-order Logic

The (Downward) Löwenheim-Skolem Theorem states that every set of sentences with an infinite model has a countable model. It, too, is a consequence of the completeness theorem: the proof of completeness generates a model for any consistent set of sentences, and that model is countable. There is also an Upward Löwenheim-Skolem Theorem, which guarantees that if a set of sentences has a countably infinite model it also has an uncountable model. Both theorems fail in second-order logic.
Theorem 7.21. The Löwenheim-Skolem Theorem fails for second-order logic: There are sentences with infinite models but no countable models.

Proof. Recall that

\[ \text{Count} \equiv \exists z \, \exists u \, \forall X \, ((X(z) \land \forall x \, (X(x) \rightarrow X(u(x)))) \rightarrow \forall x \, X(x)) \]

is true in a structure \( M \) iff \(|M|\) is countable. So Inf \( \land \neg \text{Count} \) is true in \( M \) iff \(|M|\) is both infinite and not countable. There are such structures—take any uncountable set as the domain, e.g., \( \wp(\mathbb{N}) \) or \( \mathbb{R} \). So Inf \( \land \text{Count} \) has infinite models but no countable models. □

Theorem 7.22. There are sentences with countably infinite but not with uncountable models.

Proof. Count \( \land \text{Inf} \) is true in \( \mathbb{N} \) but not in any structure \( M \) with \(|M|\) uncountable. □

7.11 Comparing Sets

Proposition 7.23. The formula \( \forall x \, (X(x) \rightarrow Y(x)) \) defines the subset relation, i.e., \( M, s \models \forall x \, (X(x) \rightarrow Y(x)) \) iff \( s(X) \subseteq S(y) \).

Proposition 7.24. The formula \( \forall x \, (X(x) \leftrightarrow Y(x)) \) defines the identity relation on sets, i.e., \( M, s \models \forall x \, (X(x) \leftrightarrow Y(x)) \) iff \( s(X) = S(y) \).
A set $X$ is no larger than a set $Y$, $X \preceq Y$, iff there is an injective function $f : X \to Y$. Since we can express that a function is injective, and also that its values for arguments in $X$ are in $Y$, we can also define the relation of being no larger than on subsets of the domain.

Proposition 7.26. The formula

$$\exists u \left( \forall x (X(x) \to Y(u(x))) \land \forall x \forall y (u(x) = u(y) \to x = y) \right)$$

defines the relation of being no larger than.

Two sets are the same size, or “equinumerous,” $X \approx Y$, iff there is a bijective function $f : X \to Y$.

Proposition 7.27. The formula

$$\exists u \left( \forall x (X(x) \to Y(u(x))) \land \forall x \forall y (u(x) = u(y) \to x = y) \land \forall y (Y(y) \to \exists x (X(x) \land y = u(x))) \right)$$

defines the relation of being equinumerous with.

We will abbreviate these formulas, respectively, as $X \subseteq Y$, $X = Y$, $X \neq \emptyset$, $X \preceq Y$, and $X \approx Y$. (This may be slightly confusing, since we use the same notation when we speak informally about sets $X$ and $Y$—but here the notation is an abbreviation for formulas in second-order logic involving one-place relation variables $X$ and $Y$.)
Proposition 7.28. The sentence $\forall X \forall Y ((X \preceq Y \land Y \preceq X) \rightarrow X \approx Y)$ is valid.

Proof. The sentence is satisfied in a structure $M$ if, for any subsets $X \subseteq |X|$ and $Y \subseteq |M|$, if $X \preceq Y$ and $Y \preceq X$ then $X \approx Y$. But this holds for any sets $X$ and $Y$—it is the Schröder-Bernstein Theorem. □

7.12 Cardinalities of Sets

Just as we can express that the domain is finite or infinite, countable or uncountable, we can define the property of a subset of $|M|$ being finite or infinite, countable or uncountable.

Proposition 7.29. The formula $\text{Inf}(X) \equiv$

$\exists u (\forall x \forall y (u(x) = u(y) \rightarrow x = y) \land$

$\exists y (X(y) \land \forall x (X(x) \rightarrow y \neq u(x)))$

is satisfied with respect to a variable assignment $s$ iff $s(X)$ is infinite.

Proposition 7.30. The formula $\text{Count}(X) \equiv$

$\exists z \exists u (X(z) \land \forall x (X(x) \rightarrow X(u(x)))) \land$

$\forall Y ((Y(z) \land \forall x (Y(x) \rightarrow Y(u(x)))) \rightarrow X = Y))$

is satisfied with respect to a variable assignment $s$ iff $s(X)$ is countable.

We know from Cantor’s Theorem that there are uncountable sets, and in fact, that there are infinitely many different levels of infinite sizes. Set theory develops an entire arithmetic of sizes of sets, and assigns infinite cardinal numbers to sets. The natural numbers serve as the cardinal numbers measuring the sizes of finite sets. The cardinality of countably infinite sets is the first infinite cardinality, called $\aleph_0$ (“aleph-nought” or “aleph-zero”).
The next infinite size is $\aleph_1$. It is the smallest size a set can be without being countable (i.e., of size $\aleph_0$). We can define “$X$ has size $\aleph_0$” as $\text{Aleph}_0(X) \leftrightarrow \text{Inf}(X) \land \text{Count}(X)$. $X$ has size $\aleph_1$ iff all its subsets are finite or have size $\aleph_0$, but is not itself of size $\aleph_0$. Hence we can express this by the formula $\text{Aleph}_1(X) \equiv \forall Y (Y \subseteq X \rightarrow (\neg \text{Inf}(Y) \lor \text{Aleph}_0(Y))) \land \neg \text{Aleph}_0(X)$. Being of size $\aleph_2$ is defined similarly, etc.

There is one size of special interest, the so-called cardinality of the continuum. It is the size of $\varphi(\mathbb{N})$, or, equivalently, the size of $\mathbb{R}$. That a set is the size of the continuum can also be expressed in second-order logic, but requires a bit more work.

### 7.13 The Power of the Continuum

In second-order logic we can quantify over subsets of the domain, but not over sets of subsets of the domain. To do this directly, we would need third-order logic. For instance, if we wanted to state Cantor’s Theorem that there is no injective function from the power set of a set to the set itself, we might try to formulate it as “for every set $X$, and every set $P$, if $P$ is the power set of $X$, then not $P \preceq X$. And to say that $P$ is the power set of $X$ would require formalizing that the elements of $P$ are all and only the subsets of $X$, so something like $\forall Y (P(Y) \leftrightarrow Y \subseteq X)$. The problem lies in $P(Y)$: that is not a formula of second-order logic, since only terms can be arguments to one-place relation variables like $P$.

We can, however, simulate quantification over sets of sets, if the domain is large enough. The idea is to make use of the fact that two-place relations $R$ relates elements of the domain to elements of the domain. Given such an $R$, we can collect all the elements to which some $x$ is $R$-related: $\{y \in |M| : R(x, y)\}$ is the set “coded by” $x$. Converseley, if $Z \subseteq \varphi(|M|)$ is some collection of subsets of $|M|$, and there are at least as many elements of $|M|$ as there are sets in $Z$, then there is also a relation $R \subseteq |M|^2$ such that every $Y \in Z$ is coded by some $x$ using $R$. 
**Definition 7.31.** If \( R \subseteq |M|^2 \), then \( x \ R\)-codes \( \{y \in |M| : R(x, y)\} \). \( Y \ R\)-codes \( \varphi(X) \) iff for every \( Z \subseteq X \), some \( x \in Y \ R\)-codes \( Y \), and every \( x \in Y \ R\)-codes some \( Y \in Z \).

**Proposition 7.32.** The formula
\[
\text{Codes}(x, R, Y) \equiv \forall y (Y(y) \leftrightarrow R(x, y))
\]
expresses that \( s(x) \ s(R)\)-codes \( s(Y) \). The formula
\[
\text{Pow}(Y, R, X) \equiv \\
\forall Z (Z \subseteq X \rightarrow \exists x (Y(x) \land \text{Codes}(x, R, Z))) \land \\
\forall x (Y(x) \rightarrow \forall Z (\text{Codes}(x, R, Z) \rightarrow Z \subseteq X))
\]
expresses that \( s(Y) \ s(R)\)-codes the power set of \( s(X) \).

With this trick, we can express statements about the power set by quantifying over the codes of subsets rather than the subsets themselves. For instance, Cantor’s Theorem can now be expressed by saying that there is no injective function from the domain of any relation that codes the power set of \( X \) to \( X \) itself.

**Proposition 7.33.** The sentence
\[
\forall X \forall R (\text{Pow}(R, X) \rightarrow \\
\neg \exists u (\forall x \forall y (u(x) = u(y) \rightarrow x = y) \land \\
\forall Y (\text{Codes}(x, R, Y) \rightarrow X(u(x))))))
\]
is valid.

The power set of a countably infinite set is uncountable, and so its cardinality is larger than that of any countably infinite set (which is \( \aleph_0 \)). The size of \( \varphi(\mathbb{R}) \) is called the “power of the continuum,” since it is the same size as the points on the real number
line, \( \mathbb{R} \). If the domain is large enough to code the power set of a countably infinite set, we can express that a set is the size of the continuum by saying that it is equinumerous with any set \( Y \) that codes the power set of set \( X \) of size \( \aleph_0 \). (If the domain is not large enough, i.e., it contains no subset equinumerous with \( \mathbb{R} \), then there can also be no relation that codes \( \wp(X) \).)

**Proposition 7.34.** If \( \mathbb{R} \leq |M| \), then the formula

\[
\text{Cont}(X) \equiv \forall X \forall Y \forall R ((\aleph_0(X) \land \text{Pow}(Y, R, X)) \rightarrow \neg Y \leq X)
\]

expresses that \( s(X) \approx \mathbb{R} \).

**Proposition 7.35.** \( |M| \approx \mathbb{R} \) iff

\[
M \models \exists X \exists Y \exists R (\aleph_0(X) \land \text{Pow}(Y, R, X) \land \exists u (\forall x \forall y (u(x) = u(y) \rightarrow x = y) \land \forall y (Y(y) \rightarrow \exists x y = u(x))))
\]

The Continuum Hypothesis is the statement that the size of the continuum is the first uncountable cardinality, i.e, that \( \wp(\mathbb{N}) \) has size \( \aleph_1 \).

**Proposition 7.36.** The Continuum Hypothesis is true iff

\[
\text{CH} \equiv \forall X (\aleph_1(X) \leftrightarrow \text{Cont}(x))
\]

is valid.

Note that it isn’t true that \( \neg \text{CH} \) is valid iff the Continuum Hypothesis is false. In a countable domain, there are no subsets of size \( \aleph_1 \) and also no subsets of the size of the continuum, so \( \text{CH} \) is always true in a countable domain. However, we can give a different sentence that is valid iff the Continuum Hypothesis is false:
Proposition 7.37. The Continuum Hypothesis is false iff

\[ \text{NCH} \equiv \forall X (\text{Cont}(X) \rightarrow \exists Y (Y \subseteq X \land \neg \text{Count}(X) \land \neg X \approx Y)) \]

is valid.

Summary

Second-order logic is an extension of first-order logic by variables for relations and functions, which can be quantified. Structures for second-order logic are just like first-order structures and give the interpretations of all non-logical symbols of the language. Variable assignments, however, also assign relations and functions on the domain to the second-order variables. The satisfaction relation is defined for second-order formulas just like in the first-order case, but extended to deal with second-order variables and quantifiers.

Second-order quantifiers make second-order logic more expressive than first-order logic. For instance, the identity relation on the domain of a structure can be defined without \(=\), by \(\forall X (X(x) \leftrightarrow X(y))\). Second-order logic can express the transitive closure of a relation, which is not expressible in first-order logic. Second-order quantifiers can also express properties of the domain, that it is finite or infinite, countable or uncountable. This means that, e.g., there is a second-order sentence Inf such that \(M \models \text{Inf} \iff |M|\) is infinite. Importantly, these are pure second-order sentences, i.e., they contain no non-logical symbols. Because of the compactness and Löwenheim-Skolem theorems, there are no first-order sentences that have these properties. It also shows that the compactness and Löwenheim-Skolem theorems fail for second-order logic.

Second-order quantification also makes it possible to replace first-order schemas by single sentences. For instance, second-order arithmetic \(\text{PA}^2\) is comprised of the axioms of \(\text{Q}\) plus the
single **induction axiom**

\[ \forall X \ ((X(0) \land \forall x \ (X(x) \rightarrow X(x')))) \rightarrow \forall x \ X(x) \].

In contrast to first-order \( \mathbf{PA} \), all second-order models of \( \mathbf{PA}^2 \) are isomorphic to the standard model. In other words, \( \mathbf{PA}^2 \) has **no non-standard models**.

Since second-order logic includes first-order logic, it is undecidable. First-order logic is at least axiomatizable, i.e., it has a sound and complete proof system. Second-order logic does not, it **is not axiomatizable**. Thus, the set of validities of second-order logic is highly non-computable. In fact, pure second-order logic can express set-theoretic claims like the **continuum hypothesis**, which are independent of set theory.

**Problems**

**Problem 7.1.** Show that \( \forall X \ (X(v_0) \rightarrow X(v_1)) \) (note: \( \rightarrow \) not \( \leftrightarrow \)) defines \( \text{Id}_{[M]} \).

**Problem 7.2.** The sentence \( \text{Inf} \land \text{Count} \) is true in all and only countably infinite domains. Adjust the definition of Count so that it becomes a different sentence that directly expresses that the domain is countably infinite, and prove that it does.

**Problem 7.3.** Prove Proposition 7.16.

**Problem 7.4.** Give an example of a set \( \Gamma \) and a sentence \( A \) so that \( \Gamma \models A \) but for every finite subset \( \Gamma_0 \subseteq \Gamma \), \( \Gamma_0 \not\models A \).
Expressions of first-order logic are built up from a basic vocabulary containing variables, constant symbols, predicate symbols and sometimes function symbols. From them, together with logical connectives, quantifiers, and punctuation symbols such as parentheses and commas, terms and formulas are formed.

Informally, predicate symbols are names for properties and relations, constant symbols are names for individual objects, and function symbols are names for mappings. These, except for the identity predicate =, are the non-logical symbols and together make up a language. Any first-order language $L$ is determined by its non-logical symbols. In the most general case, $L$ contains infinitely many symbols of each kind.

In the general case, we make use of the following symbols in first-order logic:

1. Logical symbols
   
   a) Logical connectives: $\neg$ (negation), $\land$ (conjunction),
A.1. FIRST-ORDER LANGUAGES

1. Logical symbols, making up the standard language of first-order logic

a) \( \lor \) (disjunction), \( \to \) (conditional), \( \forall \) (universal quantifier), \( \exists \) (existential quantifier).

b) The propositional constant for falsity \( \perp \).
c) The two-place identity predicate \( = \).
d) A countably infinite set of variables: \( \nu_0, \nu_1, \nu_2, \ldots \)

2. Non-logical symbols, making up the standard language of first-order logic

a) A countably infinite set of \( n \)-place predicate symbols for each \( n > 0 \): \( A^n_0, A^n_1, A^n_2, \ldots \)
b) A countably infinite set of constant symbols: \( c_0, c_1, c_2, \ldots \)
c) A countably infinite set of \( n \)-place function symbols for each \( n > 0 \): \( f^n_0, f^n_1, f^n_2, \ldots \)

3. Punctuation marks: (, ), and the comma.

Most of our definitions and results will be formulated for the full standard language of first-order logic. However, depending on the application, we may also restrict the language to only a few predicate symbols, constant symbols, and function symbols.

**Example A.1.** The language \( \mathcal{L}_A \) of arithmetic contains a single two-place predicate symbol \(<\), a single constant symbol \( 0 \), one one-place function symbol \( \prime \), and two two-place function symbols \(+\) and \(\times\).

**Example A.2.** The language of set theory \( \mathcal{L}_Z \) contains only the single two-place predicate symbol \(\in\).

**Example A.3.** The language of orders \( \mathcal{L}_\leq \) contains only the two-place predicate symbol \(\leq\).

Again, these are conventions: officially, these are just aliases, e.g., \(<\), \(\in\), and \(\leq\) are aliases for \( A^2_0 \), \(0\) for \( c_0 \), \(\prime\) for \( f^1_0 \), \(+\) for \( f^2_0 \), \(\times\) for \( f^2_1 \).
In addition to the primitive connectives and quantifiers introduced above, we also use the following defined symbols: \( \leftrightarrow \) (biconditional), truth \( \top \)

A defined symbol is not officially part of the language, but is introduced as an informal abbreviation: it allows us to abbreviate formulas which would, if we only used primitive symbols, get quite long. This is obviously an advantage. The bigger advantage, however, is that proofs become shorter. If a symbol is primitive, it has to be treated separately in proofs. The more primitive symbols, therefore, the longer our proofs.

We might treat all the propositional operators and both quantifiers as primitive symbols of the language. We might instead choose a smaller stock of primitive symbols and treat the other logical operators as defined. “Truth functionally complete” sets of Boolean operators include \( \{ \neg, \lor \} \), \( \{ \neg, \land \} \), and \( \{ \neg, \rightarrow \} \)—these can be combined with either quantifier for an expressively complete first-order language.

You may be familiar with two other logical operators: the Sheffer stroke \( | \) (named after Henry Sheffer), and Peirce’s arrow \( \downarrow \), also known as Quine’s dagger. When given their usual readings of “nand” and “nor” (respectively), these operators are truth functionally complete by themselves.

### A.2 Terms and Formulas

Once a first-order language \( \mathcal{L} \) is given, we can define expressions built up from the basic vocabulary of \( \mathcal{L} \). These include in particular terms and formulas.

**Definition A.4** (Terms). The set of terms \( \text{Trm}(\mathcal{L}) \) of \( \mathcal{L} \) is defined inductively by:

1. Every variable is a term.
2. Every constant symbol of \( \mathcal{L} \) is a term.
A.2. TERMS AND FORMULAS

3. If $f$ is an $n$-place function symbol and $t_1, \ldots, t_n$ are terms, then $f(t_1, \ldots, t_n)$ is a term.

4. Nothing else is a term.

A term containing no variables is a **closed term**.

The constant symbols appear in our specification of the language and the terms as a separate category of symbols, but they could instead have been included as zero-place function symbols. We could then do without the second clause in the definition of terms. We just have to understand $f(t_1, \ldots, t_n)$ as just $f$ by itself if $n = 0$.

**Definition A.5** (Formula). The set of formulas $\text{Frm}(\mathcal{L})$ of the language $\mathcal{L}$ is defined inductively as follows:

1. $\bot$ is an atomic formula.

2. If $R$ is an $n$-place predicate symbol of $\mathcal{L}$ and $t_1, \ldots, t_n$ are terms of $\mathcal{L}$, then $R(t_1, \ldots, t_n)$ is an atomic formula.

3. If $t_1$ and $t_2$ are terms of $\mathcal{L}$, then $= (t_1, t_2)$ is an atomic formula.

4. If $A$ is a formula, then $\neg A$ is formula.

5. If $A$ and $B$ are formulas, then $(A \land B)$ is a formula.

6. If $A$ and $B$ are formulas, then $(A \lor B)$ is a formula.

7. If $A$ and $B$ are formulas, then $(A \rightarrow B)$ is a formula.

8. If $A$ is a formula and $x$ is a variable, then $\forall x \ A$ is a formula.

9. If $A$ is a formula and $x$ is a variable, then $\exists x \ A$ is a formula.

10. Nothing else is a formula.

The definitions of the set of terms and that of formulas are
inductive definitions. Essentially, we construct the set of formulas in infinitely many stages. In the initial stage, we pronounce all atomic formulas to be formulas; this corresponds to the first few cases of the definition, i.e., the cases for \( \bot \), \( R(t_1, \ldots, t_n) \) and \( = (t_1, t_2) \). “Atomic formula” thus means any formula of this form.

The other cases of the definition give rules for constructing new formulas out of formulas already constructed. At the second stage, we can use them to construct formulas out of atomic formulas. At the third stage, we construct new formulas from the atomic formulas and those obtained in the second stage, and so on. A formula is anything that is eventually constructed at such a stage, and nothing else.

By convention, we write \( = \) between its arguments and leave out the parentheses: \( t_1 = t_2 \) is an abbreviation for \( = (t_1, t_2) \). Moreover, \( = (t_1, t_2) \) is abbreviated as \( t_1 \neq t_2 \). When writing a formula \( (B * C) \) constructed from \( B, C \) using a two-place connective \*, we will often leave out the outermost pair of parentheses and write simply \( B * C \).

**Definition A.6.** Formulas constructed using the defined operators are to be understood as follows:

1. \( \top \) abbreviates \( \neg \bot \).
2. \( A \leftrightarrow B \) abbreviates \( (A \rightarrow B) \land (B \rightarrow A) \).

If we work in a language for a specific application, we will often write two-place predicate symbols and function symbols between the respective terms, e.g., \( t_1 < t_2 \) and \( (t_1 + t_2) \) in the language of arithmetic and \( t_1 \in t_2 \) in the language of set theory. The successor function in the language of arithmetic is even written conventionally after its argument: \( t' \). Officially, however, these are just conventional abbreviations for \( A^2_0(t_1, t_2) \), \( f^2_0(t_1, t_2) \), \( A^2_0(t_1, t_2) \) and \( f^1_0(t) \), respectively.
Definition A.7 (Syntactic identity). The symbol $\equiv$ expresses syntactic identity between strings of symbols, i.e., $A \equiv B$ iff $A$ and $B$ are strings of symbols of the same length and which contain the same symbol in each place.

The $\equiv$ symbol may be flanked by strings obtained by concatenation, e.g., $A \equiv (B \lor C)$ means: the string of symbols $A$ is the same string as the one obtained by concatenating an opening parenthesis, the string $B$, the $\lor$ symbol, the string $C$, and a closing parenthesis, in this order. If this is the case, then we know that the first symbol of $A$ is an opening parenthesis, $A$ contains $B$ as a substring (starting at the second symbol), that substring is followed by $\lor$, etc.

A.3 Free Variables and Sentences

Definition A.8 (Free occurrences of a variable). The free occurrences of a variable in a formula are defined inductively as follows:

1. $A$ is atomic: all variable occurrences in $A$ are free.
2. $A \equiv \neg B$: the free variable occurrences of $A$ are exactly those of $B$.
3. $A \equiv (B \ast C)$: the free variable occurrences of $A$ are those in $B$ together with those in $C$.
4. $A \equiv \forall x B$: the free variable occurrences in $A$ are all of those in $B$ except for occurrences of $x$.
5. $A \equiv \exists x B$: the free variable occurrences in $A$ are all of those in $B$ except for occurrences of $x$. 
**Definition A.9** (Bound Variables). An occurrence of a variable in a formula \( A \) is *bound* if it is not free.

**Definition A.10** (Scope). If \( \forall x B \) is an occurrence of a subformula in a formula \( A \), then the corresponding occurrence of \( B \) in \( A \) is called the *scope* of the corresponding occurrence of \( \forall x \). Similarly for \( \exists x \).

If \( B \) is the scope of a quantifier occurrence \( \forall x \) or \( \exists x \) in \( A \), then all occurrences of \( x \) which are free in \( B \) are said to be *bound by* the mentioned quantifier occurrence.

**Example A.11.** Consider the following formula:

\[
\exists v_0 \underbrace{A^2_0(v_0, v_1)}_B
\]

\( B \) represents the scope of \( \exists v_0 \). The quantifier binds the occurrence of \( v_0 \) in \( B \), but does not bind the occurrence of \( v_1 \). So \( v_1 \) is a free variable in this case.

We can now see how this might work in a more complicated formula \( A \):

\[
\forall v_0 \underbrace{(A^1_0(v_0) \rightarrow A^2_0(v_0, v_1))}_B \rightarrow \exists v_1 \underbrace{(A^2_1(v_0, v_1) \lor \forall v_0 \neg A^1_1(v_0))}_C
\]

\( B \) is the scope of the first \( \forall v_0 \), \( C \) is the scope of \( \exists v_1 \), and \( D \) is the scope of the second \( \forall v_0 \). The first \( \forall v_0 \) binds the occurrences of \( v_0 \) in \( B \), \( \exists v_1 \) the occurrence of \( v_1 \) in \( C \), and the second \( \forall v_0 \) binds the occurrence of \( v_0 \) in \( D \). The first occurrence of \( v_1 \) and the fourth occurrence of \( v_0 \) are free in \( A \). The last occurrence of \( v_0 \) is free in \( D \), but bound in \( C \) and \( A \).
**Definition A.12** (Sentence). A formula $A$ is a *sentence* iff it contains no free occurrences of variables.

### A.4 Substitution

**Definition A.13** (Substitution in a term). We define $s[t/x]$, the result of *substituting* $t$ for every occurrence of $x$ in $s$, recursively:

1. $s \equiv c$: $s[t/x]$ is just $s$.
2. $s \equiv y$: $s[t/x]$ is also just $s$, provided $y$ is a variable and $y \not\equiv x$.
3. $s \equiv x$: $s[t/x]$ is $t$.
4. $s \equiv f(t_1, \ldots, t_n)$: $s[t/x]$ is $f(t_1[t/x], \ldots, t_n[t/x])$.

**Definition A.14.** A term $t$ is *free for* $x$ in $A$ if none of the free occurrences of $x$ in $A$ occur in the scope of a quantifier that binds a variable in $t$.

**Example A.15.**

1. $v_8$ is free for $v_1$ in $\exists v_3 A_4^2(v_3, v_1)$
2. $r_1^2(v_1, v_2)$ is *not* free for $v_0$ in $\forall v_2 A_4^2(v_0, v_2)$

**Definition A.16** (Substitution in a formula). If $A$ is a formula, $x$ is a variable, and $t$ is a term free for $x$ in $A$, then $A[t/x]$ is the result of substituting $t$ for all free occurrences of $x$ in $A$.

1. $A \equiv \bot$: $A[t/x]$ is $\bot$.
2. $A \equiv P(t_1, \ldots, t_n)$: $A[t/x]$ is $P(t_1[t/x], \ldots, t_n[t/x])$. 
3. \( A \equiv t_1 = t_2: \ A[t/x] \text{ is } t_1[t/x] = t_2[t/x]. \)

4. \( A \equiv \neg B: \ A[t/x] \text{ is } \neg B[t/x]. \)

5. \( A \equiv (B \land C): \ A[t/x] \text{ is } (B[t/x] \land C[t/x]). \)

6. \( A \equiv (B \lor C): \ A[t/x] \text{ is } (B[t/x] \lor C[t/x]). \)

7. \( A \equiv (B \rightarrow C): \ A[t/x] \text{ is } (B[t/x] \rightarrow C[t/x]). \)

8. \( A \equiv \forall y B: \ A[t/x] \text{ is } \forall y B[t/x], \) provided \( y \) is a variable other than \( x; \) otherwise \( A[t/x] \text{ is just } A. \)

9. \( A \equiv \exists y B: \ A[t/x] \text{ is } \exists y B[t/x], \) provided \( y \) is a variable other than \( x; \) otherwise \( A[t/x] \text{ is just } A. \)

Note that substitution may be vacuous: If \( x \) does not occur in \( A \) at all, then \( A[t/x] \) is just \( A. \)

The restriction that \( t \) must be free for \( x \) in \( A \) is necessary to exclude cases like the following. If \( A \equiv \exists y x < y \) and \( t \equiv y, \) then \( A[t/x] \) would be \( \exists y y < y. \) In this case the free variable \( y \) is “captured” by the quantifier \( \exists y \) upon substitution, and that is undesirable. For instance, we would like it to be the case that whenever \( \forall x B \) holds, so does \( B[t/x]. \) But consider \( \forall x \exists y x < y \) (here \( B \) is \( \exists y x < y). \) It is sentence that is true about, e.g., the natural numbers: for every number \( x \) there is a number \( y \) greater than it. If we allowed \( y \) as a possible substitution for \( x, \) we would end up with \( B[y/x] \equiv \exists y y < y, \) which is false. We prevent this by requiring that none of the free variables in \( t \) would end up being bound by a quantifier in \( A. \)

We often use the following convention to avoid cumbersome notation: If \( A \) is a formula with a free variable \( x, \) we write \( A(x) \) to indicate this. When it is clear which \( A \) and \( x \) we have in mind, and \( t \) is a term (assumed to be free for \( x \) in \( A(x) \)), then we write \( A(t) \) as short for \( A(x)[t/x]. \)
A.5 Structures for First-order Languages

First-order languages are, by themselves, uninterpreted: the constant symbols, function symbols, and predicate symbols have no specific meaning attached to them. Meanings are given by specifying a structure. It specifies the domain, i.e., the objects which the constant symbols pick out, the function symbols operate on, and the quantifiers range over. In addition, it specifies which constant symbols pick out which objects, how a function symbol maps objects to objects, and which objects the predicate symbols apply to. Structures are the basis for semantic notions in logic, e.g., the notion of consequence, validity, satisfiability. They are variously called “structures,” “interpretations,” or “models” in the literature.

Definition A.17 (Structures). A structure $M$, for a language $\mathcal{L}$ of first-order logic consists of the following elements:

1. **Domain**: a non-empty set, $|M|$

2. **Interpretation of constant symbols**: for each constant symbol $c$ of $\mathcal{L}$, an element $c^M \in |M|$

3. **Interpretation of predicate symbols**: for each $n$-place predicate symbol $R$ of $\mathcal{L}$ (other than $=$), an $n$-place relation $R^M \subseteq |M|^n$

4. **Interpretation of function symbols**: for each $n$-place function symbol $f$ of $\mathcal{L}$, an $n$-place function $f^M : |M|^n \rightarrow |M|$

Example A.18. A structure $M$ for the language of arithmetic consists of a set, an element of $|M|$, $0^M$, as interpretation of the constant symbol $0$, a one-place function $\cdot^M : |M| \rightarrow |M|$, two two-place functions $+^M$ and $\times^M$, both $|M|^2 \rightarrow |M|$, and a two-place relation $<^M \subseteq |M|^2$.

An obvious example of such a structure is the following:

1. $|N| = \mathbb{N}$
2. \(0^N = 0\)

3. \(\mathbb{N}(n) = n + 1\) for all \(n \in \mathbb{N}\)

4. \(\mathbb{N}(n, m) = n + m\) for all \(n, m \in \mathbb{N}\)

5. \(\times^N(n, m) = n \cdot m\) for all \(n, m \in \mathbb{N}\)

6. \(<^N = \{\langle n, m \rangle : n \in \mathbb{N}, m \in \mathbb{N}, n < m\}\)

The structure \(N\) for \(\mathcal{L}_A\) so defined is called the standard model of arithmetic, because it interprets the non-logical constants of \(\mathcal{L}_A\) exactly how you would expect.

However, there are many other possible structures for \(\mathcal{L}_A\). For instance, we might take as the domain the set \(\mathbb{Z}\) of integers instead of \(\mathbb{N}\), and define the interpretations of \(0, \prime, +, \times, <\) accordingly. But we can also define structures for \(\mathcal{L}_A\) which have nothing even remotely to do with numbers.

**Example A.19.** A structure \(M\) for the language \(\mathcal{L}_Z\) of set theory requires just a set and a single-two place relation. So technically, e.g., the set of people plus the relation “\(x\) is older than \(y\)” could be used as a structure for \(\mathcal{L}_Z\), as well as \(\mathbb{N}\) together with \(n \geq m\) for \(n, m \in \mathbb{N}\).

A particularly interesting structure for \(\mathcal{L}_Z\) in which the elements of the domain are actually sets, and the interpretation of \(\in\) actually is the relation “\(x\) is an element of \(y\)” is the structure \(HF\) of hereditarily finite sets:

1. \(|HF| = \emptyset \cup \varphi(\emptyset) \cup \varphi(\varphi(\emptyset)) \cup \varphi(\varphi(\varphi(\emptyset))) \cup \ldots\); 

2. \(\in^{HF} = \{\langle x, y \rangle : x, y \in |HF|, x \in y\}\).

The stipulations we make as to what counts as a structure impact our logic. For example, the choice to prevent empty domains ensures, given the usual account of satisfaction (or truth) for quantified sentences, that \(\exists x (\mathcal{A}(x) \lor \neg \mathcal{A}(x))\) is valid—that is, a logical truth. And the stipulation that all constant symbols must
refer to an object in the domain ensures that the existential generalization is a sound pattern of inference: $A(a)$, therefore $\exists x A(x)$. If we allowed names to refer outside the domain, or to not refer, then we would be on our way to a free logic, in which existential generalization requires an additional premise: $A(a)$ and $\exists x x = a$, therefore $\exists x A(x)$.

A.6  Satisfaction of a Formula in a Structure

The basic notion that relates expressions such as terms and formulas, on the one hand, and structures on the other, are those of value of a term and satisfaction of a formula. Informally, the value of a term is an element of a structure—if the term is just a constant, its value is the object assigned to the constant by the structure, and if it is built up using function symbols, the value is computed from the values of constants and the functions assigned to the functions in the term. A formula is satisfied in a structure if the interpretation given to the predicates makes the formula true in the domain of the structure. This notion of satisfaction is specified inductively: the specification of the structure directly states when atomic formulas are satisfied, and we define when a complex formula is satisfied depending on the main connective or quantifier and whether or not the immediate subformulas are satisfied. The case of the quantifiers here is a bit tricky, as the immediate subformula of a quantified formula has a free variable, and structures don’t specify the values of variables. In order to deal with this difficulty, we also introduce variable assignments and define satisfaction not with respect to a structure alone, but with respect to a structure plus a variable assignment.
**Definition A.20** (Variable Assignment). A *variable assignment* $s$ for a structure $M$ is a function which maps each variable to an element of $|M|$, i.e., $s : \text{Var} \rightarrow |M|$.

A structure assigns a value to each constant symbol, and a variable assignment to each variable. But we want to use terms built up from them to also name elements of the domain. For this we define the value of terms inductively. For constant symbols and variables the value is just as the structure or the variable assignment specifies it; for more complex terms it is computed recursively using the functions the structure assigns to the function symbols.

**Definition A.21** (Value of Terms). If $t$ is a term of the language $\mathcal{L}$, $M$ is a structure for $\mathcal{L}$, and $s$ is a variable assignment for $M$, the *value* $\text{Val}_{s}^{M}(t)$ is defined as follows:

1. $t \equiv c$: $\text{Val}_{s}^{M}(t) = c^{M}$.
2. $t \equiv x$: $\text{Val}_{s}^{M}(t) = s(x)$.
3. $t \equiv f(t_{1}, \ldots, t_{n})$:

   $$\text{Val}_{s}^{M}(t) = f^{M}(\text{Val}_{s}^{M}(t_{1}), \ldots, \text{Val}_{s}^{M}(t_{n})).$$

**Definition A.22** ($x$-Variant). If $s$ is a variable assignment for a structure $M$, then any variable assignment $s'$ for $M$ which differs from $s$ at most in what it assigns to $x$ is called an *$x$-variant* of $s$. If $s'$ is an $x$-variant of $s$ we write $s \sim_{x} s'$.

Note that an $x$-variant of an assignment $s$ does not *have* to assign something different to $x$. In fact, every assignment counts as an $x$-variant of itself.
Definition A.23 (Satisfaction). Satisfaction of a formula $A$ in a structure $M$ relative to a variable assignment $s$, in symbols: $M, s \vDash A$, is defined recursively as follows. (We write $M, s \not\vDash A$ to mean “not $M, s \vDash A$.”)

1. $A \equiv \bot$: $M, s \not\vDash A$.

2. $A \equiv R(t_1, \ldots, t_n)$: $M, s \vDash A$ iff $\langle \text{Val}_s^M(t_1), \ldots, \text{Val}_s^M(t_n) \rangle \in R^M$.

3. $A \equiv t_1 = t_2$: $M, s \vDash A$ iff $\text{Val}_s^M(t_1) = \text{Val}_s^M(t_2)$.

4. $A \equiv \neg B$: $M, s \vDash A$ iff $M, s \not\vDash B$.

5. $A \equiv (B \land C)$: $M, s \vDash A$ iff $M, s \vDash B$ and $M, s \vDash C$.

6. $A \equiv (B \lor C)$: $M, s \vDash A$ iff $M, s \not\vDash B$ or $M, s \vDash B$ (or both).

7. $A \equiv (B \rightarrow C)$: $M, s \vDash A$ iff $M, s \not\vDash B$ or $M, s \vDash C$ (or both).

8. $A \equiv \forall x B$: $M, s \vDash A$ iff for every $x$-variant $s'$ of $s$, $M, s' \vDash B$.

9. $A \equiv \exists x B$: $M, s \vDash A$ iff there is an $x$-variant $s'$ of $s$ so that $M, s' \vDash B$.

The variable assignments are important in the last two clauses. We cannot define satisfaction of $\forall x B(x)$ by “for all $a \in \lvert M \rvert$, $M \vDash B(a)$.” We cannot define satisfaction of $\exists x B(x)$ by “for at least one $a \in \lvert M \rvert$, $M \vDash B(a)$.” The reason is that $a$ is not symbol of the language, and so $B(a)$ is not a formula (that is, $B[a/x]$ is undefined). We also cannot assume that we have constant symbols or terms available that name every element of $M$, since there is nothing in the definition of structures that requires it. Even in the standard language the set of constant symbols is countably infinite, so if $\lvert M \rvert$ is not countable there aren’t even enough constant symbols to name every object.

Example A.24. Let $\{a, b, f, R\}$ where $a$ and $b$ are constant symbols, $f$ is a two-place function symbol, and $A$ is a two-place pred-
icate symbol. Consider the structure $M$ defined by:

1. $|M| = \{1, 2, 3, 4\}$
2. $a^M = 1$
3. $b^M = 2$
4. $f^M(x, y) = x + y$ if $x + y \leq 3$ and $= 3$ otherwise.
5. $R^M = \{(1, 1), (1, 2), (2, 3), (2, 4)\}$

The function $s(x) = 1$ that assigns $1 \in |M|$ to every variable is a variable assignment for $M$.

Then

$$\text{Val}_s^M(f(a, b)) = f^M(\text{Val}_s^M(a), \text{Val}_s^M(b)).$$

Since $a$ and $b$ are constant symbols, $\text{Val}_s^M(a) = a^M = 1$ and $\text{Val}_s^M(b) = b^M = 2$. So

$$\text{Val}_s^M(f(a, b)) = f^M(1, 2) = 1 + 2 = 3.$$

To compute the value of $f(f(a, b), a)$ we have to consider

$$\text{Val}_s^M(f(f(a, b), a)) = f^M(\text{Val}_s^M(f(a, b)), \text{Val}_s^M(a)) = f^M(3, 1) = 3,$$

since $3 + 1 > 3$. Since $s(x) = 1$ and $\text{Val}_s^M(x) = s(x)$, we also have

$$\text{Val}_s^M(f(f(a, b), x)) = f^M(\text{Val}_s^M(f(a, b)), \text{Val}_s^M(x)) = f^M(3, 1) = 3.$$

An atomic formula $R(t_1, t_2)$ is satisfied if the tuple of values of its arguments, i.e., $\langle \text{Val}_s^M(t_1), \text{Val}_s^M(t_2) \rangle$, is an element of $R^M$. So, e.g., we have $M, s \models R(b, f(a, b))$ since $\langle \text{Val}_s^M(b), \text{Val}_s^M(f(a, b)) \rangle = \langle 2, 3 \rangle \in R^M$, but $M \not\models R(x, f(a, b))$ since $\langle 1, 3 \rangle \notin R^M[s]$.

To determine if a non-atomic formula $A$ is satisfied, you apply the clauses in the inductive definition that applies to the main connective. For instance, the main connective in $R(a, a) \rightarrow (R(b, x) \lor R(x, b))$ is the $\rightarrow$, and

$$M, s \models R(a, a) \rightarrow (R(b, x) \lor R(x, b))$$

iff
\[ M, s \not\models R(a, a) \text{ or } M, s \models R(b, x) \lor R(x, b) \]

Since \( M, s \models R(a, a) \) (because \( \langle 1, 1 \rangle \in R^M \)) we can’t yet determine the answer and must first figure out if \( M, s \not\models R(b, x) \lor R(x, b) \):

\[
M, s \models R(b, x) \lor R(x, b) \text{ iff }
M, s \models R(b, x) \text{ or } M, s \models R(x, b)
\]

And this is the case, since \( M, s \models R(b, x) \) (because \( \langle 1, 2 \rangle \in R^M \)).

Recall that an \( x \)-variant of \( s \) is a variable assignment that differs from \( s \) at most in what it assigns to \( x \). For every element of \( |M| \), there is an \( x \)-variant of \( s \): \( s_1(x) = 1 \), \( s_2(x) = 2 \), \( s_3(x) = 3 \), \( s_4(s) = 4 \), and with \( s_i(y) = s(y) = 1 \) for all variables \( y \) other than \( x \) are all the \( x \)-variants of \( s \) for the structure \( M \). Note, in particular, that \( s_1 = s \) is also an \( x \)-variant of \( s \), i.e., \( s \) is an \( x \)-variant of itself.

To determine if an existentially quantified formula \( \exists x \, A(x) \) is satisfied, we have to determine if \( M, s' \models A(x) \) for at least one \( x \)-variant \( s' \) of \( s \). So,

\[ M, s \models \exists x \, (R(b, x) \lor R(x, b)) \]

since \( M, s_1 \models R(b, x) \lor R(x, b) \) (\( s_3 \) would also fit the bill). But,

\[ M, s \not\models \exists x \, (R(b, x) \land R(x, b)) \]

since for none of the \( s_i \), \( M, s_i \models R(b, x) \land R(x, b) \).

To determine if a universally quantified formula \( \forall x \, A(x) \) is satisfied, we have to determine if \( M, s' \models A(x) \) for all \( x \)-variants \( s' \) of \( s \). So,

\[ M, s \models \forall x \, (R(x, a) \rightarrow R(a, x)) \]

since \( M, s_i \models R(x, a) \rightarrow R(a, x) \) for all \( s_i \) (\( M, s_1 \models R(a, x) \) and \( M, s_j \not\models R(a, x) \) for \( j = 2, 3 \), and 4). But,

\[ M, s \not\models \forall x \, (R(a, x) \rightarrow R(x, a)) \]
since \( M, s_2 \not\models R(a, x) \rightarrow R(x, a) \) (because \( M, s_2 \models R(a, x) \) and \( M, s_2 \not\models R(x, a) \)).

For a more complicated case, consider

\[
\forall x (R(a, x) \rightarrow \exists y R(x, y)).
\]

Since \( M, s_3 \not\models R(a, x) \) and \( M, s_4 \not\models R(a, x) \), the interesting cases where we have to worry about the consequent of the conditional are only \( s_1 \) and \( s_2 \). Does \( M, s_1 \models \exists y R(x, y) \) hold? It does if there is at least one \( y \)-variant \( s'_1 \) of \( s_1 \) so that \( M, s'_1 \models R(x, y) \). In fact, \( s_1 \) is such a \( y \)-variant \( (s_1(x) = 1, s_1(y) = 1, \langle 1, 1 \rangle \in R^M) \), so the answer is yes. To determine if \( M, s_2 \models \exists y R(x, y) \) we have to look at the \( y \)-variants of \( s_2 \). Here, \( s_2 \) itself does not satisfy \( R(x, y) \) \( (s_2(x) = 2, s_2(y) = 1, \langle 1, 1 \rangle \not\in R^M) \). However, consider \( s'_2 \sim_y s_2 \) with \( s'_2(y) = 3 \). \( M, s'_2 \models R(x, y) \) since \( \langle 2, 3 \rangle \in R^M \), and so \( M, s_2 \models \exists y R(x, y) \). In sum, for every \( x \)-variant \( s_i \) of \( s \), either \( M, s_i \not\models R(a, x) \) \( (i = 3, 4) \) or \( M, s_i \models \exists y R(x, y) \) \( (i = 1, 2) \), and so

\[
M, s \models \forall x (R(a, x) \rightarrow \exists y R(x, y)).
\]

On the other hand,

\[
M, s \not\models \exists x (R(a, x) \land \forall y R(x, y)).
\]

The only \( x \)-variants \( s_i \) of \( s \) with \( M, s_i \models R(a, x) \) are \( s_1 \) and \( s_2 \). But for each, there is in turn a \( y \)-variant \( s'_i \sim_y s_i \) with \( s'_i(y) = 4 \) so that \( M, s'_i \not\models R(x, y) \) and so \( M, s_i \not\models \forall y R(x, y) \) for \( i = 1, 2 \). In sum, none of the \( x \)-variants \( s_i \sim_x s \) are such that \( M, s_i \models R(a, x) \land \forall y R(x, y) \).

### A.7 Variable Assignments

A variable assignment \( s \) provides a value for every variable—and there are infinitely many of them. This is of course not necessary. We require variable assignments to assign values to all variables simply because it makes things a lot easier. The value of a term \( t \), and whether or not a formula \( A \) is satisfied in a structure with respect to \( s \), only depend on the assignments \( s \) makes to
the variables in \( t \) and the free variables of \( A \). This is the content of the next two propositions. To make the idea of “depends on” precise, we show that any two variable assignments that agree on all the variables in \( t \) give the same value, and that \( A \) is satisfied relative to one iff it is satisfied relative to the other if two variable assignments agree on all free variables of \( A \).

**Proposition A.25.** If the variables in a term \( t \) are among \( x_1, \ldots, x_n \), and \( s_1(x_i) = s_2(x_i) \) for \( i = 1, \ldots, n \), then \( \text{Val}^M_{s_1}(t) = \text{Val}^M_{s_2}(t) \).

**Proof.** By induction on the complexity of \( t \). For the base case, \( t \) can be a constant symbol or one one of the variables \( x_1, \ldots, x_n \). If \( t = c \), then \( \text{Val}^M_{s_1}(t) = c^M = \text{Val}^M_{s_2}(t) \). If \( t = x_i \), \( s_1(x_i) = s_2(x_i) \) by the hypothesis of the proposition, and so \( \text{Val}^M_{s_1}(t) = s_1(x_i) = s_2(x_i) = \text{Val}^M_{s_2}(t) \).

For the inductive step, assume that \( t = f(t_1, \ldots, t_k) \) and that the claim holds for \( t_1, \ldots, t_k \). Then

\[
\text{Val}^M_{s_1}(t) = \text{Val}^M_{s_1}(f(t_1, \ldots, t_k)) = f^M(\text{Val}^M_{s_1}(t_1), \ldots, \text{Val}^M_{s_1}(t_k))
\]

For \( j = 1, \ldots, k \), the variables of \( t_j \) are among \( x_1, \ldots, x_n \). So by induction hypothesis, \( \text{Val}^M_{s_1}(t_j) = \text{Val}^M_{s_2}(t_j) \). So,

\[
\text{Val}^M_{s_1}(t) = \text{Val}^M_{s_2}(f(t_1, \ldots, t_k)) = f^M(\text{Val}^M_{s_1}(t_1), \ldots, \text{Val}^M_{s_1}(t_k)) = f^M(\text{Val}^M_{s_2}(t_1), \ldots, \text{Val}^M_{s_2}(t_k)) = \text{Val}^M_{s_2}(f(t_1, \ldots, t_k)) = \text{Val}^M_{s_2}(t).
\]

\( \square \)
Proposition A.26. If the free variables in $A$ are among $x_1, \ldots, x_n$, and $s_1(x_i) = s_2(x_i)$ for $i = 1, \ldots, n$, then $M, s_1 \models A$ iff $M, s_2 \models A$.

Proof. We use induction on the complexity of $A$. For the base case, where $A$ is atomic, $A$ can be: $\bot$, $R(t_1, \ldots, t_k)$ for a $k$-place predicate $R$ and terms $t_1, \ldots, t_k$, or $t_1 = t_2$ for terms $t_1$ and $t_2$.

1. $A \equiv \bot$: both $M, s_1 \not\models A$ and $M, s_2 \not\models A$.

2. $A \equiv R(t_1, \ldots, t_k)$: let $M, s_1 \models A$. Then

   $$\langle \text{Val}_{s_1}^M(t_1), \ldots, \text{Val}_{s_1}^M(t_k) \rangle \in R^M.$$  
   
   For $i = 1, \ldots, k$, $\text{Val}_{s_1}^M(t_i) = \text{Val}_{s_2}^M(t_i)$ by Proposition A.25. So we also have $\langle \text{Val}_{s_2}^M(t_1), \ldots, \text{Val}_{s_2}^M(t_k) \rangle \in R^M$.

3. $A \equiv t_1 = t_2$: suppose $M, s_1 \models A$. Then $\text{Val}_{s_1}^M(t_1) = \text{Val}_{s_1}^M(t_2)$.

   So,

   $$\text{Val}_{s_2}^M(t_1) = \text{Val}_{s_1}^M(t_1) \quad \text{(by Proposition A.25)}$$

   $$= \text{Val}_{s_1}^M(t_2) \quad \text{(since $M, s_1 \models t_1 = t_2$)}$$

   $$= \text{Val}_{s_2}^M(t_2) \quad \text{(by Proposition A.25)},$$

   so $M, s_2 \models t_1 = t_2$.

Now assume $M, s_1 \models B$ iff $M, s_2 \models B$ for all formulas $B$ less complex than $A$. The induction step proceeds by cases determined by the main operator of $A$. In each case, we only demonstrate the forward direction of the biconditional; the proof of the reverse direction is symmetrical. In all cases except those for the quantifiers, we apply the induction hypothesis to sub-formulas $B$ of $A$. The free variables of $B$ are among those of $A$. Thus, if $s_1$ and $s_2$ agree on the free variables of $A$, they also agree on those of $B$, and the induction hypothesis applies to $B$.

1. $A \equiv \neg B$: if $M, s_1 \models A$, then $M, s_1 \not\models B$, so by the induction hypothesis, $M, s_2 \not\models B$, hence $M, s_2 \models A$. 


2. \( A \equiv B \land C \): exercise.

3. \( A \equiv B \lor C \): if \( M, s_1 \models A \), then \( M, s_1 \models B \) or \( M, s_1 \models C \). By induction hypothesis, \( M, s_2 \models B \) or \( M, s_2 \models C \), so \( M, s_2 \models A \).

4. \( A \equiv B \rightarrow C \): exercise.

5. \( A \equiv \exists x B \): if \( M, s_1 \models A \), there is an \( x \)-variant \( s'_1 \) of \( s_1 \) so that \( M, s'_1 \models B \). Let \( s'_2 \) be the \( x \)-variant of \( s_2 \) that assigns the same thing to \( x \) as does \( s'_1 \). The free variables of \( B \) are among \( x_1, \ldots, x_n \), and \( x \). \( s'_1(x_i) = s'_2(x_i) \), since \( s'_1 \) and \( s'_2 \) are \( x \)-variants of \( s_1 \) and \( s_2 \), respectively, and by hypothesis \( s_1(x_i) = s_2(x_i) \). \( s'_1(x) = s'_2(x) \) by the way we have defined \( s'_2 \). Then the induction hypothesis applies to \( B \) and \( s'_1, s'_2 \), so \( M, s'_2 \models B \). Hence, there is an \( x \)-variant of \( s_2 \) that satisfies \( B \), and so \( M, s_2 \models A \).

6. \( A \equiv \forall x B \): exercise.

By induction, we get that \( M, s_1 \models A \) iff \( M, s_2 \models A \) whenever the free variables in \( A \) are among \( x_1, \ldots, x_n \) and \( s_1(x_i) = s_2(x_i) \) for \( i = 1, \ldots, n \). □

Sentences have no free variables, so any two variable assignments assign the same things to all the (zero) free variables of any sentence. The proposition just proved then means that whether or not a sentence is satisfied in a structure relative to a variable assignment is completely independent of the assignment. We’ll record this fact. It justifies the definition of satisfaction of a sentence in a structure (without mentioning a variable assignment) that follows.
**Corollary A.27.** If \( A \) is a sentence and \( s \) a variable assignment, then \( M, s \models A \) iff \( M, s' \models A \) for every variable assignment \( s' \).

*Proof.* Let \( s' \) be any variable assignment. Since \( A \) is a sentence, it has no free variables, and so every variable assignment \( s' \) trivially assigns the same things to all free variables of \( A \) as does \( s \). So the condition of Proposition A.26 is satisfied, and we have \( M, s \models A \) iff \( M, s' \models A \). □

**Definition A.28.** If \( A \) is a sentence, we say that a structure \( M \) satisfies \( A \), \( M \models A \), iff \( M, s \models A \) for all variable assignments \( s \).

If \( M \models A \), we also simply say that \( A \) is true in \( M \).

**Proposition A.29.** Let \( M \) be a structure, \( A \) be a sentence, and \( s \) a variable assignment. \( M \models A \) iff \( M, s \models A \).

*Proof.* Exercise. □

**Proposition A.30.** Suppose \( A(x) \) only contains \( x \) free, and \( M \) is a structure. Then:

1. \( M \models \exists x \ A(x) \) iff \( M, s \models A(x) \) for at least one variable assignment \( s \).

2. \( M \models \forall x \ A(x) \) iff \( M, s \models A(x) \) for all variable assignments \( s \).

*Proof.* Exercise. □

**A.8 Extensionality**

Extensionality, sometimes called relevance, can be expressed informally as follows: the only factors that bears upon the satisfaction of formula \( A \) in a structure \( M \) relative to a variable assignment \( s \), are the size of the domain and the assignments made
A.8. EXTENSIONALITY

by \( M \) and \( s \) to the elements of the language that actually appear in \( A \).

One immediate consequence of extensionality is that where two structures \( M \) and \( M' \) agree on all the elements of the language appearing in a sentence \( A \) and have the same domain, \( M \) and \( M' \) must also agree on whether or not \( A \) itself is true.

**Proposition A.31** (Extensionality). Let \( A \) be a formula, and \( M_1 \) and \( M_2 \) be structures with \( |M_1| = |M_2| \), and \( s \) a variable assignment on \( |M_1| = |M_2| \). If \( c^{M_1} = c^{M_2} \), \( R^{M_1} = R^{M_2} \), and \( f^{M_1} = f^{M_2} \) for every constant symbol \( c \), relation symbol \( R \), and function symbol \( f \) occurring in \( A \), then \( M_1, s \models A \) iff \( M_2, s \models A \).

**Proof.** First prove (by induction on \( t \)) that for every term, \( \text{Val}_{s}^{M_1}(t) = \text{Val}_{s}^{M_2}(t) \). Then prove the proposition by induction on \( A \), making use of the claim just proved for the induction basis (where \( A \) is atomic). \( \square \)

**Corollary A.32** (Extensionality for Sentences). Let \( A \) be a sentence and \( M_1, M_2 \) as in Proposition A.31. Then \( M_1 \models A \) iff \( M_2 \models A \).

**Proof.** Follows from Proposition A.31 by Corollary A.27. \( \square \)

Moreover, the value of a term, and whether or not a structure satisfies a formula, only depends on the values of its subterms.

**Proposition A.33.** Let \( M \) be a structure, \( t \) and \( t' \) terms, and \( s \) a variable assignment. Let \( s' \sim_x s \) be the \( x \)-variant of \( s \) given by \( s'(x) = \text{Val}_{s}^{M}(t') \). Then \( \text{Val}_{s}^{M}(t[t'/x]) = \text{Val}_{s'}^{M}(t) \).

**Proof.** By induction on \( t \).

1. If \( t \) is a constant, say, \( t \equiv c \), then \( t[t'/x] = c \), and \( \text{Val}_{s}^{M}(c) = c^{M} = \text{Val}_{s'}^{M}(c) \).

2. If \( t \) is a variable other than \( x \), say, \( t \equiv y \), then \( t[t'/x] = y \), and \( \text{Val}_{s}^{M}(y) = \text{Val}_{s'}^{M}(y) \) since \( s' \sim_x s \).
3. If $t \equiv x$, then $t[t'/x] = t'$. But $\text{Val}_{s'}^M(x) = \text{Val}_s^M(t')$ by definition of $s'$.

4. If $t \equiv f(t_1, \ldots, t_n)$ then we have:

$$\text{Val}_s^M(t[t'/x]) =$$

$$= \text{Val}_s^M(f(t_1[t'/x], \ldots, t_n[t'/x]))$$

by definition of $t[t'/x]$

$$= f^M(\text{Val}_s^M(t_1[t'/x]), \ldots, \text{Val}_s^M(t_n[t'/x]))$$

by definition of $\text{Val}_s^M(f(\ldots))$

$$= f^M(\text{Val}_{s'}^M(t_1), \ldots, \text{Val}_{s'}^M(t_n))$$

by induction hypothesis

$$= \text{Val}_{s'}^M(t) \text{ by definition of } \text{Val}_{s'}^M(f(\ldots))$$

\[\square\]

**Proposition A.34.** Let $M$ be a structure, $A$ a formula, $t$ a term, and $s$ a variable assignment. Let $s' \sim_x s$ be the $x$-variant of $s$ given by $s'(x) = \text{Val}_s^M(t)$. Then $M, s \models A[t/x]$ iff $M, s' \models A$.

**Proof.** Exercise.  \[\square\]

### A.9 Semantic Notions

Give the definition of structures for first-order languages, we can define some basic semantic properties of and relationships between sentences. The simplest of these is the notion of *validity* of a sentence. A sentence is valid if it is satisfied in every structure. Valid sentences are those that are satisfied regardless of how the non-logical symbols in it are interpreted. Valid sentences are therefore also called *logical truths*—they are true, i.e., satisfied, in any structure and hence their truth depends only on the logical symbols occurring in them and their syntactic structure, but not on the non-logical symbols or their interpretation.
**Definition A.35** (Validity). A sentence $A$ is valid, $\vDash A$, iff $M \vDash A$ for every structure $M$.

**Definition A.36** (Entailment). A set of sentences $\Gamma$ entails a sentence $A$, $\Gamma \vDash A$, iff for every structure $M$ with $M \vDash \Gamma$, $M \vDash A$.

**Definition A.37** (Satisfiability). A set of sentences $\Gamma$ is satisfiable if $M \vDash \Gamma$ for some structure $M$. If $\Gamma$ is not satisfiable it is called unsatisfiable.

**Proposition A.38.** A sentence $A$ is valid iff $\Gamma \vDash A$ for every set of sentences $\Gamma$.

*Proof.* For the forward direction, let $A$ be valid, and let $\Gamma$ be a set of sentences. Let $M$ be a structure so that $M \vDash \Gamma$. Since $A$ is valid, $M \vDash A$, hence $\Gamma \vDash A$.

For the contrapositive of the reverse direction, let $A$ be invalid, so there is a structure $M$ with $M \nvdash A$. When $\Gamma = \{\top\}$, since $\top$ is valid, $M \vDash \Gamma$. Hence, there is a structure $M$ so that $M \vDash \Gamma$ but $M \nvdash A$, hence $\Gamma$ does not entail $A$. \(\square\)

**Proposition A.39.** $\Gamma \vDash A$ iff $\Gamma \cup \{\neg A\}$ is unsatisfiable.

*Proof.* For the forward direction, suppose $\Gamma \vDash A$ and suppose to the contrary that there is a structure $M$ so that $M \vDash \Gamma \cup \{\neg A\}$. Since $M \vDash \Gamma$ and $\Gamma \vDash A$, $M \vDash A$. Also, since $M \vDash \Gamma \cup \{\neg A\}$, $M \vDash \neg A$, so we have both $M \vDash A$ and $M \nvdash A$, a contradiction. Hence, there can be no such structure $M$, so $\Gamma \cup \{A\}$ is unsatisfiable.

For the reverse direction, suppose $\Gamma \cup \{\neg A\}$ is unsatisfiable. So for every structure $M$, either $M \nvdash \Gamma$ or $M \vDash A$. Hence, for every structure $M$ with $M \vDash \Gamma$, $M \vDash A$, so $\Gamma \vDash A$. \(\square\)
**Proposition A.40.** If $\Gamma \subseteq \Gamma'$ and $\Gamma \models A$, then $\Gamma' \models A$.

*Proof.* Suppose that $\Gamma \subseteq \Gamma'$ and $\Gamma \models A$. Let $M$ be such that $M \models \Gamma'$; then $M \models \Gamma$, and since $\Gamma \models A$, we get that $M \models A$. Hence, whenever $M \models \Gamma'$, $M \models A$, so $\Gamma' \models A$. \hfill \blacksquare

**Theorem A.41** (Semantic Deduction Theorem). $\Gamma \cup \{A\} \models B$ iff $\Gamma \models A \to B$.

*Proof.* For the forward direction, let $\Gamma \cup \{A\} \models B$ and let $M$ be a structure so that $M \models \Gamma$. If $M \models A$, then $M \models \Gamma \cup \{A\}$, so since $\Gamma \cup \{A\}$ entails $B$, we get $M \models B$. Therefore, $M \models A \to B$, so $\Gamma \models A \to B$.

For the reverse direction, let $\Gamma \models A \to B$ and $M$ be a structure so that $M \models \Gamma \cup \{A\}$. Then $M \models \Gamma$, so $M \models A \to B$, and since $M \models A$, $M \models B$. Hence, whenever $M \models \Gamma \cup \{A\}$, $M \models B$, so $\Gamma \cup \{A\} \models B$. \hfill \blacksquare

**Proposition A.42.** Let $M$ be a structure, and $A(x)$ a formula with one free variable $x$, and $t$ a closed term. Then:

1. $A(t) \models \exists x A(x)$
2. $\forall x A(x) \models A(t)$

*Proof.*

1. Suppose $M \models A(t)$. Let $s$ be a variable assignment with $s(x) = \text{Val}^M(t)$. Then $M, s \models A(t)$ since $A(t)$ is a sentence. By Proposition A.34, $M, s \models A(x)$. By Proposition A.30, $M \models \exists x A(x)$.

2. Exercise. \hfill \blacksquare
A.10 Theories

**Definition A.43.** A set of sentences $\Gamma$ is closed iff, whenever $\Gamma \models A$ then $A \in \Gamma$. The closure of a set of sentences $\Gamma$ is $\{ A : \Gamma \models A \}$.

We say that $\Gamma$ is axiomatized by a set of sentences $\Delta$ if $\Gamma$ is the closure of $\Delta$.

**Example A.44.** The theory of strict linear orders in the language $\mathcal{L}_<$ is axiomatized by the set

\[
\forall x \neg x < x, \\
\forall x \forall y ((x < y \lor y < x) \lor x = y), \\
\forall x \forall y \forall z ((x < y \land y < z) \rightarrow x < z)
\]

It completely captures the intended structures: every strict linear order is a model of this axiom system, and vice versa, if $R$ is a linear order on a set $X$, then the structure $M$ with $|M| = X$ and $<_M = R$ is a model of this theory.

**Example A.45.** The theory of groups in the language $\mathcal{L}$ (constant symbol), $\cdot$ (two-place function symbol) is axiomatized by

\[
\forall x (x \cdot 1) = x \\
\forall x \forall y \forall z (x \cdot (y \cdot z)) = ((x \cdot y) \cdot z) \\
\forall x \exists y (x \cdot y) = 1
\]

**Example A.46.** The theory of Peano arithmetic is axiomatized by the following sentences in the language of arithmetic $\mathcal{L}_A$.

\[
\neg \exists x x' = 0 \\
\forall x \forall y (x' = y' \rightarrow x = y) \\
\forall x \forall y (x < y \leftrightarrow \exists z (x + z' = y)) \\
\forall x (x + 0) = x \\
\forall x \forall y (x + y') = (x + y)' \\
\forall x (x \times 0) = 0
\]
\[ \forall x \forall y (x \times y') = ((x \times y) + x) \]

plus all sentences of the form

\[ (A(\circ) \land \forall x (A(x) \rightarrow A(x'))) \rightarrow \forall x A(x) \]

Since there are infinitely many sentences of the latter form, this axiom system is infinite. The latter form is called the induction schema. (Actually, the induction schema is a bit more complicated than we let on here.)

The third axiom is an explicit definition of <.

**Summary**

A first-order language consists of constant, function, and predicate symbols. Function and constant symbols take a specified number of arguments. In the language of arithmetic, e.g., we have a single constant symbol 0, one 1-place function symbol \(\prime\), two 2-place function symbols + and \(\times\), and one 2-place predicate symbol \(<\). From variables and constant and function symbols we form the terms of a language. From the terms of a language together with its predicate symbol, as well as the identity symbol \(=\), we form the atomic formulas. And in turn from them, using the logical connectives \(\neg\), \(\lor\), \(\land\), \(\rightarrow\), \(\leftrightarrow\) and the quantifiers \(\forall\) and \(\exists\) we form its formulas. Since we are careful to always include necessary parentheses in the process of forming terms and formulas, there is always exactly one way of reading a formula. This makes it possible to define things by induction on the structure of formulas.

Occurrences of variables in formulas are sometimes governed by a corresponding quantifier: if a variable occurs in the scope of a quantifier it is considered bound, otherwise free. These concepts all have inductive definitions, and we also inductively define the operation of substitution of a term for a variable in a formula. Formulas without free variable occurrences are called sentences.
The **semantics** for a first-order language is given by a **structure** for that language. It consists of a **domain** and elements of that domain are assigned to each constant symbol. Function symbols are interpreted by functions and relation symbols by relation on the domain. A function from the set of variables to the domain is a **variable assignment**. The relation of **satisfaction** relates structures, variable assignments and formulas; $\mathcal{M} \models [s]A$ is defined by induction on the structure of $A$. $\mathcal{M} \models [s]A$ only depends on the interpretation of the symbols actually occurring in $A$, and in particular does not depend on $s$ if $A$ contains no free variables. So if $A$ is a sentence, $\mathcal{M} \models A$ if $\mathcal{M} \models [s]A$ for any (or all) $s$.

The satisfaction relation is the basis for all semantic notions. A sentence is **valid**, $A \models$, if it is satisfied in every structure. A sentence $A$ is **entailed** by set of sentences $\Gamma$, $\Gamma \models A$, iff $\mathcal{M} \models A$ for all $\mathcal{M}$ which satisfy every sentence in $\Gamma$. A set $\Gamma$ is **satisfiable** iff there is some structure that satisfies every sentence in $\Gamma$, otherwise unsatisfiable. These notions are interrelated, e.g., $\Gamma \models A$ iff $\Gamma \cup \{\neg A\}$ is unsatisfiable.

**Problems**

**Problem A.1.** Give an inductive definition of the bound variable occurrences along the lines of **Definition A.8**.

**Problem A.2.** Let $\mathcal{L} = \{c, f, A\}$ with one constant symbol, one one-place function symbol and one two-place predicate symbol, and let the structure $\mathcal{M}$ be given by

1. $|\mathcal{M}| = \{1, 2, 3\}$
2. $c^\mathcal{M} = 3$
3. $f^\mathcal{M}(1) = 2, f^\mathcal{M}(2) = 3, f^\mathcal{M}(3) = 2$
4. $A^\mathcal{M} = \{(1, 2), (2, 3), (3, 3)\}$
(a) Let $s(v) = 1$ for all variables $v$. Find out whether

$$M, s \models \exists x (A(f(z), c) \to \forall y (A(y, x) \lor A(f(y), x)))$$

Explain why or why not.

(b) Give a different structure and variable assignment in which the formula is not satisfied.

**Problem A.3.** Complete the proof of Proposition A.26.

**Problem A.4.** Prove Proposition A.29.

**Problem A.5.** Prove Proposition A.30.

**Problem A.6.** Suppose $\mathcal{L}$ is a language without function symbols. Given a structure $M$, a constant symbol and $a \in |M|$, define $M[a/c]$ to be the structure that is just like $M$, except that $c_{M[a/c]} = a$. Define $M \models A$ for sentences $A$ by:

1. $A \equiv \bot$: not $M \models A$.

2. $A \equiv R(d_1, \ldots, d_n)$: $M \models A$ iff $\langle d_1^M, \ldots, d_n^M \rangle \in R^M$.

3. $A \equiv d_1 = d_2$: $M \models A$ iff $d_1^M = d_2^M$.

4. $A \equiv \neg B$: $M \models A$ iff not $M \models B$.

5. $A \equiv (B \land C)$: $M \models A$ iff $M \models B$ and $M \models C$.

6. $A \equiv (B \lor C)$: $M \models A$ iff $M \models B$ or $M \models C$ (or both).

7. $A \equiv (B \to C)$: $M \models A$ iff not $M \models B$ or $M \models C$ (or both).

8. $A \equiv \forall x B$: $M \models A$ iff for all $a \in |M|$, $M[a/c] \models B[c/x]$, if $c$ does not occur in $B$.

9. $A \equiv \exists x B$: $M \models A$ iff there is an $a \in |M|$ such that $M[a/c] \models B[c/x]$, if $c$ does not occur in $B$. 
Let $x_1, \ldots, x_n$ be all free variables in $A$, $c_1, \ldots, c_n$ constant symbols not in $A$, $a_1, \ldots, a_n \in |M|$, and $s(x_i) = a_i$.

Show that $M, s \models A$ iff $M[a_1/c_1, \ldots, a_n/c_n] \models A[c_1/x_1] \ldots [c_n/x_n]$.

(This problem shows that it is possible to give a semantics for first-order logic that makes do without variable assignments.)

**Problem A.7.** Suppose that $f$ is a function symbol not in $A(x, y)$. Show that there is a structure $M$ such that $M \models \forall x \exists y A(x, y)$ iff there is an $M'$ such that $M' \models \forall x A(x, f(x))$.

(This problem is a special case of what’s known as Skolem’s Theorem; $\forall x A(x, f(x))$ is called a *Skolem normal form* of $\forall x \exists y A(x, y)$.)

**Problem A.8.** Carry out the proof of Proposition A.31 in detail.

**Problem A.9.** Prove Proposition A.34

**Problem A.10.**

1. Show that $\Gamma \models \bot$ iff $\Gamma$ is unsatisfiable.

2. Show that $\Gamma \cup \{A\} \models \bot$ iff $\Gamma \models \neg A$.

3. Suppose $c$ does not occur in $A$ or $\Gamma$. Show that $\Gamma \models \forall x A$ iff $\Gamma \models A[c/x]$.

**Problem A.11.** Complete the proof of Proposition A.42.
APPENDIX B

Natural Deduction

B.1 Natural Deduction

Natural deduction is a derivation system intended to mirror actual reasoning (especially the kind of regimented reasoning employed by mathematicians). Actual reasoning proceeds by a number of “natural” patterns. For instance, proof by cases allows us to establish a conclusion on the basis of a disjunctive premise, by establishing that the conclusion follows from either of the disjuncts. Indirect proof allows us to establish a conclusion by showing that its negation leads to a contradiction. Conditional proof establishes a conditional claim “if … then …” by showing that the consequent follows from the antecedent. Natural deduction is a formalization of some of these natural inferences. Each of the logical connectives and quantifiers comes with two rules, an introduction and an elimination rule, and they each correspond to one such natural inference pattern. For instance, $\rightarrow$Intro corresponds to conditional proof, and $\lor$Elim to proof by cases. A particularly simple rule is $\land$Elim which allows the inference from $A \land B$ to $A$ (or $B$).

One feature that distinguishes natural deduction from other
derivation systems is its use of assumptions. A derivation in natural deduction is a tree of formulas. A single formula stands at the root of the tree of formulas, and the “leaves” of the tree are formulas from which the conclusion is derived. In natural deduction, some leaf formulas play a role inside the derivation but are “used up” by the time the derivation reaches the conclusion. This corresponds to the practice, in actual reasoning, of introducing hypotheses which only remain in effect for a short while. For instance, in a proof by cases, we assume the truth of each of the disjuncts; in conditional proof, we assume the truth of the antecedent; in indirect proof, we assume the truth of the negation of the conclusion. This way of introducing hypothetical assumptions and then doing away with them in the service of establishing an intermediate step is a hallmark of natural deduction. The formulas at the leaves of a natural deduction derivation are called assumptions, and some of the rules of inference may “discharge” them. For instance, if we have a derivation of $B$ from some assumptions which include $A$, then the $\rightarrow$Intro rule allows us to infer $A \rightarrow B$ and discharge any assumption of the form $A$. (To keep track of which assumptions are discharged at which inferences, we label the inference and the assumptions it discharges with a number.) The assumptions that remain undischarged at the end of the derivation are together sufficient for the truth of the conclusion, and so a derivation establishes that its undischarged assumptions entail its conclusion.

The relation $\Gamma \vdash A$ based on natural deduction holds iff there is a derivation in which $A$ is the last sentence in the tree, and every leaf which is undischarged is in $\Gamma$. $A$ is a theorem in natural deduction iff there is a derivation in which $A$ is the last sentence and all assumptions are discharged. For instance, here is a derivation that shows that $\vdash (A \land B) \rightarrow A$:

\[
\begin{align*}
\frac{[A \land B]^1}{A} & \quad \text{\textit{\&Elim}} \\
1 & \quad \frac{1}{(A \land B) \rightarrow A} \quad \text{\textit{\rightarrow Intro}}
\end{align*}
\]
The label 1 indicates that the assumption $A \land B$ is discharged at the $\rightarrow$Intro inference.

A set $\Gamma$ is inconsistent iff $\Gamma \vdash \bot$ in natural deduction. The rule $\bot_I$ makes it so that from an inconsistent set, any sentence can be derived.

Natural deduction systems were developed by Gerhard Gentzen and Stanisław Jaśkowski in the 1930s, and later developed by Dag Prawitz and Frederic Fitch. Because its inferences mirror natural methods of proof, it is favored by philosophers. The versions developed by Fitch are often used in introductory logic textbooks. In the philosophy of logic, the rules of natural deduction have sometimes been taken to give the meanings of the logical operators (“proof-theoretic semantics”).

### B.2 Rules and Derivations

Natural deduction systems are meant to closely parallel the informal reasoning used in mathematical proof (hence it is somewhat “natural”). Natural deduction proofs begin with assumptions. Inference rules are then applied. Assumptions are “discharged” by the $\neg$Intro, $\rightarrow$Intro, $\lor$Elim and $\exists$Elim inference rules, and the label of the discharged assumption is placed beside the inference for clarity.

**Definition B.1** (Initial Formula). An *initial formula* or *assumption* is any formula in the topmost position of any branch.

Derivations in natural deduction are certain trees of sentences, where the topmost sentences are assumptions, and if a sentence stands below one, two, or three other sequents, it must follow correctly by a rule of inference. The sentences at the top of the inference are called the *premises* and the sentence below the *conclusion* of the inference. The rules come in pairs, an introduction and an elimination rule for each logical operator. They introduce a logical operator in the conclusion or remove a logical operator
from a premise of the rule. Some of the rules allow an assumption of a certain type to be discharged. To indicate which assumption is discharged by which inference, we also assign labels to both the assumption and the inference. This is indicated by writing the assumption as “\([A]^n\)“.

It is customary to consider rules for all logical operators, even for those (if any) that we consider as defined.

### B.3 Propositional Rules

#### Rules for ∧

<table>
<thead>
<tr>
<th>Rule</th>
<th>Formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>∧Intro</td>
<td>(\frac{A \land B}{A} \quad \frac{A \land B}{B})</td>
</tr>
<tr>
<td>∧Elim</td>
<td>(\frac{A}{A \land B} \quad \frac{B}{A \land B})</td>
</tr>
</tbody>
</table>

#### Rules for ∨

<table>
<thead>
<tr>
<th>Rule</th>
<th>Formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>∨Intro</td>
<td>(\frac{A}{A \lor B} \quad \frac{B}{A \lor B})</td>
</tr>
<tr>
<td>∨Elim</td>
<td>(\frac{A \lor B}{C} \quad \frac{C}{C} \quad \frac{C}{A \lor B})</td>
</tr>
</tbody>
</table>

#### Rules for →

<table>
<thead>
<tr>
<th>Rule</th>
<th>Formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>→Intro</td>
<td>(\frac{A \to B}{A} \quad \frac{B}{A \to B})</td>
</tr>
<tr>
<td>→Elim</td>
<td>(\frac{A \to B}{B} \quad \frac{A}{A \to B})</td>
</tr>
</tbody>
</table>

#### Rules for ¬
APPENDIX B. NATURAL DEDUCTION

\[ [A]^n \]
\[ \cdots \]
\[ \cdots \]
\[ n \quad \vdash \quad \neg A \quad \neg \text{Intro} \]
\[ \neg A \quad A \quad \neg \text{Elim} \]

Rules for \( \bot \)

\[ \frac{\bot}{A} \quad \bot \text{I} \]
\[ \frac{[\neg A]^n}{\cdots} \]
\[ n \quad \vdash \quad A \quad \bot \text{C} \]

Note that \( \neg \text{Intro} \) and \( \bot \text{C} \) are very similar: The difference is that \( \neg \text{Intro} \) derives a negated sentence \( \neg A \) but \( \bot \text{C} \) a positive sentence \( A \).

\subsection*{B.4 Quantifier Rules}

Rules for \( \forall \)

\[ \frac{A(a)}{\forall x \ A(x)} \quad \forall \text{Intro} \]
\[ \frac{\forall x \ A(x)}{A(t)} \quad \forall \text{Elim} \]

In the rules for \( \forall \), \( t \) is a ground term (a term that does not contain any variables), and \( a \) is a constant symbol which does not occur in the conclusion \( \forall x \ A(x) \), or in any assumption which is undischarged in the derivation ending with the premise \( A(a) \). We call \( a \) the \textit{eigenvariable} of the \( \forall \text{Intro} \) inference.

Rules for \( \exists \)
Again, \( t \) is a ground term, and \( a \) is a constant which does not occur in the premise \( \exists x A(x) \), in the conclusion \( C \), or any assumption which is undischarged in the derivations ending with the two premises (other than the assumptions \( A(a) \)). We call \( a \) the *eigenvariable* of the \( \exists \text{Elim} \) inference.

The condition that an eigenvariable neither occur in the premises nor in any assumption that is undischarged in the derivations leading to the premises for the \( \forall \text{Intro} \) or \( \exists \text{Elim} \) inference is called the *eigenvariable condition*.

We use the term “eigenvariable” even though \( a \) in the above rules is a constant. This has historical reasons.

In \( \exists \text{Intro} \) and \( \forall \text{Elim} \) there are no restrictions, and the term \( t \) can be anything, so we do not have to worry about any conditions. On the other hand, in the \( \exists \text{Elim} \) and \( \forall \text{Intro} \) rules, the eigenvariable condition requires that the constant symbol \( a \) does not occur anywhere in the conclusion or in an undischarged assumption. The condition is necessary to ensure that the system is sound, i.e., only derives sentences from undischarged assumptions from which the follow. Without this condition, the following would be allowed:

\[
\exists x A(x) \quad [A(a)]^1 \\
\forall x A(x) \quad \forall \text{Intro}
\]

\[
\forall x A(x) \quad \exists \text{Elim}
\]

However, \( \exists x A(x) \not\equiv \forall x A(x) \).

**B.5 Derivations**

We’ve said what an assumption is, and we’ve given the rules of
inference. Derivations in natural deduction are inductively generated from these: each derivation either is an assumption on its own, or consists of one, two, or three derivations followed by a correct inference.

**Definition B.2** (Derivation). A *derivation* of a sentence $A$ from assumptions $\Gamma$ is a tree of sentences satisfying the following conditions:

1. The topmost sentences of the tree are either in $\Gamma$ or are discharged by an inference in the tree.
2. The bottommost sentence of the tree is $A$.
3. Every sentence in the tree except $A$ is a premise of a correct application of an inference rule whose conclusion stands directly below that sentence in the tree.

We then say that $A$ is the *conclusion* of the derivation and that $A$ is *derivable* from $\Gamma$.

**Example B.3.** Every assumption on its own is a derivation. So, e.g., $C$ by itself is a derivation, and so is $D$ by itself. We can obtain a new derivation from these by applying, say, the $\land$Intro rule,

$$
\frac{A \quad B}{A \land B} \land\text{Intro}
$$

These rules are meant to be general: we can replace the $A$ and $B$ in it with any sentences, e.g., by $C$ and $D$. Then the conclusion would be $C \land D$, and so

$$
\frac{C \quad D}{C \land D} \land\text{Intro}
$$

is a correct derivation. Of course, we can also switch the assumptions, so that $D$ plays the role of $A$ and $C$ that of $B$. Thus,

$$
\frac{D \quad C}{D \land C} \land\text{Intro}
$$
is also a correct derivation.

We can now apply another rule, say, →Intro, which allows us to conclude a conditional and allows us to discharge any assumption that is identical to the conclusion of that conditional. So both of the following would be correct derivations:

\[
\begin{array}{c}
[C]^1 \quad D \\
\hline
C \land D \\
\hline
C 
\end{array} \quad \land\text{Intro} \\
\begin{array}{c}
\hline
C \rightarrow (C \land D) \\
\hline
C 
\end{array} \quad \rightarrow\text{Intro}
\]

\[
\begin{array}{c}
\hline
[C]^1 \quad D \\
\hline
C \land D \\
\hline
D \rightarrow (C \land D) \\
\hline
D 
\end{array} \quad \land\text{Intro} \\
\begin{array}{c}
\hline
C \land D \\
\hline
C 
\end{array} \quad \rightarrow\text{Intro}
\]

### B.6 Examples of Derivations

**Example B.4.** Let’s give a derivation of the sentence \((A \land B) \rightarrow A\).

We begin by writing the desired conclusion at the bottom of the derivation.

\[
(A \land B) \rightarrow A
\]

Next, we need to figure out what kind of inference could result in a sentence of this form. The main operator of the conclusion is →, so we’ll try to arrive at the conclusion using the →Intro rule. It is best to write down the assumptions involved and label the inference rules as you progress, so it is easy to see whether all assumptions have been discharged at the end of the proof.

\[
\begin{array}{c}
[A \land B]^1 \\
\vdots \\
\vdots \\
\hline
A \\
\hline
(A \land B) \rightarrow A \\
\hline
\end{array} \quad \rightarrow\text{Intro}
\]

We now need to fill in the steps from the assumption \(A \land B\) to \(A\). Since we only have one connective to deal with, \(\land\), we must use the \(\land\) elim rule. This gives us the following proof:

\[
\begin{array}{c}
[A \land B]^1 \\
\hline
A \\
\hline
(A \land B) \rightarrow A \\
\hline
\end{array} \quad \land\text{Elim} \\
\begin{array}{c}
\hline
(A \land B) \rightarrow A \\
\hline
\end{array} \quad \rightarrow\text{Intro}
\]

We now have a correct derivation of \((A \land B) \rightarrow A\).
**Example B.5.** Now let’s give a derivation of \((\neg A \lor B) \rightarrow (A \rightarrow B)\).

We begin by writing the desired conclusion at the bottom of the derivation.

\[
\overline{(\neg A \lor B) \rightarrow (A \rightarrow B)}
\]

To find a logical rule that could give us this conclusion, we look at the logical connectives in the conclusion: \(\neg\), \(\lor\), and \(\rightarrow\). We only care at the moment about the first occurrence of \(\rightarrow\) because it is the main operator of the sentence in the end-sequent, while \(\neg\), \(\lor\) and the second occurrence of \(\rightarrow\) are inside the scope of another connective, so we will take care of those later. We therefore start with the \(\rightarrow\)Intro rule. A correct application must look as follows:

\[
[\neg A \lor B]^1
\]

\[
\vdots
\]

\[
\vdots
\]

\[
1 \quad A \rightarrow B
\]

\[
\frac{\neg A \lor B}{} \rightarrow Intro
\]

This leaves us with two possibilities to continue. Either we can keep working from the bottom up and look for another application of the \(\rightarrow\)Intro rule, or we can work from the top down and apply a \(\lor\)Elim rule. Let us apply the latter. We will use the assumption \(\neg A \lor B\) as the leftmost premise of \(\lor\)Elim. For a valid application of \(\lor\)Elim, the other two premises must be identical to the conclusion \(A \rightarrow B\), but each may be derived in turn from another assumption, namely the two disjuncts of \(\neg A \lor B\). So our derivation will look like this:

\[
[\neg A]^2
\]

\[
\vdots
\]

\[
\vdots
\]

\[
2 \quad \neg A \lor B
\]

\[
\vdots
\]

\[
\vdots
\]

\[
1 \quad A \rightarrow B
\]

\[
\frac{A \rightarrow B \quad A \rightarrow B}{} \lor Elim
\]

\[
\frac{A \rightarrow B}{} \rightarrow Intro
\]
In each of the two branches on the right, we want to derive $A \rightarrow B$, which is best done using $\rightarrow$Intro.

\[
\begin{array}{cccc}
\vdots & \vdots & \vdots & \vdots \\
\end{array}
\]

\[
\begin{array}{c}
2 \quad \frac{[-A \lor B]^1}{A \rightarrow B} \quad \rightarrow Intro \\
\end{array}
\]

\[
\begin{array}{c}
3 \quad \frac{B}{A \rightarrow B} \quad \rightarrow Intro \\
4 \quad \frac{A \rightarrow B}{(\neg A \lor B) \rightarrow (A \rightarrow B)} \quad \vee Elim \\
\end{array}
\]

For the two missing parts of the derivation, we need derivations of $B$ from $\neg A$ and $A$ in the middle, and from $A$ and $B$ on the left. Let’s take the former first. $\neg A$ and $A$ are the two premises of $\neg$Elim:

\[
\begin{array}{c}
\vdots & \vdots & \vdots \\
\end{array}
\]

\[
\begin{array}{c}
\frac{\neg A}{\bot} \quad \neg Elim \\
\end{array}
\]

\[
\begin{array}{c}
\vdots \\
\end{array}
\]

\[
\begin{array}{c}
\frac{B}{A \rightarrow B} \quad \rightarrow Intro \\
\end{array}
\]

By using $\bot_I$, we can obtain $B$ as a conclusion and complete the branch.

\[
\begin{array}{c}
\vdots & \vdots & \vdots \\
\end{array}
\]

\[
\begin{array}{c}
\frac{\bot}{B} \quad \bot Intro \\
\end{array}
\]

\[
\begin{array}{c}
\frac{B}{A \rightarrow B} \quad \rightarrow Intro \\
\end{array}
\]

\[
\begin{array}{c}
\frac{A \rightarrow B}{(\neg A \lor B) \rightarrow (A \rightarrow B)} \quad \vee Elim \\
\end{array}
\]

Let’s now look at the rightmost branch. Here it’s important to realize that the definition of derivation allows assumptions to be discharged but does not require them to be. In other words, if we can derive $B$ from one of the assumptions $A$ and $B$ without using the other, that’s ok. And to derive $B$ from $B$ is trivial: $B$ by itself is such a derivation, and no inferences are needed. So we can
simply delete the assumption $A$.

\[
\begin{align*}
&\frac{\neg A}{\perp} \quad \frac{\perp}{\neg A} \quad \frac{\neg A \lor B}{\perp} \quad \text{Intro} \\
&\frac{\neg A \lor B}{\perp} \quad \text{Intro} \quad \frac{A \lor B}{\neg A} \quad \text{Intro} \quad \frac{B}{A \rightarrow B} \quad \text{Intro} \\
&\frac{A \rightarrow B}{\neg A \lor B \rightarrow (A \rightarrow B)} \quad \text{Intro}
\end{align*}
\]

Note that in the finished derivation, the rightmost $\rightarrow$Intro inference does not actually discharge any assumptions.

**Example B.6.** So far we have not needed the $\perp_C$ rule. It is special in that it allows us to discharge an assumption that isn’t a sub-formula of the conclusion of the rule. It is closely related to the $\perp_I$ rule. In fact, the $\perp_I$ rule is a special case of the $\perp_C$ rule—there is a logic called “intuitionistic logic” in which only $\perp_I$ is allowed. The $\perp_C$ rule is a last resort when nothing else works. For instance, suppose we want to derive $A \lor \neg A$. Our usual strategy would be to attempt to derive $A \lor \neg A$ using $\lor$Intro. But this would require us to derive either $A$ or $\neg A$ from no assumptions, and this can’t be done. $\perp_C$ to the rescue!

\[
\begin{align*}
&\frac{\neg (A \lor \neg A)}{\perp} \\
&\frac{\perp}{\neg (A \lor \neg A)} \\
&\frac{A \lor \neg A}{\neg (A \lor \neg A)} \quad \text{Intro} \\
&\frac{\perp}{\neg (A \lor \neg A)} \quad \text{Intro} \\
&\frac{A \lor \neg A}{\neg (A \lor \neg A)} \quad \text{Intro}
\end{align*}
\]

Now we’re looking for a derivation of $\perp$ from $\neg (A \lor \neg A)$. Since $\perp$ is the conclusion of $\neg$Elim we might try that:

\[
\begin{align*}
&\frac{\neg (A \lor \neg A)}{\perp} \\
&\frac{\perp}{\neg (A \lor \neg A)} \\
&\frac{A \lor \neg A}{\neg (A \lor \neg A)} \quad \text{Intro} \\
&\frac{\perp}{\neg (A \lor \neg A)} \quad \text{Intro} \\
&\frac{A \lor \neg A}{\neg (A \lor \neg A)} \quad \text{Intro}
\end{align*}
\]

Our strategy for finding a derivation of $\neg A$ calls for an application of $\neg\text{Intro}$:

\[
\begin{array}{c}
\neg\neg (A \lor \neg A) \quad \left[\neg\neg (A \lor \neg A)\right]^1, [A]^2 \\
\vdots \\
2 \quad \frac{\bot}{\neg A} \quad \neg\text{Intro} \\
\frac{\bot}{A} \quad \neg\text{Elim} \\
\vdots \\
1 \quad \frac{A \lor \neg A}{\bot} \quad \bot_C
\end{array}
\]

Here, we can get $\bot$ easily by applying $\neg\text{Elim}$ to the assumption $\neg(A \lor \neg A)$ and $A \lor \neg A$ which follows from our new assumption $A$ by $\lor\text{Intro}$:

\[
\begin{array}{c}
\neg\neg (A \lor \neg A) \quad \left[\neg\neg (A \lor \neg A)\right]^1 \\
\vdots \\
2 \quad \frac{\bot}{\neg A} \quad \neg\text{Intro} \\
\frac{[A]^2}{A \lor \neg A} \quad \lor\text{Intro} \\
\frac{\neg\neg (A \lor \neg A) \quad [\neg\neg (A \lor \neg A)]^1}{\vdots} \\
\frac{\bot}{A} \quad \neg\text{Elim} \\
\vdots \\
1 \quad \frac{A \lor \neg A}{\bot} \quad \bot_C
\end{array}
\]

On the right side we use the same strategy, except we get $A$ by $\bot_C$:

\[
\begin{array}{c}
\neg\neg (A \lor \neg A) \quad \left[\neg\neg (A \lor \neg A)\right]^1 \\
\vdots \\
2 \quad \frac{\bot}{\neg A} \quad \neg\text{Intro} \\
\frac{[A]^2}{A \lor \neg A} \quad \lor\text{Intro} \\
\frac{\neg\neg (A \lor \neg A) \quad [\neg\neg (A \lor \neg A)]^1}{\vdots} \\
\frac{\bot}{A} \quad \neg\text{Elim} \\
3 \quad \frac{\bot}{A} \quad \bot_C \\
\frac{\bot}{A \lor \neg A} \quad \bot_C \\
\vdots \\
1 \quad \frac{A \lor \neg A}{\bot} \quad \bot_C
\end{array}
\]

### B.7 Derivations with Quantifiers

**Example B.7.** When dealing with quantifiers, we have to make sure not to violate the eigenvariable condition, and sometimes this requires us to play around with the order of carrying out certain inferences. In general, it helps to try and take care of rules
subject to the eigenvariable condition first (they will be lower
down in the finished proof).

Let’s see how we’d give a derivation of the formula $\exists x \neg A(x) \rightarrow \neg \forall x A(x)$. Starting as usual, we write

$$
\exists x \neg A(x) \rightarrow \neg \forall x A(x)
$$

We start by writing down what it would take to justify that last step using the $\rightarrow$Intro rule.

$$
[\exists x \neg A(x)]^1 \\
\vdots \\
\vdots \\
\neg \forall x A(x) \\
\overline{\exists x \neg A(x) \rightarrow \neg \forall x A(x)} \rightarrow \text{Intro}
$$

Since there is no obvious rule to apply to $\neg \forall x A(x)$, we will proceed by setting up the derivation so we can use the $\exists$Elim rule. Here we must pay attention to the eigenvariable condition, and choose a constant that does not appear in $\exists x A(x)$ or any assumptions that it depends on. (Since no constant symbols appear, however, any choice will do fine.)

$$
[\neg A(a)]^2 \\
\vdots \\
\overline{\exists x \neg A(x) \rightarrow \neg \forall x A(x)} \rightarrow \text{Intro}
$$

In order to derive $\neg \forall x A(x)$, we will attempt to use the $\neg$Intro rule: this requires that we derive a contradiction, possibly using $\forall x A(x)$ as an additional assumption. Of course, this contradiction may involve the assumption $\neg A(a)$ which will be discharged
by the $\rightarrow$Intro inference. We can set it up as follows:

\[
\begin{array}{c}
[\neg A(a)]_2, [\forall x A(x)]_3 \\
\vdots \\
\vdots \\
2 \quad [\exists x \neg A(x)]_1 \\
\end{array}
\]

\[
\begin{array}{c}
3 \quad \neg \forall x A(x) \\
\forall x A(x) \\
\exists x \neg A(x) \rightarrow \neg \forall x A(x) \\
\rightarrow \text{Intro} \\
\end{array}
\]

It looks like we are close to getting a contradiction. The easiest rule to apply is the $\forall$Elim, which has no eigenvariable conditions. Since we can use any term we want to replace the universally quantified $x$, it makes the most sense to continue using $a$ so we can reach a contradiction.

\[
\begin{array}{c}
[\neg A(a)]_2 \\
\vdots \\
3 \quad \neg \forall x A(x) \\
\forall x A(x) \\
\exists x \neg A(x) \rightarrow \neg \forall x A(x) \\
\rightarrow \text{Intro} \\
\end{array}
\]

It is important, especially when dealing with quantifiers, to double check at this point that the eigenvariable condition has not been violated. Since the only rule we applied that is subject to the eigenvariable condition was $\exists$Elim, and the eigenvariable $a$ does not occur in any assumptions it depends on, this is a correct derivation.

**Example B.8.** Sometimes we may derive a formula from other formulas. In these cases, we may have undischarged assumptions. It is important to keep track of our assumptions as well as the end goal.

Let’s see how we’d give a derivation of the formula $\exists x C(x, b)$ from the assumptions $\exists x (A(x) \land B(x))$ and $\forall x (B(x) \rightarrow C(x, b))$. Starting as usual, we write the conclusion at the bottom.

\[
\exists x C(x, b)
\]
We have two premises to work with. To use the first, i.e., try to find a derivation of $\exists x\ C(x, b)$ from $\exists x\ (A(x) \land B(x))$ we would use the $\exists$Elim rule. Since it has an eigenvariable condition, we will apply that rule first. We get the following:

\[
\begin{array}{c}
\exists x\ (A(x) \land B(x)) \quad 1 \\
\end{array}
\]

\[
\begin{array}{c}
\therefore \exists x\ C(x, b) \\
\exists\text{Elim}
\end{array}
\]

The two assumptions we are working with share $B$. It may be useful at this point to apply $\land$Elim to separate out $B(a)$.

\[
\begin{array}{c}
[A(a) \land B(a)]^1 \\
\end{array}
\]

\[
\begin{array}{c}
B(a) \\
\land\text{Elim}
\end{array}
\]

\[
\begin{array}{c}
\exists x\ (A(x) \land B(x)) \\
\exists\text{Elim}
\end{array}
\]

The second assumption we have to work with is $\forall x\ (B(x) \rightarrow C(x, b))$. Since there is no eigenvariable condition we can instantiate $x$ with the constant symbol $a$ using $\forall$Elim to get $B(a) \rightarrow C(a, b)$. We now have both $B(a) \rightarrow C(a, b)$ and $B(a)$. Our next move should be a straightforward application of the $\rightarrow$Elim rule.

\[
\begin{array}{c}
\forall x\ (B(x) \rightarrow C(x, b)) \\
\forall\text{Elim}
\end{array}
\]

\[
\begin{array}{c}
[A(a) \land B(a)]^1 \\
\forall\text{Elim}
\end{array}
\]

\[
\begin{array}{c}
B(a) \\
\rightarrow\text{Elim}
\end{array}
\]

\[
\begin{array}{c}
\therefore C(a, b) \\
\rightarrow\text{Elim}
\end{array}
\]

\[
\begin{array}{c}
\exists x\ (A(x) \land B(x)) \\
\exists\text{Elim}
\end{array}
\]

\[
\begin{array}{c}
\exists x\ C(x, b) \\
\exists\text{Elim}
\end{array}
\]

We are so close! One application of $\exists$Intro and we have reached
our goal.

\[
\begin{align*}
& \forall x (B(x) \to C(x, b)) \quad \forall \text{Elim} \\
& \frac{\begin{array}{c} B(a) \to C(a, b) \\ \forall \text{Elim} \quad \frac{\begin{array}{c} [A(a) \land B(a)]^1 \\ \land \text{Elim} \end{array}}{B(a)} \end{array}}{\begin{array}{c} C(a, b) \\ \exists \text{Intro} \quad \frac{\exists x C(x, b)}{\exists x C(x, b)} \quad \exists \text{Elim} \end{array}} \\
& \frac{1 \quad \exists x (A(x) \land B(x))}{\exists x C(x, b)} \\
& \frac{1}{\exists x C(x, b)}
\end{align*}
\]

Since we ensured at each step that the eigenvariable conditions were not violated, we can be confident that this is a correct derivation.

**Example B.9.** Give a derivation of the formula \(\neg \forall x A(x)\) from the assumptions \(\forall x A(x) \to \exists y B(y)\) and \(\neg \exists y B(y)\). Starting as usual, we write the target formula at the bottom.

\[
\neg \forall x A(x)
\]

The last line of the derivation is a negation, so let’s try using \(\neg \text{Intro}\). This will require that we figure out how to derive a contradiction.

\[
\begin{align*}
[\forall x A(x)]^1 \\
\vdots \\
\vdots \\
1 \quad \neg \forall x A(x) \quad \neg \text{Intro}
\end{align*}
\]

So far so good. We can use \(\forall \text{Elim}\) but it’s not obvious if that will help us get to our goal. Instead, let’s use one of our assumptions. \(\forall x A(x) \to \exists y B(y)\) together with \(\forall x A(x)\) will allow us to use the \(\to \text{Elim}\) rule.

\[
\begin{align*}
& \forall x A(x) \to \exists y B(y) \\
& \frac{[\forall x A(x)]^1}{\exists y B(y)} \quad \to \text{Elim} \\
& \vdots \\
& \vdots \\
& 1 \quad \neg \forall x A(x) \quad \neg \text{Intro}
\end{align*}
\]
We now have one final assumption to work with, and it looks like this will help us reach a contradiction by using \( \neg\text{Elim} \).

\[
\begin{align*}
\neg \exists y B(y) & \quad \forall x A(x) \rightarrow \exists y B(y) \\
\exists y B(y) & \quad [\forall x A(x)]^1 \\
\rightarrow\text{Elim} & \\
\hline
\downarrow & \\
\neg\text{Intro} & \\
1 & \neg\forall x A(x) \quad \neg\text{Intro}
\end{align*}
\]

### B.8 Derivations with Identity predicate

Derivations with identity predicate require additional inference rules.

\[
\begin{align*}
t_1 = t_2 & \quad A(t_1) \\
A(t_2) & \quad =\text{Elim} \\
\hline
t_1 = t_2 & \quad A(t_2) \\
A(t_1) & \quad =\text{Elim}
\end{align*}
\]

In the above rules, \( t, t_1, \) and \( t_2 \) are closed terms. The \( =\text{Intro} \) rule allows us to derive any identity statement of the form \( t = t \) outright, from no assumptions.

**Example B.10.** If \( s \) and \( t \) are closed terms, then \( A(s), s = t \vdash A(t) \):

\[
s = t \quad A(s) \\
\hline
A(t) \quad =\text{Elim}
\]

This may be familiar as the “principle of substitutability of identicals,” or Leibniz’ Law.

**Example B.11.** We derive the sentence

\[
\forall x \forall y ((A(x) \land A(y)) \rightarrow x = y)
\]

from the sentence

\[
\exists x \forall y (A(y) \rightarrow y = x)
\]

We develop the derivation backwards:
\[ \exists x \forall y (A(y) \rightarrow y = x) \quad [A(a) \land A(b)]^1 \]

[ \[ \begin{align*}
&\vdots \\
&1 \quad a = b \\
&((A(a) \land A(b)) \rightarrow a = b) \quad \rightarrow \text{Intro} \\
&\forall y ((A(a) \land A(y)) \rightarrow a = y) \quad \forall \text{Intro} \\
&\forall x \forall y ((A(x) \land A(y)) \rightarrow x = y) \quad \forall \text{Intro}
\end{align*} \]

We’ll now have to use the main assumption: since it is an existential formula, we use \( \exists \text{Elim} \) to derive the intermediary conclusion \( a = b \).

\[ [\forall y (A(y) \rightarrow y = c)]^2 \]
\[ [A(a) \land A(b)]^1 \]
\[ \vdots \]
\[ \vdots \]
\[ 2 \quad \exists x \forall y (A(y) \rightarrow y = x) \quad a = b \quad \exists \text{Elim} \\
\[ \begin{align*}
&1 \quad a = b \\
&((A(a) \land A(b)) \rightarrow a = b) \quad \rightarrow \text{Intro} \\
&\forall y ((A(a) \land A(y)) \rightarrow a = y) \quad \forall \text{Intro} \\
&\forall x \forall y ((A(x) \land A(y)) \rightarrow x = y) \quad \forall \text{Intro}
\end{align*} \]

The sub-derivation on the top right is completed by using its assumptions to show that \( a = c \) and \( b = c \). This requires two separate derivations. The derivation for \( a = c \) is as follows:

\[ [\forall y (A(y) \rightarrow y = c)]^2 \quad \forall \text{Elim} \]
\[ [A(a) \land A(b)]^1 \quad \land \text{Elim} \]
\[ A(a) \rightarrow a = c \quad \forall \text{Elim} \]
\[ A(a) \rightarrow a = c \quad \land \text{Elim} \]
\[ a = c \]
\[ a = c \quad \rightarrow \text{Elim} \]

From \( a = c \) and \( b = c \) we derive \( a = b \) by \( = \text{Elim} \).

**B.9 Proof-Theoretic Notions**

Just as we’ve defined a number of important semantic notions (validity, entailment, satisfiability), we now define corresponding
proof-theoretic notions. These are not defined by appeal to satisfaction of sentences in structures, but by appeal to the derivability or non-derivability of certain sentences from others. It was an important discovery that these notions coincide. That they do is the content of the soundness and completeness theorems.

**Definition B.12** (Theorems). A sentence $A$ is a *theorem* if there is a derivation of $A$ in natural deduction in which all assumptions are discharged. We write $\vdash A$ if $A$ is a theorem and $\not\vdash A$ if it is not.

**Definition B.13** (Derivability). A sentence $A$ is *derivable from* a set of sentences $\Gamma$, $\Gamma \vdash A$, if there is a derivation with conclusion $A$ and in which every assumption is either discharged or is in $\Gamma$. If $A$ is not derivable from $\Gamma$ we write $\not\Gamma \vdash A$.

**Definition B.14** (Consistency). A set of sentences $\Gamma$ is *inconsistent* iff $\Gamma \vdash \bot$. If $\Gamma$ is not inconsistent, i.e., if $\not\Gamma \vdash \bot$, we say it is *consistent*.

**Proposition B.15** (Reflexivity). *If* $A \in \Gamma$, *then* $\Gamma \vdash A$.

*Proof.* The assumption $A$ by itself is a derivation of $A$ where every undischarged assumption (i.e., $A$) is in $\Gamma$.

**Proposition B.16** (Monotony). *If* $\Gamma \subseteq \Delta$ *and* $\Gamma \vdash A$, *then* $\Delta \vdash A$.

*Proof.* Any derivation of $A$ from $\Gamma$ is also a derivation of $A$ from $\Delta$. 

$\square$
**Proposition B.17** (Transitivity). If $\Gamma \vdash A$ and $\{A\} \cup \Delta \vdash B$, then $\Gamma \cup \Delta \vdash B$.

**Proof.** If $\Gamma \vdash A$, there is a derivation $\delta_0$ of $A$ with all undischarged assumptions in $\Gamma$. If $\{A\} \cup \Delta \vdash B$, then there is a derivation $\delta_1$ of $B$ with all undischarged assumptions in $\{A\} \cup \Delta$. Now consider:

$\Delta, [A]^1$

\[ \vdots \]

\[ \vdots \]

\[ \vdots \]

\[ 1 \frac{B}{A \rightarrow B \rightarrow \text{Intro}} \]

\[ \frac{A}{B} \rightarrow \text{Elim} \]

The undischarged assumptions are now all among $\Gamma \cup \Delta$, so this shows $\Gamma \cup \Delta \vdash B$. □

Note that this means that in particular if $\Gamma \vdash A$ and $A \vdash B$, then $\Gamma \vdash B$. It follows also that if $A_1, \ldots, A_n \vdash B$ and $\Gamma \vdash A_i$ for each $i$, then $\Gamma \vdash B$.

**Proposition B.18.** $\Gamma$ is inconsistent iff $\Gamma \vdash A$ for every sentence $A$.

**Proof.** Exercise. □

**Proposition B.19** (Compactness). 1. If $\Gamma \vdash A$ then there is a finite subset $\Gamma_0 \subseteq \Gamma$ such that $\Gamma_0 \vdash A$.

2. If every finite subset of $\Gamma$ is consistent, then $\Gamma$ is consistent.

**Proof.** 1. If $\Gamma \vdash A$, then there is a derivation $\delta$ of $A$ from $\Gamma$. Let $\Gamma_0$ be the set of undischarged assumptions of $\delta$. Since any derivation is finite, $\Gamma_0$ can only contain finitely many sentences. So, $\delta$ is a derivation of $A$ from a finite $\Gamma_0 \subseteq \Gamma$.

2. This is the contrapositive of (1) for the special case $A \equiv \bot$. □
Summary

Proof systems provide purely syntactic methods for characterizing consequence and compatibility between sentences. Natural deduction is one such proof system. A derivation in it consists of a tree of formulas. The topmost formula a derivation are assumptions. All other formulas, for the derivation to be correct, must be correctly justified by one of a number of inference rules. These come in pairs; an introduction and an elimination rule for each connective and quantifier. For instance, if a formula $A$ is justified by a $\rightarrow$Elim rule, the preceding formulas (the premises) must be $B \rightarrow A$ and $B$ (for some $B$). Some inference rules also allow assumptions to be discharged. For instance, if $A \rightarrow B$ is inferred from $B$ using $\rightarrow$Intro, any occurrences of $A$ as assumptions in the derivation leading to the premise $B$ may be discharged, given a label that is also recorded at the inference.

If there is a derivation with end formula $A$ and all assumptions are discharged, we say $A$ is a theorem and write $\vdash A$. If all undischarged assumptions are in some set $\Gamma$, we say $A$ is derivable from $\Gamma$ and write $\Gamma \vdash A$. If $\Gamma \vdash \bot$ we say $\Gamma$ is inconsistent, otherwise consistent. These notions are interrelated, e.g., $\Gamma \vdash A$ iff $\Gamma \cup \{\neg A\} \vdash \bot$. They are also related to the corresponding semantic notions, e.g., if $\Gamma \vdash A$ then $\Gamma \vDash A$. This property of natural deduction—what can be derived from $\Gamma$ is guaranteed to be entailed by $\Gamma$—is called soundness. The soundness theorem is proved by induction on the length of derivations, showing that each individual inference preserves entailment of its conclusion from open assumptions provided its premises are entailed by their open assumptions.

Problems

Problem B.1. Give derivations of the following:

1. $\neg(A \rightarrow B) \rightarrow (A \land \neg B)$
2. \((A \rightarrow C) \lor (B \rightarrow C)\) from the assumption \((A \land B) \rightarrow C\)

**Problem B.2.** Give derivations of the following:

1. \(\exists y \ A(y) \rightarrow B\) from the assumption \(\forall x \ (A(x) \rightarrow B)\)

2. \(\exists x \ (A(x) \rightarrow \forall y \ A(y))\)

**Problem B.3.** Prove that \(=\) is both symmetric and transitive, i.e.,
give derivations of \(\forall x \ \forall y \ (x = y \rightarrow y = x)\) and \(\forall x \ \forall y \ \forall z ((x = y \land y = z) \rightarrow x = z)\)

**Problem B.4.** Give derivations of the following formulas:

1. \(\forall x \ \forall y ((x = y \land A(x)) \rightarrow A(y))\)

2. \(\exists x \ A(x) \land \forall y \ A(z) ((A(y) \land A(z)) \rightarrow y = z) \rightarrow \exists x \ A(x) \land \forall y \ A(y) \rightarrow y = x)\)

**Problem B.5.** Prove Proposition B.18
C.1 Alonzo Church

Alonzo Church was born in Washington, DC on June 14, 1903. In early childhood, an air gun incident left Church blind in one eye. He finished preparatory school in Connecticut in 1920 and began his university education at Princeton that same year. He completed his doctoral studies in 1927. After a couple years abroad, Church returned to Princeton. Church was known exceedingly polite and careful. His blackboard writing was immaculate, and he would preserve important papers by carefully covering them in Duco cement. Outside of his academic pursuits, he enjoyed reading science fiction magazines and was not afraid to write to the editors if he spotted any inaccuracies in the writing.

Church’s academic achievements were great. Together with his students Stephen Kleene and Barkley Rosser, he developed a theory of effective calculability, the lambda calculus, independently of Alan Turing’s development of the Turing machine. The
two definitions of computability are equivalent, and give rise to what is now known as the Church-Turing Thesis, that a function of the natural numbers is effectively computable if and only if it is computable via Turing machine (or lambda calculus). He also proved what is now known as Church’s Theorem: The decision problem for the validity of first-order formulas is unsolvable. Church continued his work into old age. In 1967 he left Princeton for UCLA, where he was professor until his retirement in 1990. Church passed away on August 1, 1995 at the age of 92.

Further Reading For a brief biography of Church, see Ender- ton (forthcoming). Church’s original writings on the lambda calculus and the Entscheidungsproblem (Church’s Thesis) are Church (1936a,b). Aspray (1984) records an interview with Church about the Princeton mathematics community in the 1930s. Church wrote a series of book reviews of the Journal of Symbolic Logic from 1936 until 1979. They are all archived on John MacFarlane’s web-site (MacFarlane, 2015).

C.2 Kurt Gödel

Kurt Gödel (gör-dle) was born on April 28, 1906 in Brünn in the Austro-Hungarian empire (now Brno in the Czech Republic). Due to his inquisitive and bright nature, young Kurtele was often called “Der kleine Herr Warum” (Little Mr. Why) by his family. He excelled in academics from primary school onward, where he got less than the highest grade only in mathematics. Gödel was often absent from school due to poor health and was exempt from physical education. He was diagnosed with rheumatic fever during his childhood. Throughout his life, he believed this permanently affected his heart despite medical assessment saying otherwise.

Gödel began studying at the University of Vienna in 1924 and completed his doctoral studies in 1929. He first intended to study physics, but his interests soon moved to mathematics and
especially logic, in part due to the influence of the philosopher Rudolf Carnap. His dissertation, written under the supervision of Hans Hahn, proved the completeness theorem of first-order predicate logic with identity (Gödel, 1929). Only a year later, he obtained his most famous results—the first and second incompleteness theorems (published in Gödel 1931). During his time in Vienna, Gödel was heavily involved with the Vienna Circle, a group of scientifically-minded philosophers that included Carnap, whose work was especially influenced by Gödel’s results.

In 1938, Gödel married Adele Nimbursky. His parents were not pleased: not only was she six years older than him and already divorced, but she worked as a dancer in a nightclub. Social pressures did not affect Gödel, however, and they remained happily married until his death.

After Nazi Germany annexed Austria in 1938, Gödel and Adele emigrated to the United States, where he took up a position at the Institute for Advanced Study in Princeton, New Jersey. Despite his introversion and eccentric nature, Gödel’s time at Princeton was collaborative and fruitful. He published essays in set theory, philosophy and physics. Notably, he struck up a particularly strong friendship with his colleague at the IAS, Albert Einstein.

In his later years, Gödel’s mental health deteriorated. His wife’s hospitalization in 1977 meant she was no longer able to cook his meals for him. Having suffered from mental health issues throughout his life, he succumbed to paranoia. Deathly afraid of being poisoned, Gödel refused to eat. He died of starvation on January 14, 1978, in Princeton.
Further Reading  For a complete biography of Gödel’s life is available, see John Dawson (1997). For further biographical pieces, as well as essays about Gödel’s contributions to logic and philosophy, see Wang (1990), Baaz et al. (2011), Takeuti et al. (2003), and Sigmund et al. (2007).

Gödel’s PhD thesis is available in the original German (Gödel, 1929). The original text of the incompleteness theorems is (Gödel, 1931). All of Gödel’s published and unpublished writings, as well as a selection of correspondence, are available in English in his Collected Papers Feferman et al. (1986, 1990).

For a detailed treatment of Gödel’s incompleteness theorems, see Smith (2013). For an informal, philosophical discussion of Gödel’s theorems, see Mark Linsenmayer’s podcast (Linsenmayer, 2014).

C.3  Rózsa Péter

Rózsa Péter was born Rósza Politzer, in Budapest, Hungary, on February 17, 1905. She is best known for her work on recursive functions, which was essential for the creation of the field of recursion theory.

Péter was raised during harsh political times—WWI raged when she was a teenager—but was able to attend the affluent Maria Terezia Girls’ School in Budapest, from where she graduated in 1922. She then studied at Pázmány Péter University (later renamed Loránd Eötvös University) in Budapest. She began studying chemistry at the insistence of her father, but later switched to mathematics, and graduated in 1927. Although she had the credentials to teach high school mathematics, the economic situation at the time was dire as the Great Depression
affected the world economy. During this time, Péter took odd jobs as a tutor and private teacher of mathematics. She eventually returned to university to take up graduate studies in mathematics. She had originally planned to work in number theory, but after finding out that her results had already been proven, she almost gave up on mathematics altogether. She was encouraged to work on Gödel’s incompleteness theorems, and unknowingly proved several of his results in different ways. This restored her confidence, and Péter went on to write her first papers on recursion theory, inspired by David Hilbert’s foundational program. She received her PhD in 1935, and in 1937 she became an editor for the *Journal of Symbolic Logic*.

Péter’s early papers are widely credited as founding contributions to the field of recursive function theory. In Péter (1935a), she investigated the relationship between different kinds of recursion. In Péter (1935b), she showed that a certain recursively defined function is not primitive recursive. This simplified an earlier result due to Wilhelm Ackermann. Péter’s simplified function is what’s now often called the Ackermann function—and sometimes, more properly, the Ackermann-Péter function. She wrote the first book on recursive function theory (Péter, 1951).

Despite the importance and influence of her work, Péter did not obtain a full-time teaching position until 1945. During the Nazi occupation of Hungary during World War II, Péter was not allowed to teach due to anti-Semitic laws. In 1944 the government created a Jewish ghetto in Budapest; the ghetto was cut off from the rest of the city and attended by armed guards. Péter was forced to live in the ghetto until 1945 when it was liberated. She then went on to teach at the Budapest Teachers Training College, and from 1955 onward at Eötvös Loránd University. She was the first female Hungarian mathematician to become an Academic Doctor of Mathematics, and the first woman to be elected to the Hungarian Academy of Sciences.

Péter was known as a passionate teacher of mathematics, who preferred to explore the nature and beauty of mathematical problems with her students rather than to merely lecture. As a result,
she was affectionately called “Aunt Rosa” by her students. Péter died in 1977 at the age of 71.

Further Reading For more biographical reading, see (O’Connor and Robertson, 2014) and (Andrásfai, 1986). Tamassy (1994) conducted a brief interview with Péter. For a fun read about mathematics, see Péter’s book *Playing With Infinity* (Péter, 2010).

C.4 Julia Robinson

Julia Bowman Robinson was an American mathematician. She is known mainly for her work on decision problems, and most famously for her contributions to the solution of Hilbert’s tenth problem. Robinson was born in St. Louis, Missouri on December 8, 1919. At a young age Robinson recalls being intrigued by numbers (Reid, 1986, 4). At age nine she contracted scarlet fever and suffered from several recurrent bouts of rheumatic fever. This forced her to spend much of her time in bed, putting her behind in her education. Although she was able to catch up with the help of private tutors, the physical effects of her illness had a lasting impact on her life.

Despite her childhood struggles, Robinson graduated high school with several awards in mathematics and the sciences. She started her university career at San Diego State College, and transferred to the University of California, Berkeley as a senior. There she was highly influenced by mathematician Raphael Robinson. They quickly became good friends, and married in 1941. As a spouse of a faculty member, Robinson was barred from teaching in the mathematics department at Berkeley. Although she
continued to audit mathematics classes, she hoped to leave university and start a family. Not long after her wedding, however, Robinson contracted pneumonia. She was told that there was substantial scar tissue build up on her heart due to the rheumatic fever she suffered as a child. Due to the severity of the scar tissue, the doctor predicted that she would not live past forty and she was advised not to have children (Reid, 1986, 13).

Robinson was depressed for a long time, but eventually decided to continue studying mathematics. She returned to Berkeley and completed her PhD in 1948 under the supervision of Alfred Tarski. The first-order theory of the real numbers had been shown to be decidable by Tarski, and from Gödel’s work it followed that the first-order theory of the natural numbers is undecidable. It was a major open problem whether the first-order theory of the rationals is decidable or not. In her thesis (1949), Robinson proved that it was not.

Interested in decision problems, Robinson next attempted to find a solution Hilbert’s tenth problem. This problem was one of a famous list of 23 mathematical problems posed by David Hilbert in 1900. The tenth problem asks whether there is an algorithm that will answer, in a finite amount of time, whether or not a polynomial equation with integer coefficients, such as $3x^2 - 2y + 3 = 0$, has a solution in the integers. Such questions are known as Diophantine problems. After some initial successes, Robinson joined forces with Martin Davis and Hilary Putnam, who were also working on the problem. They succeeded in showing that exponential Diophantine problems (where the unknowns may also appear as exponents) are undecidable, and showed that a certain conjecture (later called “J.R.”) implies that Hilbert’s tenth problem is undecidable (Davis et al., 1961). Robinson continued to work on the problem for the next decade. In 1970, the young Russian mathematician Yuri Matijasevich finally proved the J.R. hypothesis. The combined result is now called the Matijasevich-Robinson-Davis-Putnam theorem, or MDRP theorem for short. Matijasevich and Robinson became friends and collaborated on several papers. In a letter to Matijasevich, Robinson once wrote that “actually I am
very pleased that working together (thousands of miles apart) we are obviously making more progress than either one of us could alone” (Matijasevich, 1992, 45).

Robinson was the first female president of the American Mathematical Society, and the first woman to be elected to the National Academy of Science. She died on July 30, 1985 at the age of 65 after being diagnosed with leukemia.

Further Reading Robinson’s mathematical papers are available in her Collected Works (Robinson, 1996), which also includes a reprint of her National Academy of Sciences biographical memoir (Feferman, 1994). Robinson’s older sister Constance Reid published an “Autobiography of Julia,” based on interviews (Reid, 1986), as well as a full memoir (Reid, 1996). A short documentary about Robinson and Hilbert’s tenth problem was directed by George Csicsery (Csicsery, 2016). For a brief memoir about Yuri Matijasevich’s collaborations with Robinson, and her influence on his work, see (Matijasevich, 1992).

C.5 Alfred Tarski

Alfred Tarski was born on January 14, 1901 in Warsaw, Poland (then part of the Russian Empire). Often described as “Napoleonic,” Tarski was boisterous, talkative, and intense. His energy was often reflected in his lectures—he once set fire to a wastebasket while disposing of a cigarette during a lecture, and was forbidden from lecturing in that building again.

Tarski had a thirst for knowledge from a young age. Although later in life he would tell students
that he studied logic because it was the only class in which he got a B, his high school records show that he got A’s across the board—even in logic. He studied at the University of Warsaw from 1918 to 1924. Tarski first intended to study biology, but became interested in mathematics, philosophy, and logic, as the university was the center of the Warsaw School of Logic and Philosophy. Tarski earned his doctorate in 1924 under the supervision of Stanisław Leśniewski.

Before emigrating to the United States in 1939, Tarski completed some of his most important work while working as a secondary school teacher in Warsaw. His work on logical consequence and logical truth were written during this time. In 1939, Tarski was visiting the United States for a lecture tour. During his visit, Germany invaded Poland, and because of his Jewish heritage, Tarski could not return. His wife and children remained in Poland until the end of the war, but were then able to emigrate to the United States as well. Tarski taught at Harvard, the College of the City of New York, and the Institute for Advanced Study at Princeton, and finally the University of California, Berkeley. There he founded the multidisciplinary program in Logic and the Methodology of Science. Tarski died on October 26, 1983 at the age of 82.

Further Reading  For more on Tarski’s life, see the biography *Alfred Tarski: Life and Logic* (Feferman and Feferman, 2004). Tarski’s seminal works on logical consequence and truth are available in English in (Corcoran, 1983). All of Tarski’s original works have been collected into a four volume series, (Tarski, 1981).
Alonzo Church, p. 212: Portrait of Alonzo Church, undated, photographer unknown. Alonzo Church Papers; 1924–1995, (C0948) Box 60, Folder 3. Manuscripts Division, Department of Rare Books and Special Collections, Princeton University Library. © Princeton University. The Open Logic Project has obtained permission to use this image for inclusion in non-commercial OLP-derived materials. Permission from Princeton University is required for any other use.

Kurt Gödel, p. 214: Portrait of Kurt Gödel, ca. 1925, photographer unknown. From the Shelby White and Leon Levy Archives Center, Institute for Advanced Study, Princeton, NJ, USA, on deposit at Princeton University Library, Manuscript Division, Department of Rare Books and Special Collections, Kurt Gödel Papers, (Co282), Box 14b, #110000. The Open Logic Project has obtained permission from the Institute’s Archives Center to use this image for inclusion in non-commercial OLP-derived materials. Permission from the Archives Center is required for any other use.


Julia Robinson, p. 217: Portrait of Julia Robinson, unknown photographer, courtesy of Neil D. Reid. The Open Logic Project has obtained permission to use this image for inclusion in non-commercial OLP-derived materials. Permission is required for
any other use.

Alfred Tarski, p. 219: Passport photo of Alfred Tarski, 1939. Cropped and restored from a scan of Tarski’s passport by Joel Fuller. Original courtesy of Bancroft Library, University of California, Berkeley. Alfred Tarski Papers, Banc MSS 84/49. The Open Logic Project has obtained permission to use this image for inclusion in non-commercial OLP-derived materials. Permission from Bancroft Library is required for any other use.


MacFarlane, John. 2015. Alonzo Church’s JSL reviews. URL http://johnmacfarlane.net/church.html.


O’Connor, John J. and Edmund F. Robertson. 2014. Rózsa Péter. URL http://www-groups.dcs.st-and.ac.uk/~history/Biographies/Peter.html.


About the Open Logic Project

The Open Logic Text is an open-source, collaborative textbook of formal meta-logic and formal methods, starting at an intermediate level (i.e., after an introductory formal logic course). Though aimed at a non-mathematical audience (in particular, students of philosophy and computer science), it is rigorous.

The Open Logic Text is a collaborative project and is under active development. Coverage of some topics currently included may not yet be complete, and many sections still require substantial revision. We plan to expand the text to cover more topics in the future. We also plan to add features to the text, such as a glossary, a list of further reading, historical notes, pictures, better explanations, sections explaining the relevance of results to philosophy, computer science, and mathematics, and more problems and examples. If you find an error, or have a suggestion, please let the project team know.

The project operates in the spirit of open source. Not only is the text freely available, we provide the LaTeX source under
the Creative Commons Attribution license, which gives anyone the right to download, use, modify, re-arrange, convert, and re-distribute our work, as long as they give appropriate credit.

Please see the Open Logic Project website at openlogicproject.org for additional information.