MATHS 315 Mathematical Logic

Second Semester, 2007

Contents

1 Informal Statement Logic **1**

1.1 Statements and truth tables 1

1.2 Tautologies, logical equivalence and logical implication 4

1.3 Logical laws 7

1.4 Soundness and adequacy 8

1.5 Soundness of $\approx$ 9

1.6 Truth Functions and Disjunctive Normal Form 11

1.7 Adequacy of our method 14

1.8 A method that is not adequate* 15

1.9 Arguments and validity 18

2 Formal Statement Logic **21**

2.1 Post production systems 21

2.2 The system L 25

2.3 The Deduction Theorem for L 27

2.4 Truth assignments 32

2.5 Adequacy of L 34

2.6 Countable sets 37

2.7 The Adequacy Theorem for L 40

3 Informal Predicate Logic **43**

3.1 First order languages 45

3.2 Interpretations, satisfaction and truth 48

3.3 Logical validity and logical implication 55

3.4 Free and bound variables 58

3.5 Logical equivalence 61

3.6 Changing bound variables 63

3.7 Substitutions 66

3.8 Prenex normal form 71

4 Formal Predicate Logic **74**

4.1 The formal system $K_{\mathcal{L}}$ 74

4.2 The soundness theorem for $K_{\mathcal{L}}$ 77

4.3 The Deduction Theorem for $K_{\mathcal{L}}$ 79

4.4 Models and consistency 84

4.5 Compactness 87

5 First Order Systems **90**

5.1 First order systems with equality 90

5.2 Normal models 93

5.3 The Peano Postulates 98

5.4 First order arithmetic 101

6 First Order Arithmetic **103**

6.1 Basic consequences of the axioms of N 103

6.2 Using adequacy of $K_{\mathcal{L}}$ to deduce theorems of N 106

6.3 Expressibility in $\mathcal{L}_N$ 109

7 Gödel's Incompleteness Theorem **112**

8 Axiomatic Set Theory **117**

Preface

This course endeavours to build a basic understanding of first order predicate logic, and how models of a first order system relate to mathematical structures.

We are motivated by the following question: is it possible to write a computer program which will tell the truth, the whole truth and nothing but the truth? Of course we cannot expect the program to truthfully tell us next week's winning lottery numbers: we restrict our attention to some precisely defined area. In fact, we will limit ourselves to the natural numbers $\mathbb{N} = \{0, 1, 2 \dots\}$ and to statements which can be made about them. So we are interested in statements such as "2391 is a prime number" and "for any natural numbers m and n, $m + n = n + m$".

There are two types of program we would be interested in.

The best thing we could have would be an *oracle:* a program which could decide if a given statement is true or false. We would supply a statement as input, and the program would respond with an output of "true" or "false". Can we write a program which would do this and which would give the correct answer every time?

Failing that, we could have a program which outputs true statements one after another. We could sit and wait until the program outputs either the statement "2391 is a prime number" or "2391 is not a prime number". If we know that the program will eventually output every true statement, then we know it will output either one or the other of these two, so in fact we can use it as an oracle.

We won't be doing any actual computer programming in this course. However, you should keep at the back of your mind the idea that we are looking for a system which generates true statements in a methodical way. We can see the rules, the method by which the true statements are generated, and so convince ourselves that every statement really is true.

Part of the process will be to develop a way of making the true statements in a precise way which can be stored and manipulated by the computer. To do this we will introduce the notation of *predicate logic*. We will then introduce rules for making deductions, in such a way that if we believe all the assumptions to be true then we must also believe the conclusion to be true. We will then describe a set of assumptions which should be enough to describe the arithmetic of the natural numbers, and we will be able to address the question of whether every true statement about the

natural numbers can indeed be deduced from these assumptions.

How to use these notes

The recommended text book for this course is *L*ogic for Mathemeticians by A.G. Hamilton. At the end of most sections there is a reference to this book (referred to as "text") in square brackets.

These notes are intended to cover all the material needed for this course, but supplementary handouts may also be required. These handouts change from year to year.

Anything that is marked by an asterisk is intended to be extension work and will not be examined. They are included for completeness and/or interest.

Chapter 1

Informal Statement Logic

We have to learn to walk before we can run. We will start with a simpler system, called *statement logic* (or *propositional logic* or *truth table logic*). In the first section we will discuss the meaning of statement logic, learn about truth tables, logical equivalence and logical implication.

1.1 Statements and truth tables

A *statement* is an assertion which is either true or false. For example:

1. "Rodney Hide is a member of the National Party."
2. "The polar ice caps are melting."
3. "20 is divisible by 5."
4. "Every even integer greater than 4 is the sum of two prime numbers."
5. "$10 > 5$ and $2 + 2 = 11$."
6. "if 7 is an even number then 7 is divisible by 2."
7. "If the world is flat then $2 + 2 = 4$."

We only insist that the assertion is true or false, we do not need to know which it is. Notice that the last three examples are formed by combining simpler statements, namely "$10 > 5$", "$2+2 = 11$", "7 is an even number", "7 is divisible by 2", "The world is flat", and "$2 + 2 = 4$". A statement which has been built up in this way is called a *compound* statement, as opposed to a *simple* statement which cannot be split into simpler statements.

Connectives

When we want to build more complicated statements out of simpler ones, we use *connectives*. If A and B are statements, then so are the following:

$$\neg A \quad (A \wedge B) \quad (A \vee B) \quad (A \to B) \quad (A \leftrightarrow B)$$

For example, if A denotes the statement "It is raining", B denotes the statement "The sun is shining" and C denotes the statement "There is a rainbow", then $((A \wedge B) \to C)$ represents the statement "If it is raining and the sun is shining, then there is a rainbow", while $((A \wedge \neg C) \to \neg B)$ represents the statement "If it is raining and there is no rainbow, then the sun is not shining".

We are not usually interested in a particular statement, such as "If it is raining and there is no rainbow, then the sun is not shining", but rather with the way that the simple statements A, B and C were put together to form the compound statement $((A \wedge \neg C) \to \neg B)$. Thus we will consider *statement forms* rather than particular statements, and we will use *statement variables* to represent arbitrary simple statements.

We will use capital letters like A, B and C to represent particular statements, lower case letters like p, q and r to represent statement variables, and script capital letters like $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ to represent statement forms.

Definition 1.1. *A* statement form *is a string of statement variables, brackets and connectives which can be formed using the following rules:*

1. *Every statement variable is a statement form.*
2. *If $\mathcal{A}$ and $\mathcal{B}$ are statement forms, then so are $\neg\mathcal{A}$, $(\mathcal{A} \wedge \mathcal{B})$, $(\mathcal{A} \vee \mathcal{B})$, $(\mathcal{A} \to \mathcal{B})$ and $(\mathcal{A} \leftrightarrow \mathcal{B})$.*

Truth values

We refer to the truth or falsity of a statement as its *truth value*. The truth value of a compound statement depends only on the truth or falsity of the statements from which it was built, according to rules which we can give for each connective. We can summarise them in the following table, which we call a "truth table". In this table, 0 represents "false" and 1 represents "true".

A	B	$\neg A$	$(A \wedge B)$	$(A \vee B)$	$(A \to B)$	$(A \leftrightarrow B)$
0	0	1	0	0	1	1
0	1	1	0	1	1	0
1	0	0	0	1	0	0
1	1	0	1	1	1	1

For more complicated statement forms, we will use the following style of truth tables:

p	q	r	$((p$	$\rightarrow$	$q)$	$\vee$	$(q$	$\rightarrow$	$r))$
0	0	0	0	1	0	1	0	1	0
0	0	1	0	1	0	1	0	1	1
0	1	0	0	1	1	1	1	0	0
0	1	1	0	1	1	1	1	1	1
1	0	0	1	0	0	1	0	1	0
1	0	1	1	0	0	1	0	1	1
1	1	0	1	1	1	1	1	0	0
1	1	1	1	1	1	1	1	1	1

[Text: p 1-8]

1.2 Tautologies, logical equivalence and logical implication

A *tautology* is a statement form which is always true, no matter what truth values we assign to the statement variables it contains.

A *contradiction* is a statement form which is always false, no matter what truth values we assign to the statement variables it contains.

A statement form which is neither a tautology nor a contradiction is said to be *contingent*.

Here are some examples.

1. $(p \vee \neg p)$ is a tautology.
2. $(p \rightarrow p)$ is
3. $(p \wedge \neg p)$ is
4. $(\neg(p \rightarrow q) \wedge \neg p)$ is
5. $(p \rightarrow (q \rightarrow r))$ is
6. $(p \wedge (q \rightarrow p))$ is

To decide whether a statement form is a tautology, a contradiction, or a contingent statement form, we form a truth table for the statement form and check whether the last column contains all 1s, all 0s, or a mixture of 1s and 0s.

Exercises

1. Find truth tables for the following statement forms:
   (a) $(p \vee (p \rightarrow q))$
   (b) $((p \rightarrow q) \rightarrow (\neg p \rightarrow r))$
   (c) $(p \rightarrow (q \rightarrow p))$.
   (d) $((p \vee \neg q) \rightarrow (q \wedge p))$.
   (e) $(((p \vee q) \wedge (q \vee r)) \rightarrow q)$.

   Are these statement forms tautologies, contradictions or contingent?

Logical equivalence and logical implication

Two statement forms are *logically equivalent*, written $\mathcal{A} \Leftrightarrow \mathcal{B}$, if $(\mathcal{A} \leftrightarrow \mathcal{B})$ is a tautology. In other words, $\mathcal{A} \Leftrightarrow \mathcal{B}$ holds if and only if, whenever we assign truth values to the statement variables they contain, $\mathcal{A}$ and $\mathcal{B}$ have the same truth value. [Note that $\mathcal{A} \leftrightarrow \mathcal{B}$ is a compound statement, where as $\mathcal{A} \Leftrightarrow \mathcal{B}$ states a relationship between $\mathcal{A}$ and $\mathcal{B}$, it is a statement *about* $\mathcal{A}$ and $\mathcal{B}$.]

To decide whether or not two statement forms are logically equivalent, we can use truth tables. For example, to show that

$$(p \to q) \Leftrightarrow (\neg q \to \neg p),$$

we build up the following truth table:

p	q	$(p \to q)$	$(\neg q$	$\to$	$\neg p)$
0	0	1	1	1	1
0	1	1	0	1	1
1	0	0	1	0	0
1	1	1	0	1	0

We see that both $(p \to q)$ and $(\neg q \to \neg p)$ have the same truth table.

A statement form $\mathcal{A}$ *logically implies* another statement form $\mathcal{B}$, written $\mathcal{A} \Rightarrow \mathcal{B}$, if $(\mathcal{A} \to \mathcal{B})$ is a tautology. In other words, $\mathcal{A} \Rightarrow \mathcal{B}$ holds if and only if, whenever we assign truth values to the statement variables they contain, if $\mathcal{A}$ is true then $\mathcal{B}$ must also be true.

Again, to decide whether or not $\mathcal{A} \Rightarrow \mathcal{B}$ holds, we can use a truth table. For example, to show that $\neg(p \to q) \Rightarrow \neg q$, we consider the truth table below:

p	q	$\neg$	$(p$	$\to$	$q)$	$\neg q$	
0	0	0	0	1	0	1	
0	1	0	0	1	1	0	
1	0	1	1	0	0	1	$*$
1	1	0	1	1	1	0	

There is only one row in which $\neg(p \to q)$ is true (marked with a $*$), and we note that $\neg q$ is also true in that row.

Example 1.2. Let $\mathcal{A}$ and B be statement forms. Show that $\mathcal{A} \Rightarrow \mathcal{B}$ if and only if $(\mathcal{A} \wedge \mathcal{B}) \Leftrightarrow \mathcal{A}$.

Solution. Suppose first that $\mathcal{A} \Rightarrow \mathcal{B}$. We must show that $(\mathcal{A} \wedge \mathcal{B}) \Leftrightarrow \mathcal{A}$. So suppose we assign truth values to all the statement variables in $\mathcal{A}$ and $\mathcal{B}$. If $\mathcal{A}$ is true then $\mathcal{B}$ must also be true, so $(\mathcal{A} \wedge \mathcal{B})$ is true. On the other hand, if $\mathcal{A}$ is false then $(\mathcal{A} \wedge \mathcal{B})$ is false. So $(\mathcal{A} \wedge \mathcal{B})$ and $\mathcal{A}$ have the same truth value, as required.

Conversely, suppose that $(\mathcal{A} \wedge \mathcal{B}) \Leftrightarrow \mathcal{A}$. We must show that $\mathcal{A} \Rightarrow \mathcal{B}$. So suppose we assign truth values to all the statement variables in $\mathcal{A}$ and $\mathcal{B}$, and that $\mathcal{A}$ is true. Then $(\mathcal{A} \wedge \mathcal{B})$ must also be true, so $\mathcal{B}$ must be true, as required. □

[Text: p 8-10]

Exercises

2. Use truth tables to show that $(p \to (q \wedge r)) \Rightarrow (p \to r)$.
3. Use truth tables to show that $((p \to q) \wedge (q \to r)) \Rightarrow (p \to r)$.
4. Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be statement forms. Show that $(\mathcal{A} \wedge \mathcal{B}) \Rightarrow \mathcal{C}$ if and only if $\mathcal{A} \Rightarrow (\mathcal{B} \to \mathcal{C})$.

1.3 Logical laws

In the previous section we used truth tables to determine when two statement forms are logically equivalent or when one logically implies the other. Another method we can use is to use standard logical equivalences and logical implications from the following list. Each of these can be justified by checking the truth table.

Table: Logical laws

$\neg\neg\mathcal{A}$	$\Leftrightarrow$	$\mathcal{A}$	Double negation
$(\mathcal{A} \wedge \mathcal{B})$	$\Leftrightarrow$	$(\mathcal{B} \wedge \mathcal{A})$	Commutative law
$(\mathcal{A} \vee \mathcal{B})$	$\Leftrightarrow$	$(\mathcal{B} \vee \mathcal{A})$	Commutative law
$(\mathcal{A} \wedge (\mathcal{B} \wedge \mathcal{C}))$	$\Leftrightarrow$	$((\mathcal{A} \wedge \mathcal{B}) \wedge \mathcal{C})$	Associative law
$(\mathcal{A} \vee (\mathcal{B} \vee \mathcal{C}))$	$\Leftrightarrow$	$((\mathcal{A} \vee \mathcal{B}) \vee \mathcal{C})$	Associative law
$(\mathcal{A} \wedge (\mathcal{B} \vee \mathcal{C}))$	$\Leftrightarrow$	$((\mathcal{A} \wedge \mathcal{B}) \vee (\mathcal{A} \wedge \mathcal{C}))$	Distributive law
$(\mathcal{A} \vee (\mathcal{B} \wedge \mathcal{C}))$	$\Leftrightarrow$	$((\mathcal{A} \vee \mathcal{B}) \wedge (\mathcal{A} \vee \mathcal{C}))$	Distributive law
$(\mathcal{A} \wedge \mathcal{A})$	$\Leftrightarrow$	$\mathcal{A}$	Idempotent law
$(\mathcal{A} \vee \mathcal{A})$	$\Leftrightarrow$	$\mathcal{A}$	Idempotent law
$\neg(\mathcal{A} \wedge \mathcal{B})$	$\Leftrightarrow$	$(\neg\mathcal{A} \vee \neg\mathcal{B})$	De Morgan's law
$\neg(\mathcal{A} \vee \mathcal{B})$	$\Leftrightarrow$	$(\neg\mathcal{A} \wedge \neg\mathcal{B})$	De Morgan's law
$(\mathcal{A} \to \mathcal{B})$	$\Leftrightarrow$	$(\neg\mathcal{A} \vee \mathcal{B})$	Implication law
$(\mathcal{A} \to \mathcal{B})$	$\Leftrightarrow$	$\neg(\mathcal{A} \wedge \neg\mathcal{B})$	Implication law
$(\mathcal{A} \vee (\mathcal{B} \wedge \neg\mathcal{B}))$	$\Leftrightarrow$	$\mathcal{A}$	Contradiction law
$\mathcal{A}$	$\Leftrightarrow$	$(\mathcal{A} \wedge (\mathcal{A} \vee \mathcal{B}))$	Absorption law
$\mathcal{A}$	$\Leftrightarrow$	$(\mathcal{A} \vee (\mathcal{A} \wedge \mathcal{B}))$	Absorption law
$(\mathcal{A} \leftrightarrow \mathcal{B})$	$\Leftrightarrow$	$((\mathcal{A} \to \mathcal{B}) \wedge (\mathcal{B} \to \mathcal{A}))$	Equivalence law
$(\mathcal{A} \leftrightarrow \mathcal{B})$	$\Leftrightarrow$	$((\mathcal{A} \wedge \mathcal{B}) \vee (\neg\mathcal{A} \wedge \neg\mathcal{B}))$	Equivalence law

Example 1.3. Show that $(p \to q) \Leftrightarrow (\neg q \to \neg p)$.

Solution. We have

$$\begin{aligned}
(p \to q) &\Leftrightarrow (\neg p \vee q) && \text{(implication)}\\
&\Leftrightarrow (q \vee \neg p) && \text{(commutative)}\\
&\Leftrightarrow (\neg\neg q \vee \neg p) && \text{(double negation)}\\
&\Leftrightarrow (\neg q \to \neg p) && \text{(implication)}
\end{aligned}$$

□

Exercises

5. Use logical laws to show that

   (a) $((p \wedge q) \to r) \Leftrightarrow (p \to (q \to r))$.

   (b) $((p \to r) \wedge (q \to \neg p)) \Leftrightarrow (p \to (r \wedge \neg q))$.

   (c) $((\neg p \wedge \neg q) \to \neg r) \Leftrightarrow (r \to (q \vee p))$.

   [You must make sure that each line follows **directly** from the previous one by one of the laws. It is OK to make two applications to separate parts of the statement form, for example to go from $((\neg p \vee q) \vee (\neg p \vee r))$ to $((p \to q) \vee (p \to r))$ by the implication law (twice). Other combinations of steps, for example going from $(\neg p \vee (q \vee r))$ to $(q \vee (\neg p \vee r))$ by the associative and commutative laws, or going from $(\neg p \vee ((q \vee \neg p) \vee r))$ to $((\neg p \vee q) \vee (\neg p \vee r))$ by the associative law (twice), should be broken into separate steps.]

6. Let $\mathcal{A}$ and $\mathcal{B}$ be statement forms. Show that $\mathcal{A} \approx (\mathcal{A} \wedge (\mathcal{B} \vee \neg\mathcal{B}))$.

1.4 Soundness and adequacy

We now have two methods of proving that two statement forms $\mathcal{A}$ and $\mathcal{B}$ are logically equivalent. As you advance in your mathematical maturity, you should be learning that it is not always appropriate to simply accept that these two approaches are the same "because teacher says so", but rather you should be asking what the justification is: why does the logical laws method give us the same logical equivalences as the truth table method? Indeed, does it really always give us the same answers?

By analogy, consider the problem of solving the quadratic equation

$$x^2 - 8x + 15 = 0.$$

What does it really mean to "solve" the equation? The answer is that it means "find all values of x such that if you work out x^2, subtract $8x$ and add 15, you get 0. However, we often pretend that it means "work out the formula

$$\frac{-(-8) \pm \sqrt{(-8)^2 - 4 \cdot 1 \cdot 15}}{2 \cdot 1}$$

and the two numbers you get are the answer". We have a problem ("find the x's...") and a method (work out the formula...). We must ask whether the formula always gives us the right x's.

There are two desirable properties the method might have:

Soundness: The numbers the method hands us really are solutions of the equation.

Adequacy: If there are any solutions to be found, then the method will find them.

In fact, this method has a third desirable property:

Decidability: If there are no solutions to be found, the method will tell us that fact.

The quadratic formula method tells us that if the number we try to find the square root of is negative, then there are no solutions. In contrast, the factorising method (where we try to spot the factorisation

$$x^2 - 8x + 15 = (x - 3)(x - 5)$$

from which we see that the solutions are 3 and 5) will give us correct answers if we find the factorisation, but does not tell us that there is no factorisation simply because we cannot spot one.

To return to our problem of finding logical equivalences. In this case we have:

Soundness: If the method says that two statement forms are logically equivalent then they really are logically equivalent.

Adequacy: If two statement forms are logically equivalent then we can use the method to demonstrate it.

To avoid ambiguity we should introduce a new symbol so we can distinguish between "$\mathcal{A}$ and $\mathcal{B}$ are logically equivalent" and "the method says that $\mathcal{A}$ and $\mathcal{B}$ are logically equivalent". In fact we will introduce two symbols. For statement forms $\mathcal{A}$ and $\mathcal{B}$, we write $\mathcal{A} \sim \mathcal{B}$ if one of the logical laws states that $\mathcal{A}$ and $\mathcal{B}$ are logically equivalent. In other words if $\mathcal{C}$ occurs in $\mathcal{A}$, $\mathcal{C} \Leftrightarrow \mathcal{D}$ is a logical law and $\mathcal{B}$ is obtained from $\mathcal{A}$ by replacing an occurrence of $\mathcal{C}$ by $\mathcal{D}$, then $\mathcal{A} \sim \mathcal{B}$. We write $\mathcal{A} \approx \mathcal{B}$ if there is a sequence $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ with $\mathcal{A}_1 = \mathcal{A}$, $\mathcal{A}_n = \mathcal{B}$, and $\mathcal{A}_i \sim \mathcal{A}_{i+1}$ for $1 \leqslant i < n$.

To show that the method is sound, we have to show that if $\mathcal{A} \approx \mathcal{B}$, then $\mathcal{A} \Leftrightarrow \mathcal{B}$. To show that it is adequate, we must show that if $\mathcal{A} \Leftrightarrow \mathcal{B}$ then $\mathcal{A} \approx \mathcal{B}$.

1.5 Soundness of $\approx$

First we will show that our method is sound. To do this we must show that if $\mathcal{A} \approx \mathcal{B}$ then $\mathcal{A} \Leftrightarrow \mathcal{B}$. We will first prove some lemmas—preliminary results which are useful in proving our main result.

First we will show that it is OK to apply the rules to parts of formulas. For example, this result lets us use the rule $\neg\neg q \leftrightarrow q$ to deduce that $(\neg\neg q \vee r) \Leftrightarrow (q \vee r)$.

Note that if $\mathcal{A}$ is a statement form that has n different variables, then the truth table for $\mathcal{A}$ must have 2^n rows. We will call the column that is completed last, the *main column*.

Lemma 1.4. *If $\mathcal{A} \sim \mathcal{B}$ then $\mathcal{A} \Leftrightarrow \mathcal{B}$.*

Proof. To prove this we first need to check that each of the logical equivalences in our table of logical laws is correct, which we can do by drawing up a truth table for each one. We leave this as an exercise.

Now suppose that $\mathcal{C} \Leftrightarrow \mathcal{D}$ is one of the logical laws, $\mathcal{C}$ occurs in $\mathcal{A}$ and $\mathcal{B}$ is obtained from $\mathcal{A}$ by replacing an occurrence of $\mathcal{C}$ by $\mathcal{D}$.

We just need to show that $(\mathcal{A} \leftrightarrow \mathcal{B})$ is a tautology. Consider a row in the truth table of $(\mathcal{A} \leftrightarrow \mathcal{B})$, the ith row say. We must show that the main column for $\mathcal{A}$ has the same value as the main column for $\mathcal{B}$ in row i. Now we know that the main columns for $\mathcal{C}$ and $\mathcal{D}$ have the same values (because $\mathcal{C} \Leftrightarrow \mathcal{D}$) and all the other columns that $\mathcal{A}$ is built from (all the ones that are not part of $\mathcal{C}$), are the same as the corresponding columns for $\mathcal{B}$. So $\mathcal{A}$ and $\mathcal{B}$ have the same value in the ith row of the main column, so $(\mathcal{A} \leftrightarrow \mathcal{B})$ has the value 1 in the ith row of the main column. Since this is true in every row, $(\mathcal{A} \leftrightarrow \mathcal{B})$ is a tautology as required. □

Lemma 1.5. *If $\mathcal{A} \Leftrightarrow \mathcal{B}$ and $\mathcal{B} \Leftrightarrow \mathcal{C}$ then $\mathcal{A} \Leftrightarrow \mathcal{C}$.*

Proof. Suppose $\mathcal{A} \Leftrightarrow \mathcal{B}$ and $\mathcal{B} \Leftrightarrow \mathcal{C}$. Suppose we assign truth values to all the proposition variables. Then $\mathcal{A}$ and $\mathcal{B}$ must have the same truth value, and $\mathcal{B}$ and $\mathcal{C}$ must have the same truth value, so $\mathcal{A}$ and $\mathcal{C}$ must have the same truth value. □

Theorem 1.6. *Let $\mathcal{A}$ and $\mathcal{B}$ be statement forms with $\mathcal{A} \approx \mathcal{B}$. Then $\mathcal{A} \Leftrightarrow \mathcal{B}$.*

Proof. We prove by induction on n that if there is a sequence $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ with $\mathcal{A}_i \sim \mathcal{A}_{i+1}$ for $1 \leqslant i < n$ then $\mathcal{A}_1 \Leftrightarrow \mathcal{A}_n$.

Base: The least possible value of n is 2. When $n = 2$ we need to check that $\mathcal{A}_1 \Leftrightarrow \mathcal{A}_2$, which is true by Lemma 1.4.

Inductive step: Suppose $n \geq 2$ and the result holds for all sequences of length n. Suppose we have a sequence

$$\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n, \mathcal{A}_{n+1}$$

with $\mathcal{A}_i \sim \mathcal{A}_{i+1}$ for $1 \leqslant i < n+1$. By inductive hypothesis, $\mathcal{A}_1 \Leftrightarrow \mathcal{A}_n$, and by Lemma 1.4 $\mathcal{A}_n \Leftrightarrow \mathcal{A}_{n+1}$. So by Lemma 1.5 we know that $\mathcal{A}_1 \Leftrightarrow \mathcal{A}_{n+1}$.

Hence, by induction, the result holds for sequences of all finite lengths. □

We turn now to the question of whether or not our method is adequate. But before tackling this question, we will introduce a new concept, that of *disjunctive normal form*. We will show that for

every statement form $\mathcal{A}$ there is a statement form $\mathcal{A}^*$ with $\mathcal{A} \Leftrightarrow \mathcal{A}^*$ such that $\mathcal{A}^*$ is in disjunctive normal form. We will then show that if $\mathcal{A} \Leftrightarrow \mathcal{B}$ then $\mathcal{A}^* = \mathcal{B}^*$.

[Text p 10-13]

Exercises

7. Define a relation $\star$ on $\mathbb{Z} \times \mathbb{Z}$ by declaring that $(m, n) \star (p, q)$ if there is a sequence $(m_1, n_1), (m_2, n_2), \ldots, (m_k, n_k)$ of pairs of integers such that $(m, n) = (m_1, n_1)$, $(p, q) = (m_k, n_k)$ and for each i, $m_{i+1} = m_i + 1$ and $n_{i+1} = n_i + 2m_i + 1$.

   Suppose that $(1, 1) \star (m, n)$. Show that $m^2 = n$.

1.6 Truth Functions and Disjunctive Normal Form

An n-ary *truth function* is a function

$$f : \{0, 1\}^n \to \{0, 1\}.$$

We can represent a truth function by writing down a truth table giving the value of $f(t_1, t_2, \ldots, t_n)$ for all possible combinations of values for t_1, t_2, $\ldots$, t_n (where each t_i is either 0 or 1).

Given a truth function, it is reasonable to ask whether or not there is a statement form $\mathcal{A}$ involving statement variables p_1, p_2, $\ldots$, p_n which has the same truth table. If this is the case, we say that the truth function can be *represented* by $\mathcal{A}$. Notice that statement forms $\mathcal{A}$ and $\mathcal{B}$ (involving statement variables p_1, p_2, $\ldots$, p_n) represent the same truth function if and only if they are logically equivalent.

We will see that this is indeed always possible, in other words that every truth function f can be represented by some statement form $\mathcal{A}$.

First, we will introduce some notation. If $\mathcal{A}_i$ is a statement form for each $i = 1, 2, \ldots, n$ then

$$\Big(\bigvee_{i=1}^{n} \mathcal{A}_i\Big) = ((\cdots(\mathcal{A}_1 \vee \mathcal{A}_2) \vee \mathcal{A}_3)\cdots) \vee \mathcal{A}_n)$$

and

$$\Big(\bigwedge_{i=1}^{n} \mathcal{A}_i\Big) = ((\cdots(\mathcal{A}_1 \wedge \mathcal{A}_2) \wedge \mathcal{A}_3)\cdots) \wedge \mathcal{A}_n)$$

A statement form $\mathcal{A}$ involving the variables $p_1, p_2, \ldots, p_n$ is in *disjunctive normal form* if has the

form

$$\mathcal{A} = (\bigvee_{i=1}^{k} (\bigwedge_{j=1}^{n} \mathcal{B}_{i,j})),$$

where each $\mathcal{B}_{i,j}$ is either p_j or $\neg p_j$.

Theorem 1.7. *Let f be an n-ary truth function. If $f(\mathbf{v}) = 1$ for at least one $\mathbf{v} \in \{0,1\}^n$ then f can be represented by a statement form which is in disjunctive normal form.*

Proof. For each $\mathbf{v} = \langle t_1, t_2, \ldots, t_n \rangle \in \{0,1\}^n$, let $\mathcal{B}_{\mathbf{v}}$ be the statement form $(\bigwedge_{j=1}^{n} \mathcal{B}_{\mathbf{v},j})$, where

$$\mathcal{B}_{\mathbf{v},j} = \begin{cases} p_j & \text{if } t_j = 1, \\ \neg p_j & \text{if } t_j = 0 \end{cases}$$

Notice that if we assign each p_j the truth value t_j, then $\mathcal{B}_{\mathbf{v},j}$ gets the truth value 1, so $\mathcal{B}_{\mathbf{v}}$ also gets the truth value 1.

However, if we change any of the truth values of the p_j's then that $\mathcal{B}_{\mathbf{v},j}$ gets the value 0, so $\mathcal{B}_{\mathbf{v}}$ gets the value 0.

Thus $\mathcal{B}_{\mathbf{v}}$ takes the value 1 on the row of the truth table corresponding to $\mathbf{v}$, and takes the value 0 for every other row.

Now let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be all the n-tuples in $\{0,1\}^n$ for which f takes the value 1. Put

$$\mathcal{A} = (\bigvee_{i=1}^{k} \mathcal{B}_{\mathbf{v}_i}) = (\bigvee_{i=1}^{k} (\bigwedge_{j=1}^{n} \mathcal{B}_{\mathbf{v}_i,j})).$$

Then $\mathcal{A}$ is certainly in disjunctive normal form. Also, $\mathcal{A}$ takes the value 1 if and only if one of the $\mathcal{B}_{\mathbf{v}_i}$'s takes the value 1, which happens on exactly the rows corresponding to $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$, in other words on exactly the rows in which f takes the value 1.

So $\mathcal{A}$ represents f, as required. □

Example 1.8. Find a statement form in disjunctive normal form which represents the following truth function:

t_1	t_2	t_3	$f(t_1, t_2, t_3)$
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	1

Theorem 1.7 shows that for any statement form $\mathcal{A}$, we can find the truth function f represented by $\mathcal{A}$ and (provided $\mathcal{A}$ is not a contradiction) we can find a statement form $\mathcal{A}^*$ in disjunctive normal form which also represents f. Thus, in every row of the truth table, $\mathcal{A}$ and A^* have the same truth value. Thus $\mathcal{A}$ and $\mathcal{A}^*$ are logically equivalent. We call $\mathcal{A}^*$ the disjunctive normal form of $\mathcal{A}$.

Actually, to call $\mathcal{A}^*$ "the" disjunctive normal form of $\mathcal{A}$ is misleading: there are various choices for $\mathcal{A}^*$. For example, consider $\mathcal{A} = p$. This is already in disjunctive normal form if we are only considering the variable p, or we could replace it with $((p \wedge q) \vee (p \wedge \neg q))$ if we have variables p and q, or even

$$(((((p \wedge q) \wedge r) \vee ((p \wedge q) \wedge \neg r)) \vee ((p \wedge \neg q) \wedge r)) \vee ((p \wedge \neg q) \wedge \neg r))$$

with variables p, q and r. So we must fix the variables we are considering. We also have a choice of the order—assuming we are dealing with variables p and q we could have $((p \wedge q) \vee (p \wedge \neg q))$ or $((p \wedge \neg q) \vee (p \wedge q))$, or even $((q \wedge p) \vee (\neg q \wedge p))$.

To avoid this ambiguity we will assume we have fixed some set of variables (say $\{p_1, p_2, \ldots, p_n\}$), that we have fixed some order that the variables should be listed in (say in the order p_1, p_2,$\ldots$,p_n), and then the disjuncts $\left(\bigwedge_{j=1}^{n} \mathcal{B}_{i,j}\right)$ should occur in the order that the rows would appear in the truth table.

Once we have removed any ambiguities like this, we can say that if two statement forms are logically equivalent then they have exactly the same truth table, and therefore they have exactly the same disjunctive normal form, in other words if $\mathcal{A} \Leftrightarrow \mathcal{B}$ then $\mathcal{A}^* = \mathcal{B}^*$.

[Text: p 13-19, and for extra reading there is a discussion on adequate sets of connectives p 19-21]

Exercises

8. Let $\mathcal{A}$ be the statement form $\neg(p \leftrightarrow q)$.
   (a) Find a statement form $\mathcal{B}$ in disjunctive normal form which is logically equivalent to $\mathcal{A}$.
   (b) Show that $\mathcal{A} \approx \mathcal{B}$.
9. Let $\mathcal{A}$ be the statement form $\neg((p \leftrightarrow q) \rightarrow r)$.
   (a) Find a statement form $\mathcal{B}$ in disjunctive normal form which is logically equivalent to $\mathcal{A}$.
   (b) Show that $\mathcal{A} \approx \mathcal{B}$.
10. For every statement form $\mathcal{A}$ using only the connectives $\neg$, $\wedge$ and $\vee$, let $\overline{\mathcal{A}}$ be the statement form we get by replacing every statement variable p with $\neg p$, every $\vee$ with $\wedge$. So for example

we have

$$\begin{aligned}\overline{p} &= \neg p\\ \overline{(p \vee q)} &= (\neg p \wedge \neg q)\\ \overline{(p \vee \neg(q \wedge \neg r))} &= (\neg p \wedge \neg(\neg q \wedge \neg\neg r))\end{aligned}$$

Show that for every such $\mathcal{A}$, $\overline{\mathcal{A}} \Leftrightarrow \neg\mathcal{A}$.

1.7 Adequacy of our method

We return now to the problem of showing that our method of showing two statement forms to be logically equivalent is adequate, in other words that if $\mathcal{A} \Leftrightarrow \mathcal{B}$ then $\mathcal{A} \approx \mathcal{B}$. We won't actually be proving this result, we will just describe how the proof would go and fill in a few of the details.

Suppose that $\mathcal{A} \Leftrightarrow \mathcal{B}$. Our plan is to show that for every $\mathcal{A}$, we have $\mathcal{A} \approx \mathcal{A}^*$, where $\mathcal{A}^*$ is the disjunctive normal form of $\mathcal{A}$. We know that $\mathcal{A} \approx \mathcal{A}^*$ and $\mathcal{B} \approx \mathcal{B}^* = \mathcal{A}^*$, so there are sequences $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ and $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_m$ with $\mathcal{A}_1 = \mathcal{A}$, $\mathcal{A}_n = \mathcal{A}^*$, $\mathcal{A}_i \sim \mathcal{A}_{i+1}$ for $1 \leqslant i < n$ and $\mathcal{B}_1 = \mathcal{B}$, $\mathcal{B}_m = \mathcal{B}^* = \mathcal{A}^*$ and $\mathcal{B}_i \sim \mathcal{B}_{i+1}$ for $1 \leqslant i < m$. But then the sequence

$$\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n, \mathcal{B}_{m-1}, \mathcal{B}_{m-2} \ldots, \mathcal{B}_2, \mathcal{B}_1$$

establishes that $\mathcal{A} \approx \mathcal{B}$, as required.

So we "only" need to show that if that $\mathcal{A} \approx \mathcal{A}^*$ for every $\mathcal{A}$. This certainly seems plausible.

To do this, we would want to be able to replace, for example, p with $(p \wedge (q \vee \neg q))$. Notice that this is not a rule of our system. However, the next lemma shows that we can safely pretend that it is a rule.

Lemma 1.9. *Let $\mathcal{A}$ and $\mathcal{B}$ be statement forms and let $\mathcal{C}$ be a statement form which contains $\mathcal{A}$. Let $\mathcal{D}$ be the statement form obtained from $\mathcal{C}$ by replacing $\mathcal{A}$ with $(\mathcal{A} \wedge (\mathcal{B} \vee \neg\mathcal{B}))$. Then $\mathcal{C} \approx \mathcal{D}$.*

Proof. First we need to show that $\mathcal{A} \approx (\mathcal{A} \wedge (\mathcal{B} \vee \neg\mathcal{B}))$. This is an exercise (Assignment 2 Question 4). So there is a sequence $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ with $\mathcal{A}_1 = \mathcal{A}$, $\mathcal{A}_n = (\mathcal{A} \wedge (\mathcal{B} \vee \neg\mathcal{B}))$ and $\mathcal{A}_i \sim \mathcal{A}_{i+1}$ for $1 \leqslant i < n$. For each i, let $\mathcal{C}_i$ be the statement form obtained from $\mathcal{C}$ by replacing $\mathcal{A}$ with $\mathcal{A}_i$. Then we have $\mathcal{C}_1 = \mathcal{C}$, $\mathcal{C}_n = \mathcal{D}$ and $\mathcal{C}_i \sim \mathcal{C}_{i+1}$ for $1 \leqslant i < n$, so $\mathcal{C} \approx \mathcal{D}$, as required. □

To show that $\mathcal{A} \approx \mathcal{A}^*$, the idea would be to first use the Equivalence Law and the Implication Law to get rid of any $\leftrightarrow$ and $\rightarrow$, leaving only $\neg$, $\wedge$ and $\vee$. Then use De Morgan's Law to make sure that the main connective is $\wedge$ or $\vee$. Finally, repeatedly use the Distributive Law and De Morgan's Law to rewrite the statement form until it has the form $\left(\bigvee_{i=1}^{k} \mathcal{B}_i\right)$, where each $\mathcal{B}_i$ contains no

connectives except $\neg$ and $\wedge$. Finally, if any $\mathcal{B}_i$ does not mention a statement variable p_j, replace it with $(\mathcal{B}_i \wedge (p_j \vee \neg p_j))$ using Lemma 1.9 and expand using the Distributive Law.

While the above may sound plausible, the details would be horrible to work out. For this reason we will abandon this system as too complex and move on in the next chapter to a different system which is much more austere: we are severely limited in what statement forms are allowed and we have only one rule of inference instead of the 18 laws of our present system. The new system we study is not so easy to work with as this one, but it is much easier to prove theorems about.

Exercises

11. (a) Show that $\Leftrightarrow$ and $\approx$ are both equivalence relations on the set of all statement forms.

    (b) Let $[\mathcal{A}]_{\Leftrightarrow}$ and $[\mathcal{A}]_{\approx}$ denote the equivalence classes of $\mathcal{A}$ under $\Leftrightarrow$ and $\approx$ respectively. Express soundness and adequacy of $\approx$ in terms of $[\mathcal{A}]_{\Leftrightarrow}$ and $[\mathcal{A}]_{\approx}$ and $\subseteq$.

12. (a) Define a relation $\bowtie$ on the set of statement forms by declaring that $\mathcal{A} \bowtie \mathcal{B}$ iff there is a sequence $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ with $\mathcal{A}_1 = \mathcal{A}$, $\mathcal{A}_n = \mathcal{B}$ and $\mathcal{A}_i \approx \mathcal{A}_{i+1}$ for $1 \leqslant i < n$. Show that $\mathcal{A} \approx \mathcal{B}$ if and only if $\mathcal{A} \bowtie \mathcal{B}$.

    [The significance of this result is that once we have shown that $\mathcal{A} \approx \mathcal{B}$ we can pretend that this is one of our original laws. This can make it easier to show that two statement forms are $\approx$-equivalent.]

    (b) Show that for any $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$

    $$((\mathcal{A} \vee \mathcal{B}) \wedge \mathcal{C}) \approx ((\mathcal{A} \wedge \mathcal{C}) \vee (\mathcal{B} \wedge \mathcal{C}))$$

    [So by part (a) we can pretend that this law was one of the rules of our system]

1.8 A method that is not adequate*.

Suppose that as a first attempt at choosing an adequate set of logical we fall short, and our method is *not* adequate. How would we know? For example, suppose we only choose the first 16 logical laws, i.e. all but the equivalence laws:

$\neg\neg\mathcal{A}$	$\Leftrightarrow$	$\mathcal{A}$	Double negation
$(\mathcal{A} \wedge \mathcal{B})$	$\Leftrightarrow$	$(\mathcal{B} \wedge \mathcal{A})$	Commutative law
$(\mathcal{A} \vee \mathcal{B})$	$\Leftrightarrow$	$(\mathcal{B} \vee \mathcal{A})$	Commutative law
$(\mathcal{A} \wedge (\mathcal{B} \wedge \mathcal{C}))$	$\Leftrightarrow$	$((\mathcal{A} \wedge \mathcal{B}) \wedge \mathcal{C})$	Associative law
$(\mathcal{A} \vee (\mathcal{B} \vee \mathcal{C}))$	$\Leftrightarrow$	$((\mathcal{A} \vee \mathcal{B}) \vee \mathcal{C})$	Associative law
$(\mathcal{A} \wedge (\mathcal{B} \vee \mathcal{C}))$	$\Leftrightarrow$	$((\mathcal{A} \wedge \mathcal{B}) \vee (\mathcal{A} \wedge \mathcal{C}))$	Distributive law
$(\mathcal{A} \vee (\mathcal{B} \wedge \mathcal{C}))$	$\Leftrightarrow$	$((\mathcal{A} \vee \mathcal{B}) \wedge (\mathcal{A} \vee \mathcal{C}))$	Distributive law
$(\mathcal{A} \wedge \mathcal{A})$	$\Leftrightarrow$	$\mathcal{A}$	Idempotent law
$(\mathcal{A} \vee \mathcal{A})$	$\Leftrightarrow$	$\mathcal{A}$	Idempotent law
$\neg(\mathcal{A} \wedge \mathcal{B})$	$\Leftrightarrow$	$(\neg\mathcal{A} \vee \neg\mathcal{B})$	De Morgan's law
$\neg(\mathcal{A} \vee \mathcal{B})$	$\Leftrightarrow$	$(\neg\mathcal{A} \wedge \neg\mathcal{B})$	De Morgan's law
$(\mathcal{A} \to \mathcal{B})$	$\Leftrightarrow$	$(\neg\mathcal{A} \vee \mathcal{B})$	Implication law
$(\mathcal{A} \to \mathcal{B})$	$\Leftrightarrow$	$\neg(\mathcal{A} \wedge \neg\mathcal{B})$	Implication law
$(\mathcal{A} \vee (\mathcal{B} \wedge \neg\mathcal{B}))$	$\Leftrightarrow$	$\mathcal{A}$	Contradiction law
$\mathcal{A}$	$\Leftrightarrow$	$(\mathcal{A} \wedge (\mathcal{A} \vee \mathcal{B}))$	Absorption law
$\mathcal{A}$	$\Leftrightarrow$	$(\mathcal{A} \vee (\mathcal{A} \wedge \mathcal{B}))$	Absorption law

If we are only allowed to use these laws for our method, then it turns out that method is not adequate. But how do we show this? Let us use some new notation so that we don't get mixed up with our real method. For statement forms $\mathcal{A}$ and $\mathcal{B}$, we write $\mathcal{A} \sim' \mathcal{B}$ if one of the logical laws above, states that $\mathcal{A}$ and $\mathcal{B}$ are logically equivalent. (In other words if $\mathcal{C}$ occurs in $\mathcal{A}$, $\mathcal{C} \Leftrightarrow \mathcal{D}$ is a logical law and $\mathcal{B}$ is obtained from $\mathcal{A}$ by replacing an occurrence of $\mathcal{C}$ by $\mathcal{D}$, then $\mathcal{A} \sim \mathcal{B}$.) We write $\mathcal{A} \approx' \mathcal{B}$ if there is a sequence $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ with $\mathcal{A}_1 = \mathcal{A}$, $\mathcal{A}_n = \mathcal{B}$, and $\mathcal{A}_i \sim' \mathcal{A}_{i+1}$ for $1 \leqslant i < n$.

To show that our system is not adequate, we will show that there is some pair $\mathcal{A}$ and $\mathcal{B}$ such that $\mathcal{A} \Leftrightarrow \mathcal{B}$ but there can be no sequence establishing that $\mathcal{A} \approx' \mathcal{B}$. It is easy to see how we will show that $\mathcal{A} \Leftrightarrow \mathcal{B}$ (with truth tables), but how will we show that there can be no such sequence?

The answer is to find some new relation $\cong$ such that if $\mathcal{A} \approx' \mathcal{B}$ then $\mathcal{A} \cong \mathcal{B}$, and show that for the pair $\mathcal{A}$ and $\mathcal{B}$ that we have chosen, we have $\mathcal{A} \Leftrightarrow \mathcal{B}$ but $\mathcal{A} \ncong \mathcal{B}$, from which it follows that $\mathcal{A} \not\approx' \mathcal{B}$. As with the proof of soundness, to show that $\mathcal{A} \cong \mathcal{B}$ whenever $\mathcal{A} \approx' \mathcal{B}$, it is enough to show that $\mathcal{A} \cong \mathcal{B}$ whenever $\mathcal{A} \sim \mathcal{B}$, and that if $\mathcal{A} \cong \mathcal{B}$ and $\mathcal{B} \cong \mathcal{C}$ then $\mathcal{A} \cong \mathcal{C}$.

The weakness of our set of laws is that there is nothing telling us how to deal with the connective $\leftrightarrow$. This does not mean that statement forms which use $\leftrightarrow$ may not appear, just that there are no rules to change $(\mathcal{A} \leftrightarrow \mathcal{B})$ to logically equivalent forms such as $((\mathcal{A} \wedge \mathcal{B}) \vee (\neg\mathcal{A} \wedge \neg\mathcal{B}))$.

Although this idea may seem obvious, it is surprisingly tricky to capture precisely. However, it can be done.

For each statement form $\mathcal{A}$, let $\overline{\mathcal{A}}$ be the statement form we get by replacing every instance of $\leftrightarrow$ with an $\wedge$. So for example we have $\overline{(p \leftrightarrow (q \vee r))} = (p \wedge (q \vee r))$. For statement forms $\mathcal{A}$ and $\mathcal{B}$, we write $\mathcal{A} \cong \mathcal{B}$ iff $\overline{\mathcal{A}} \Leftrightarrow \overline{\mathcal{B}}$.

Lemma 1.10. *If $\mathcal{A} \sim \mathcal{B}$ then $\mathcal{A} \cong \mathcal{B}$.*

Proof. We must go through each rule and check that it respects $\cong$.

Notice that for any $\mathcal{A}$ we have $\overline{\neg\mathcal{A}} = \neg\overline{\mathcal{A}}$. So we have

$$\begin{aligned}\overline{\neg\neg\mathcal{A}} &= \neg\overline{\neg\mathcal{A}} \\ &= \neg\neg\overline{\mathcal{A}} \\ &\Leftrightarrow \overline{\mathcal{A}},\end{aligned}$$

so $\overline{\neg\neg\mathcal{A}} \Leftrightarrow \overline{\mathcal{A}}$, i.e. $\neg\neg\mathcal{A} \cong \mathcal{A}$.

Similarly $\overline{(\mathcal{A} \wedge \mathcal{B})} = (\overline{\mathcal{A}} \wedge \overline{\mathcal{B}})$, so

$$\begin{aligned}\overline{(\mathcal{A} \wedge \mathcal{B})} &= (\overline{\mathcal{A}} \wedge \overline{\mathcal{B}}) \\ &\Leftrightarrow (\overline{\mathcal{B}} \wedge \overline{\mathcal{A}}) \\ &= \overline{(\mathcal{B} \wedge \mathcal{A})},\end{aligned}$$

so $(\mathcal{A} \wedge \mathcal{B}) \cong (\mathcal{B} \wedge \mathcal{A})$.

We do this for each of the laws, using the facts that $\overline{\neg\mathcal{A}} = \neg\overline{\mathcal{A}}$, $\overline{(\mathcal{A} \wedge \mathcal{B})} = (\overline{\mathcal{A}} \wedge \overline{\mathcal{B}})$, $\overline{(\mathcal{A} \vee \mathcal{B})} = (\overline{\mathcal{A}} \vee \overline{\mathcal{B}})$ and $\overline{(\mathcal{A} \to \mathcal{B})} = (\overline{\mathcal{A}} \to \overline{\mathcal{B}})$.

Finally, as with the soundness proof in the previous section, we must check that the result still holds if we apply a law to a part of a statement form rather than to the whole statement form. So suppose $\mathcal{C}$ contains $\mathcal{A}$, that one of the laws asserts directly that $\mathcal{A} \sim \mathcal{B}$, and that $\mathcal{D}$ is obtained from $\mathcal{C}$ by replacing $\mathcal{A}$ with $\mathcal{B}$. Then $\overline{\mathcal{D}}$ is obtained from $\overline{\mathcal{C}}$ by replacing $\overline{\mathcal{A}}$ with $\overline{\mathcal{B}}$. From the above we know that $\overline{\mathcal{A}} \Leftrightarrow \overline{\mathcal{B}}$ so by Lemma 1.4 we also know that $\overline{\mathcal{C}} \Leftrightarrow \overline{\mathcal{D}}$, as required. □

Lemma 1.11. *If $\mathcal{A} \cong \mathcal{B}$ and $\mathcal{B} \cong \mathcal{C}$ then $\mathcal{A} \cong \mathcal{C}$.*

Proof. Suppose $\mathcal{A} \cong \mathcal{B}$ and $\mathcal{B} \cong \mathcal{C}$. Then $\overline{\mathcal{A}} \Leftrightarrow \overline{\mathcal{B}}$ and $\overline{\mathcal{B}} \Leftrightarrow \overline{\mathcal{C}}$, so by Lemma 1.5 we know that $\overline{\mathcal{A}} \Leftrightarrow \overline{\mathcal{C}}$, i.e. $\mathcal{A} \cong \mathcal{C}$, as required. □

From these two lemmas we can mimic the proof of soundness to show that if $\mathcal{A} \approx' \mathcal{B}$ then $\mathcal{A} \cong \mathcal{B}$. However, consider the statement forms $(p \leftrightarrow q)$ and $((p \wedge q) \vee (\neg p \wedge \neg q))$. We can easily check that

$$(p \leftrightarrow q) \Leftrightarrow ((p \wedge q) \vee (\neg p \wedge \neg q)).$$

However, we have $\overline{(p \leftrightarrow q)} = (p \wedge q)$ which is not logically equivalent to $\overline{((p \wedge q) \vee (\neg p \wedge \neg q))}$. Thus we have

$$(p \leftrightarrow q) \not\cong ((p \wedge q) \vee (\neg p \wedge \neg q)).$$

so the logical equivalence

$$(p \leftrightarrow q) \Leftrightarrow ((p \wedge q) \vee (\neg p \wedge \neg q)).$$

can not be established by our method.

Exercises

13. Show that the method in this section is sound, that is, if $\mathcal{A} \approx' \mathcal{B}$ then $\mathcal{A} \Leftrightarrow \mathcal{B}$.

1.9 Arguments and validity

An *argument* is a sequence of statements such as the following:

> "If it is not raining I go for a walk. If I go for a walk I feed the ducks. I did not feed the ducks today. Therefore, it was raining today."

We want to decide whether the reasoning that this represents is *valid*. In other words, does the conclusion ("it was raining today") follow from the assumptions ("If it is not raining I go for a walk. If I go for a walk I feed the ducks. I did not feed the ducks today.").

Just as we consider statement forms rather than specific statements, we consider argument forms rather than specific arguments.

Definition 1.12. *An* argument form *is a sequence*

$$\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n, \therefore \mathcal{B}$$

where $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ and $\mathcal{B}$ are statement forms. We refer to $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ as the premisses *(or* hypotheses*) of the argument form and $\mathcal{B}$ as the* conclusion *of the argument form.*

For example, we could represent the previous argument with the argument form

$$(\neg p \rightarrow q), (q \rightarrow r), \neg r, \therefore p$$

Definition 1.13. *An argument form*

$$\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n, \therefore \mathcal{B}$$

is invalid *if it is possible to assign truth values to the statement variables in its premisses and conclusion in such a way that* $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ *are all true but* $\mathcal{B}$ *is false. Otherwise it is* valid.

Equivalently, the argument form is valid if and only if

$$\big((\bigwedge_{i=1}^{n} \mathcal{A}_i) \to \mathcal{B}\big)$$

is a tautology.

For example, $p, (p \to q), \therefore q$ is a valid argument form: suppose we assign truth values to p and 1 in such a way that p and $(p \to q)$ are both true. If q were false, then (since p is true) $(p \to q)$ would be false. So q cannot be false, in other words it must be true.

On the other hand, $q, (p \to q), \therefore p$ is invalid. To see this, note that if q is true and p is false then both premisses are true, but the conclusion is false.

To determine whether or not a given argument form is valid, we can use a truth table.

Example 1.14. Determine whether or not the argument form

$$(p \to q), (q \to r), \therefore (p \to r)$$

is valid.

Solution. We construct the following truth table:

p	q	r	$(p \to q)$	$(q \to r)$	$(p \to r)$	
0	0	0	1	1	1	$*$
0	0	1	1	1	1	$*$
0	1	0	1	0	1	
0	1	1	1	1	1	$*$
1	0	0	0	1	0	
1	0	1	0	1	1	
1	1	0	1	0	0	
1	1	1	1	1	1	$*$

We have marked with a $*$ the rows in which both premisses are true. Note that the conclusion is also true in all these rows. Therefore the argument is valid. □

Example 1.15. Determine whether or not the argument form

$$(p \to q), (p \to r), \therefore (q \to r)$$

is valid.

Theorem 1.16. *Let $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$, $\mathcal{B}$ and $\mathcal{C}$ be statement forms. Then the argument form*

$$\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n, \mathcal{B}, \therefore \mathcal{C}$$

is valid if and only if the argument form

$$\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n, \therefore (\mathcal{B} \to \mathcal{C})$$

is valid.

Proof. Suppose first that $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n, \mathcal{B}, \therefore \mathcal{C}$ is valid. Suppose we assign truth values to all the relevant statement variables so that $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ are all true. If $\mathcal{B}$ also happens to be true, then $\mathcal{C}$ must also be true, so $(\mathcal{B} \to \mathcal{C})$ is true. On the other hand, if $\mathcal{B}$ is false then $(\mathcal{B} \to \mathcal{C})$ is true no matter what the truth value of $\mathcal{C}$. Either way, we know that whenever $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ are all true, $(\mathcal{B} \to \mathcal{C})$ is also true. Thus $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n, \therefore (\mathcal{B} \to \mathcal{C})$ is valid.

Conversely, suppose that

$$\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n, \therefore (\mathcal{B} \to \mathcal{C})$$

is valid. Suppose we assign truth values to all the relevant statement variables in such a way that $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ and $\mathcal{B}$ are all true. In particular, $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ are all true, so $(\mathcal{B} \to \mathcal{C})$ must be true. Since $\mathcal{B}$ is true, this implies that $\mathcal{C}$ must also be true. Thus

$$\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n, \mathcal{B} \therefore \mathcal{C}$$

is also valid. □

Exercises

14. Let $\mathcal{A}$ be a statement form which uses only the connectives $\wedge$ and $\vee$ and only the statement variable p. Show that $p \Rightarrow \mathcal{A}$. [Hint: use induction on the complexity of $\mathcal{A}$.]

15. Determine whether or not the argument form $(p \to (q \vee r)), (q \to r), \therefore r$ is valid.

16. (a) Let $\mathcal{A}$ be a statement form which uses only the connectives $\wedge$, $\vee$, $\to$ and $\leftrightarrow$. Prove (by induction on the complexity of $\mathcal{A}$) that if every statement variable is true then $\mathcal{A}$ is true.

    (b) Deduce from (a) that $\{\wedge, \vee, \to, \leftrightarrow\}$ is not adequate.

17. Determine whether or not the argument form $((p \wedge q) \to r), (p \to \neg r), \therefore (p \to \neg q)$ is valid.

Chapter 2

Formal Statement Logic

In our discussion so far we have seen two approaches to proof of logical equivalence. The first was a "semantic" approach, where we give meaning to the statement forms (via truth tables) and to what it means for the statement forms to be logically equivalent (that they have the same truth table). The second is to give a "syntactic" approach where we gave rules for manipulating the statement forms (the logical laws) and gave a syntactic meaning to logical equivalence (that it was possible to use the logical laws to transform one statement form into the other). This raises the question of whether the two approaches are equivalent: whether our syntactic method is sound and adequate to capture the semantic meaning. We will now push this syntactic approach a bit further.

2.1 Post production systems

We will give what is called a *Post production system* for statement logic. To specify such a system S we must provide

- a set of symbols, called the *alphabet* of S;
- rules for deciding which strings of symbols from the alphabet are *well-formed formulas* of S (or *wffs* of S);
- a set of well-formed formulas, which are the *axioms* of S; and
- one or more *rules of inference* in S.

Let $\Sigma \cup \{\mathcal{A}\}$ be a set of wffs. A *derivation of* $\mathcal{A}$ *from* Σ is a sequence $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n$ of wffs of S such that each $\mathcal{B}_i$ is an axiom, an element of Σ, or follows from one or more earlier terms in the sequence by one of the rules of deduction, and such that $\mathcal{B}_n = \mathcal{A}$.

If there exists such a derivation, we write $\Sigma \vdash_S \mathcal{A}$. If $\Sigma = \varnothing$, we just write $\vdash_S \mathcal{A}$. In this case we say that $\mathcal{A}$ is a *theorem* of S.

Examples of Post production systems

1. The formal system $\mathcal{F}$ is as follows:

Alphabet: $\{A, U, B\}$

Wff's: all strings

Single axiom: U

Rules:

- From x infer Ax
- From x infer xB.

Then x is a theorem if and only if x is of the form $\underbrace{AA\ldots A}_{i} U \underbrace{BB\ldots B}_{j}$, which we abbreviate to A^iUB^j for some $i, j \in \mathbb{N}$.

- Consider the derivation: $U, AU, AAU, \ldots A^iU, A^iUB, \ldots A^iUB^j$. Hence A^iUB^j is a theorem.
- Now suppose $\mathcal{B}_1, \ldots, \mathcal{B}_n = x$ is a derivation. We show by induction that each $\mathcal{B}_k$, $1 \leqslant k \leqslant n$ is of the required form.

  For $k = 1$, $\mathcal{B}_1$ is U. So it is of the required form where $i = j = 0$.

  For $k > 1$, if $\mathcal{B}_k = U$ we are done. Otherwise, $\mathcal{B}_k$ was obtained from $\mathcal{B}_l$, $l < k$ via some rule. By inductive hypothesis $\mathcal{B}_l = A^rUB^s$. Then $\mathcal{B}_k = Ar + 1UB^s$ if it was the first rule, and $\mathcal{B}_k = A^rUB^{s+1}$ if it was the second. Either way, $\mathcal{B}_k$ has the required form.

Describe the set of theorems of $\mathcal{F}$. Prove your claim.

2. The MU puzzle

The MU puzzle is from Douglas Hofstadter's *Gödel, Escher, Bach: an Eternal Golden Braid*. We will call this system M.

- The alphabet of M is $\{M, I, U\}$.

- The well-formed formulas are all strings of the form Mx where x is a non-empty string of Is and Us.
- The only axiom is MI.
- There are four rules of inference: here x and y represent any string (possibly empty) of symbols from the alphabet.

  Rule 1: From xI infer xIU.

  Rule 2: From Mx infer Mxx.

  Rule 3: From $MxIIIy$ infer $MxUy$.

  Rule 4: From $MxUUy$ infer Mxy.

For example, $MIUIU$ is a theorem of M. To see this, we provide the following derivation:

1.	MI	Axiom
2.	MII	From 1 by rule 2.
3.	$MIIII$	From 2 by rule 2.
4.	MIU	From 1 by rule 1.
5.	$MIUIU$	From 4 by rule 2.

Some observations:

- Strictly speaking, the derivation is just the sequence

  $$MI, MII, MIIII, MIU, MIUIU.$$

  However, whenever we present a derivation we will also present the number and the justification for each term in this way.
- The derivation could be made shorter by omitting 2 and 3, which are not used later. In practice, we usually make sure that our derivations do not include any redundant steps like this.

Example 2.1. Show that $MIIIUII$ and $MUIU$ are theorems of the post production system M.

The MU puzzle is the following question: is the string MU a theorem of M?

One important point is that there is no systematic way of finding a derivation of a particular wff which we hope might be a theorem. We can systematically consider all possible derivations, and if we ever find a derivation of our favourite wff then we know at once that it is a theorem. However, if we just keep trying to find a derivation, but have not yet succeeded, that does not tell us that we will never find a derivation if we keep trying for long enough.

To show that a wff is *not* a theorem of M, we must find some property which every theorem of M has but which our wff does not have. For example, the following result shows us that MU is not a theorem of M.

Theorem 2.2. *Let x be a theorem of* MIU*. Then the number of Is in x is not a multiple of 3.*

Proof. For a wff $\mathcal{A}$, let $i(\mathcal{A})$ be the number of Is in $\mathcal{A}$.

Let $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n$ be a derivation in MIU. We prove by induction on k that $i(\mathcal{B}_k)$ is congruent to 1 or 2 modulo 3.

Base step: B_1 must be the axiom MI, so $i(\mathcal{B}_1) = 1 \equiv 1 \pmod 3$.

Inductive step: Suppose $1 < k \leqslant n$, and the result holds for each $\mathcal{B}_j$ with $1 \leqslant j < k$. $\mathcal{B}_k$ must be either the axiom MI (in which case the result holds as for the base step) or must follow from some $\mathcal{B}_j$ for $j < k$ by one of the four rules of inference.

Case 1: If $\mathcal{B}_k$ follows from $\mathcal{B}_j$ by Rule 1, then $i(\mathcal{B}_k) = i(\mathcal{B}_j)$, so the result holds for $\mathcal{B}_k$.

Case 2: If $\mathcal{B}_k$ follows from $\mathcal{B}_j$ by Rule 2, then $i(\mathcal{B}_k) = 2i(\mathcal{B}_j)$. If $i(\mathcal{B}_j) \equiv 1 \pmod 3$ then $i(\mathcal{B}_k) \equiv 2 \pmod 3$, and if $i(\mathcal{B}_j) \equiv 2 \pmod 3$ then $i(\mathcal{B}_k) \equiv 4 \equiv 1 \pmod 3$, so either way the result holds for $\mathcal{B}_k$.

Case 3: If $\mathcal{B}_k$ follows from $\mathcal{B}_j$ by Rule 3, then $i(\mathcal{B}_k) = i(\mathcal{B}_j) - 3$, so $i(\mathcal{B}_k) \equiv i(\mathcal{B}_j) \pmod 3$, and so the result holds for $\mathcal{B}_k$.

Case 4: If $\mathcal{B}_k$ follows from $\mathcal{B}_j$ by Rule 4, then $i(\mathcal{B}_k) = i(\mathcal{B}_j)$, so the result holds for $\mathcal{B}_k$.

So in any case, $i(\mathcal{B}_k)$ is congruent to 1 or 2 modulo 3.

Hence, by induction, $i(\mathcal{B}_k)$ is congruent to 1 or 2 modulo 3 for all k. In particular, $i(\mathcal{B}_n)$ is not a multiple of 3. □

Exercises

1. The formal system $\mathcal{G}$ is as follows:

   Alphabet: $\{A, B, M\}$

   Wff's: all strings

   Axiom: M

   Rule: From x infer AxB

   Prove that x is a theorem of $\mathcal{G}$ if and only if x is of the form A^iUB^i for some $i \in \mathbb{N}$.

2.* (a) If k is a power of 2, show that $MI \ldots I$ (k many I's) is a theorem in MIU.

   (b) Show that for each string x consisting of I's and U's only, $\{MxIII\} \vdash_{\mathsf{MIU}} Mx$.

   (c) Show that if 3 does not divide m, then $MI \ldots I$ (m many I's) is a theorem in MIU.

   (d) Use (a) to (c) to prove the following characterisation of the theorems in MIU:
   Let x be any string of symbols I, U. If 3 does not divide the number of I's in x, then Mx is a theorem.

2.2 The system L

We will now consider a Post production system L which is a bit more relevant to our studies.

- The alphabet of L is $\{\rightarrow, \neg, (,)\} \cup \{\, p_n \mid n \in \mathbb{N} \,\}$. We refer to the symbols p_n as *statement variables*.
- The wffs are those strings of symbols which can be built using the following rules:
  - for every $n \in \mathbb{N}$, p_n is a wff
  - if $\mathcal{A}$ is a wff then so is $\neg\mathcal{A}$
  - if $\mathcal{A}$ and $\mathcal{B}$ are wffs then so is $(\mathcal{A} \rightarrow \mathcal{B})$.
- The axioms are given by the following three schemes. Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be wffs. Then

  L1 $(\mathcal{A} \rightarrow (\mathcal{B} \rightarrow \mathcal{A}))$ is an axiom

  L2 $((\mathcal{A} \rightarrow (\mathcal{B} \rightarrow \mathcal{C})) \rightarrow ((\mathcal{A} \rightarrow \mathcal{B}) \rightarrow (\mathcal{A} \rightarrow \mathcal{C})))$ is an axiom

  L3 $((\neg\mathcal{A} \rightarrow \neg\mathcal{B}) \rightarrow ((\neg\mathcal{A} \rightarrow \mathcal{B}) \rightarrow \mathcal{A}))$ is an axiom
- The only rule of inference is modus ponens: from $\mathcal{A}$ and $(\mathcal{A} \rightarrow \mathcal{B})$ infer $\mathcal{B}$.

Example 2.3. Show that **L1**, **L2** and **L3** are tautologies.

The following derivation establishes that
$\vdash_L ((p_1 \rightarrow p_2) \rightarrow (p_1 \rightarrow p_1))$:

1.	$(p_1 \rightarrow (p_2 \rightarrow p_1))$	L1
2.	$((p_1 \rightarrow (p_2 \rightarrow p_1)) \rightarrow ((p_1 \rightarrow p_2) \rightarrow (p_1 \rightarrow p_1)))$	L2
3.	$((p_1 \rightarrow p_2) \rightarrow (p_1 \rightarrow p_1))$	1, 2, MP

Line 1 states that $(p_1 \rightarrow (p_2 \rightarrow p_1))$ is an instance of L1 (with p_1 as $\mathcal{A}$ and p_2 as $\mathcal{B}$). Line 2 states that $((p_1 \rightarrow (p_2 \rightarrow p_1) \rightarrow ((p_1 \rightarrow p_2) \rightarrow (p_1 \rightarrow p_1)))$ is an instance of L2 (with p_1 as $\mathcal{A}$, p_2 as $\mathcal{B}$ and p_1 as $\mathcal{C}$). Line 3 states that $((p_1 \rightarrow p_2) \rightarrow (p_1 \rightarrow p_2))$ follows from lines 1 and 2 by the rule MP (with $(p_1 \rightarrow (p_2 \rightarrow p_1))$ as $\mathcal{A}$ and $((p_1 \rightarrow p_2) \rightarrow (p_1 \rightarrow p_2))$ as $\mathcal{B}$).

Example 2.4. Show that for wffs $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ of L,

$$\{(\mathcal{B} \rightarrow \mathcal{C})\} \vdash_L ((\mathcal{A} \rightarrow \mathcal{B}) \rightarrow (\mathcal{A} \rightarrow \mathcal{C})).$$

Solution. We have the following derivation:

1.	$(\mathcal{B} \to \mathcal{C})$	Hyp.
2.	$((\mathcal{B} \to \mathcal{C}) \to (\mathcal{A} \to (\mathcal{B} \to \mathcal{C})))$	L1
3.	$(\mathcal{A} \to (\mathcal{B} \to \mathcal{C}))$	1, 2, MP
4.	$((\mathcal{A} \to (\mathcal{B} \to \mathcal{C})) \to$ $((\mathcal{A} \to \mathcal{B}) \to (\mathcal{A} \to \mathcal{C})))$	L2
5.	$((\mathcal{A} \to \mathcal{B}) \to (\mathcal{A} \to \mathcal{C}))$	3, 4, MP

□

Example 2.5. Show that for any wff $\mathcal{D}$ of L, $\vdash_L (\mathcal{D} \to \mathcal{D})$.

Solution. Hint: use the following template.

1.	L1
2.	L2
3.	1, 2, MP
4.	L1
5.	3, 4, MP

□

We have chosen an alphabet and wffs that reflect our discussion in Chapter 1. Keep in mind that these symbols and wffs have no presupposed meaning. All results must be derivable through the rules and axioms of L.

Exercises

3. Prove the following results.

   (a) $\{p_2\} \vdash_L (\neg p_1 \to p_2)$.

   (b) $\{(\neg p_1 \to \neg p_2)\} \vdash_L ((\neg p_1 \to p_2) \to p_1)_1$.

   (c) $\{p_2, \neg p_2\} \vdash_L p_1$. [Hint: your proof should fit the following template:

1.		Hyp.
2.		L1
3.		1, 2, MP
4.	$((\neg p_1 \to \neg p_2) \to ((\neg p_1 \to p_2) \to p_1))$	L3
5.		3, 4, MP
6.		Hyp.
7.		L1
8.		
9.		8, 5, MP]

4. Show that $\{(\neg p_1 \to \neg p_2), p_2\} \vdash_L p_1$. [Hint: your proof should fit the following template:

1.		Hyp.
2.	$((\neg p_1 \to \neg p_2) \to ((\neg p_1 \to p_2) \to p_1))$	L3
3.		1, 2, MP
4.		Hyp.
5.		L1
6.		4, 5, MP
7.		6, 3, MP]

5. Let $\mathcal{A}, \mathcal{B}, \mathcal{C} \in \text{form}(L)$. Complete the following derivation showing that $\{(\mathcal{A} \to (\mathcal{B} \to \mathcal{C})), \mathcal{B}, \mathcal{A}\} \vdash_L \mathcal{C}$:

1.	Hyp
2.	Hyp
3.	2, 1, MP
4.	Hyp
5.	4, 3, MP

6. Show that $\{(\mathcal{F} \to (\mathcal{G} \to \mathcal{H})), \mathcal{G}\} \vdash_L (\mathcal{F} \to \mathcal{H})$. [Hint: the derivation has the form

1.	Hyp.
2.	L2
3.	1, 2, MP
4.	Hyp
5.	L1
6.	4, 5, MP
7.	6, 3 MP]

2.3 The Deduction Theorem for L

In this section we will see some more examples of derivations in L, and learn about a tool called the Deduction Theorem which makes them easier to find.

Example 2.6. Show that for any wffs $\mathcal{A}$ and $\mathcal{B}$ we have $\{\mathcal{A}, \neg\mathcal{A}\} \vdash_L \mathcal{B}$.

Solution. Hint: use the following template.

1.		Hyp.
2.		L1
3.		1, 2, MP
4.	$((\neg\mathcal{B} \to \neg\mathcal{A}) \to ((\neg\mathcal{B} \to \mathcal{A}) \to \mathcal{B}))$	L3
5.		3, 4, MP
6.		Hyp.
7.		L1
8.		6, 7, MP
9.		8, 5, MP

□

In contrast to the above derivation, here is a derivation of $\{\neg\mathcal{A}\} \vdash_L (\mathcal{A} \to \mathcal{B})$:

1.	$\neg\mathcal{A}$	Hyp.
2.	$(\neg\mathcal{A} \to (\neg\mathcal{B} \to \neg\mathcal{A}))$	L1
3.	$(\neg\mathcal{B} \to \neg\mathcal{A})$	1, 2, MP.
4.	$((\neg\mathcal{B} \to \neg\mathcal{A}) \to (\mathcal{A} \to (\neg\mathcal{B} \to \neg\mathcal{A})))$	L1
5.	$(\mathcal{A} \to (\neg\mathcal{B} \to \neg\mathcal{A}))$	3, 4, MP
6.	$((\neg\mathcal{B} \to \neg\mathcal{A}) \to ((\neg\mathcal{B} \to \mathcal{A}) \to \mathcal{B}))$	L3
7.	$(((\neg\mathcal{B} \to \neg\mathcal{A}) \to ((\neg\mathcal{B} \to \mathcal{A}) \to \mathcal{B})) \to (\mathcal{A} \to ((\neg\mathcal{B} \to \neg\mathcal{A}) \to ((\neg\mathcal{B} \to \mathcal{A}) \to \mathcal{B})))$	L1
8.	$(\mathcal{A} \to ((\neg\mathcal{B} \to \neg\mathcal{A}) \to ((\neg\mathcal{B} \to \mathcal{A}) \to \mathcal{B}))$	6, 7, MP
9.	$((\mathcal{A} \to (\neg\mathcal{B} \to \neg\mathcal{A}) \to ((\neg\mathcal{B} \to \mathcal{A}) \to \mathcal{B})) \to ((\mathcal{A} \to (\neg\mathcal{B} \to \neg\mathcal{A})) \to (\mathcal{A} \to ((\neg\mathcal{B} \to \mathcal{A}) \to \mathcal{B})))$	L2
10.	$((\mathcal{A} \to (\neg\mathcal{B} \to \neg\mathcal{A})) \to (\mathcal{A} \to ((\neg\mathcal{B} \to \mathcal{A}) \to \mathcal{B})))$	8, 9, MP
11.	$(\mathcal{A} \to (\neg\mathcal{B} \to \mathcal{A}))$	L1
12.	$(\mathcal{A} \to ((\neg\mathcal{B} \to \mathcal{A}) \to \mathcal{B}))$	5, 10, MP
13.	$(\mathcal{A} \to ((\neg\mathcal{B} \to \mathcal{A}) \to \mathcal{B}) \to ((\mathcal{A} \to (\neg\mathcal{B} \to \mathcal{A})) \to (\mathcal{A} \to \mathcal{B})))$	L2
14.	$((\mathcal{A} \to (\neg\mathcal{B} \to \mathcal{A})) \to (\mathcal{A} \to \mathcal{B}))$	12, 13, MP
15.	$(\mathcal{A} \to \mathcal{B})$	11, 14, MP

Observe that derivations with more hypotheses tend to be much simpler and easier to find. The Deduction Theorem let's us exploit this idea.

Theorem 2.7 (The Deduction Theorem for L)**.** *Let $\Sigma \cup \{\mathcal{A}, \mathcal{B}\}$ be a set of wffs of L. Then*

$$\Sigma \cup \{\mathcal{A}\} \vdash_L \mathcal{B} \quad \textit{if and only if} \quad \Sigma \vdash_L (\mathcal{A} \to \mathcal{B}).$$

Proof. First, suppose that $\Sigma \vdash_L (\mathcal{A} \to \mathcal{B})$. Then

$$\Sigma \cup \{\mathcal{A}\} \vdash_L (\mathcal{A} \to \mathcal{B}).$$

We also have $\Sigma \cup \{\mathcal{A}\} \vdash_L \mathcal{A}$, so by modus ponens we have $\Sigma \cup \{\mathcal{A}\} \vdash_L \mathcal{B}$.

Conversely, suppose that $\Sigma \cup \{\mathcal{A}\} \vdash_L \mathcal{B}$. Let $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n$ be a derivation of $\mathcal{B}$ from $\Sigma \cup \{\mathcal{A}\}$. We prove by induction on k that $\Sigma \vdash_L (\mathcal{A} \to \mathcal{B}_k)$ for $1 \leqslant k \leqslant n$.

Base Step: $\mathcal{B}_1$ must be an axiom or a hypothesis: in the latter case, $\mathcal{B}_1$ could be either in Σ or could be $\mathcal{A}$ itself.

Case 1: If $\mathcal{B}_1$ is an axiom then $\Sigma \vdash_L \mathcal{B}_1$. We also have

$$\Sigma \vdash_L (\mathcal{B}_1 \to (\mathcal{A} \to \mathcal{B}_1))$$

since $(\mathcal{B}_1 \to (\mathcal{A} \to \mathcal{B}_1))$ is an instance of L1, so by modus ponens we have $\Sigma \vdash_L (\mathcal{A} \to \mathcal{B}_1)$.

Case 2: If $\mathcal{B}_1 \in \Sigma$ then $\Sigma \vdash_L \mathcal{B}_1$, so as in Case 1 we have $\Sigma \vdash_L (\mathcal{A} \to \mathcal{B}_1)$.

Case 3: If $\mathcal{B}_1 = \mathcal{A}$ then we use our previous example to deduce that $\vdash_L (\mathcal{A} \to \mathcal{B}_1)$, and therefore $\Sigma \vdash_L (\mathcal{A} \to \mathcal{B}_1)$.

Inductive Step: Suppose $1 < k \leqslant n$ and $\Sigma \vdash_L (\mathcal{A} \to \mathcal{B}_i)$ for $1 \leqslant i < k$. There are four cases to consider: $\mathcal{B}_k$ is an axiom, $\mathcal{B}_k \in \Sigma$, $\mathcal{B}_k = \mathcal{A}$ or $\mathcal{B}_k$ follows from two earlier terms in the sequence by modus ponens.

Cases 1–3: Exactly as for the Base Step.

Case 4: Suppose that $\mathcal{B}_k$ follows from two earlier terms by modus ponens. For this to happen, the two earlier terms must have the form $\mathcal{C}$ and $(\mathcal{C} \to \mathcal{B}_k)$. In other words, there are $1 \leqslant i, j < k$ such that $\mathcal{B}_j$ is $(\mathcal{B}_i \to \mathcal{B}_k)$.

By inductive hypothesis,

$$\Sigma \vdash_L (\mathcal{A} \to \mathcal{B}_i)$$

and

$$\Sigma \vdash_L (\mathcal{A} \to (\mathcal{B}_i \to \mathcal{B}_k)).$$

But we also have

$$\Sigma \vdash_L ((\mathcal{A} \to (\mathcal{B}_i \to \mathcal{B}_k)) \to ((\mathcal{A} \to \mathcal{B}_i) \to (\mathcal{A} \to \mathcal{B}_k))),$$

since the latter is an axiom, so by modus ponens we have

$$\Sigma \vdash_L ((\mathcal{A} \to \mathcal{B}_i) \to (\mathcal{A} \to \mathcal{B}_k)),$$

and by modus ponens again we have $\Sigma \vdash_L (\mathcal{A} \to \mathcal{B}_k)$.

Hence, in any case, $\Sigma \vdash_L (\mathcal{A} \to \mathcal{B}_k)$.

Hence, by induction $\Sigma \vdash_L (\mathcal{A} \to \mathcal{B}_k)$ for all k, and in particular $\Sigma \vdash_L (\mathcal{A} \to \mathcal{B}_n)$, i.e.

$$\Sigma \vdash_L (\mathcal{A} \to \mathcal{B}).$$

□

Example 2.8. Show that for any wffs $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ of L, $\vdash_L ((\mathcal{A} \to (\mathcal{B} \to \mathcal{C})) \to (\mathcal{B} \to (\mathcal{A} \to \mathcal{C})))$.

Solution: We will first show that $\{(\mathcal{A} \to (\mathcal{B} \to \mathcal{C})), \mathcal{A}, \mathcal{B}\} \vdash_L \mathcal{C}$. We have the following derivation:

1.	$\{(\mathcal{A} \to (\mathcal{B} \to \mathcal{C}))$	Hyp.
2.	$\mathcal{A}$	Hyp.
3.	$(\mathcal{B} \to \mathcal{C}))$	2,1, MP
4.	$\mathcal{B}$	Hyp.
5.	$\mathcal{C}$	4,3, MP

From this we have

$$\{(\mathcal{A} \to (\mathcal{B} \to \mathcal{C})), \mathcal{A}, \mathcal{B}\} \vdash_L \mathcal{C},$$

so by the Deduction Theorem we have

$$\{(\mathcal{A} \to (\mathcal{B} \to \mathcal{C})), \mathcal{B}\} \vdash_L (\mathcal{A} \to \mathcal{C}).$$

Hence, by the Deduction Theorem again we have

$$\{(\mathcal{A} \to (\mathcal{B} \to \mathcal{C}))\} \vdash_L (\mathcal{B} \to (\mathcal{A} \to \mathcal{C})),$$

and by the Deduction Theorem yet again we have

$$\vdash_L ((\mathcal{A} \to (\mathcal{B} \to \mathcal{C})) \to (\mathcal{B} \to (\mathcal{A} \to \mathcal{C}))),$$

as required. □

Here is another example:

Example 2.9. Show that for any wffs $\mathcal{A}$ and $\mathcal{B}$ of L, $\vdash_L ((\mathcal{A} \to (\mathcal{A} \to \mathcal{B})) \to (\mathcal{A} \to \mathcal{B}))$.

Solution. We show first that $\{(\mathcal{A} \to (\mathcal{A} \to \mathcal{B})), \mathcal{A}\} \vdash_L \mathcal{B}$. We have the derivation

1.	$(\mathcal{A} \to (\mathcal{A} \to \mathcal{B}))$	Hyp.
2.	$\mathcal{A}$	Hyp.
3.	$(\mathcal{A} \to \mathcal{B})$	2, 1, MP
4.	$\mathcal{B}$	2, 3, MP

From this we have

$$\{(\mathcal{A} \to (\mathcal{A} \to \mathcal{B})), \mathcal{A}\} \vdash_L \mathcal{B},$$

so by DT we have

$$\{(\mathcal{A} \to (\mathcal{A} \to \mathcal{B}))\} \vdash_L \{(A \to \mathcal{B}).$$

Hence by DT again we have

$$\vdash_L ((\mathcal{A} \to (\mathcal{A} \to \mathcal{B})) \to (\mathcal{A} \to \mathcal{B})).$$

□

Compare the above proof with the following derivation, which is entirely within L:

1.	$(\mathcal{A} \to ((\mathcal{A} \to \mathcal{B}) \to \mathcal{A}))$	L1
2.	$((\mathcal{A} \to ((\mathcal{A} \to \mathcal{B}) \to \mathcal{A}))$ $\to ((\mathcal{A} \to ((\mathcal{A} \to \mathcal{B}) \to (\mathcal{A} \to \mathcal{A})))$	L2
3.	$((\mathcal{A} \to ((\mathcal{A} \to \mathcal{B}) \to (\mathcal{A} \to \mathcal{A}))$	1, 2, MP
4.	$((\mathcal{A} \to (\mathcal{A} \to \mathcal{B})) \to ((\mathcal{A} \to \mathcal{A}) \to (\mathcal{A} \to \mathcal{B})))$	L2
5.	$((\mathcal{A} \to (\mathcal{A} \to \mathcal{B})) \to ((\mathcal{A} \to \mathcal{A}) \to (\mathcal{A} \to \mathcal{B}))$ $\to (((\mathcal{A} \to (\mathcal{A} \to \mathcal{B})) \to (\mathcal{A} \to \mathcal{A}))$ $\to (\mathcal{A} \to (\mathcal{A} \to \mathcal{B})) \to (\mathcal{A} \to \mathcal{B}))))$	L2
6.	$(((\mathcal{A} \to (\mathcal{A} \to \mathcal{B})) \to (\mathcal{A} \to \mathcal{A}))$ $\to (\mathcal{A} \to (\mathcal{A} \to \mathcal{B})) \to (\mathcal{A} \to \mathcal{B}))$	4, 5, MP
7.	$((\mathcal{A} \to (\mathcal{A} \to \mathcal{B})) \to (\mathcal{A} \to \mathcal{B}))$	3, 6, MP

Of course, it is one thing to check that the above derivation stays within the rules of the game of L, but it is quite another to find a derivation like the one above. In fact, the proof of the Deduction Theorem actually gives us a method of finding the derivation, or to be precise a method of converting the derivation of $\mathcal{B}$ from $\{(\mathcal{A} \to (\mathcal{A} \to \mathcal{B})), \mathcal{A}\}$ into a derivation of

$$((\mathcal{A} \to (\mathcal{A} \to \mathcal{B})) \to (\mathcal{A} \to \mathcal{B}))$$

from $\varnothing$.

Exercises

7. Let $\mathcal{A}$ and $\mathcal{B}$ be wffs of L.

   (a) Complete the following derivation showing that $\{(\neg\mathcal{A} \to \neg\mathcal{B}), \mathcal{B}\} \vdash_L \mathcal{A}$.

1.	$((\neg\mathcal{A} \to \neg\mathcal{B}) \to ((\neg\mathcal{A} \to \mathcal{B}) \to \mathcal{A}))$	L3
2.		Hyp
3.		2, 1, MP
4.		Hyp.
5.		L1
6.		4, 5, MP
7.		6, 3, MP

   (b) Deduce that $\vdash_L (\mathcal{B} \to ((\neg\mathcal{A} \to \neg\mathcal{B}) \to \mathcal{A}))$.

8. Let $\mathcal{A}$ be a wff of L. Show that $\vdash_L (\neg\neg\mathcal{A} \to \mathcal{A})$.

9. Let $\mathcal{A}$ and $\mathcal{B}$ be wffs of L. Use Example 2.6 and the Deduction Theorem to deduce that $\vdash_L (\neg\mathcal{B} \to (\mathcal{B} \to \mathcal{A}))$.

10. Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be wffs of L.

    (a) Find a derivation showing that $\{(\mathcal{A} \to \mathcal{B}), (\mathcal{B} \to \mathcal{C}), \mathcal{A}\} \vdash_L \mathcal{C}$.

    (b) Use the Deduction Theorem to deduce that $\vdash_L ((\mathcal{A} \to \mathcal{B}) \to ((\mathcal{B} \to \mathcal{C}) \to (\mathcal{A} \to \mathcal{C})))$.

    (c) Find a derivation showing that $\{(\mathcal{A} \to \mathcal{B}), (\mathcal{B} \to \mathcal{C})\} \vdash_L (\mathcal{A} \to \mathcal{C})$.

2.4 Truth assignments

We now move onto the issue of the semantic interpretation of L. We know the definition of $\Sigma \vdash_L \mathcal{A}$, but what does it actually *mean*? Why do we care?

We will show that, for statement forms $\mathcal{A}_1$, $\mathcal{A}_2$, ..., $\mathcal{A}_n$ and $\mathcal{B}$, the argument form

$$\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n, \therefore \mathcal{B}$$

is valid if and only if

$$\{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n\} \vdash_L \mathcal{B}.$$

In fact we will extend our notion of valid arguments to allow arbitrary (possibly infinite) sets of premisses.

We use $\text{form}(L)$ to denote the set of all wffs of L.

Definition 2.10. *A* truth assignment *is a function* $v : \text{form}(L) \to \{0, 1\}$ *such that for any* $\mathcal{A}, \mathcal{B} \in \text{form}(L)$, $v(\neg\mathcal{A}) = 1 - v(\mathcal{A})$, *and*

$$v((\mathcal{A} \to \mathcal{B})) = \begin{cases} 0 & \text{if } v(\mathcal{A}) = 1 \text{ and } v(\mathcal{B}) = 0 \\ 1 & \text{otherwise} \end{cases}$$

In other words, a truth assignment is a way of assigning truth values to the statement variables, and extending these truth values to all the wffs of L.

Definition 2.11. *A wff* $\mathcal{A}$ *is a* tautology *if* $v(\mathcal{A}) = 1$ *for every truth assignment* v.

Definition 2.12. *Let* $\Sigma \cup \{\mathcal{A}\} \subseteq \text{form}(L)$. *We say that* Σ entails $\mathcal{A}$, *written* $\Sigma \models \mathcal{A}$, *if for every truth assignment* v *such that* $v(\mathcal{B}) = 1$ *for all* $\mathcal{B} \in \Sigma$, $v(\mathcal{A}) = 1$.

Notice that the argument form

$$\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n, \therefore \mathcal{B}$$

is a valid argument form if and only if

$$\{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n\} \models \mathcal{B}.$$

Our goal now, in showing that our system L represents our intentions in terms of the meaning of all the symbols, is to show that for any $\Sigma \cup \{\mathcal{A}\} \subseteq \text{form}(L)$,

$$\Sigma \vdash_L \mathcal{A} \quad \text{if and only if} \quad \Sigma \models \mathcal{A}.$$

There are two parts to this. First, we must show that L is *sound*, in other words that if $\Sigma \vdash_L \mathcal{A}$ then $\Sigma \models \mathcal{A}$. Secondly, we must show that it is *adequate*, in other words that if $\Sigma \models \mathcal{A}$ then $\Sigma \vdash_L \mathcal{A}$.

Lemma 2.13. *All the axioms of L are tautologies.*

Proof. Exercise. □

Theorem 2.14 (The Soundness Theorem for L)**.** *Let $\Sigma \cup \{\mathcal{A}\}$ be a set of wffs of L. If $\Sigma \vdash_L \mathcal{A}$ then $\Sigma \models \mathcal{A}$.*

Proof. Let $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n$ be a derivation of $\mathcal{A}$ from Σ. We prove by induction on k that $\Sigma \models \mathcal{B}_k$ for $1 \leqslant k \leqslant n$.

Base step: $\mathcal{B}_1$ must be either an axiom or an element of Σ.

Let v be a truth assignment such that $v(\mathcal{C}) = 1$ for all $\mathcal{C} \in \Sigma$.

If $\mathcal{B}_1$ is an axiom then it is a tautology so $v(\mathcal{B}_1) = 1$. Otherwise it is in Σ, so again we have $v(\mathcal{B}_1) = 1$. Thus $\Sigma \models \mathcal{B}_1$.

Inductive step Suppose $1 < k \leqslant n$ and $\Sigma \models \mathcal{B}_i$ for $1 \leqslant i < k$. $\mathcal{B}_k$ must be an axiom, in Σ or follow from two earlier terms by modus ponens.

In the first two cases, we know that $\Sigma \models \mathcal{B}_k$ as in the base step.

So suppose that $\mathcal{B}_k$ follows from two earlier terms by modus ponens. In other words, there are i, j with $1 \leqslant i, j < k$ such that $\mathcal{B}_j = (\mathcal{B}_i \to \mathcal{B}_k)$.

Let v be a truth assignment with $v(\mathcal{C}) = 1$ for all $\mathcal{C} \in \Sigma$.

By inductive hypothesis we know that $\Sigma \models \mathcal{B}_i$ and $\Sigma \models (\mathcal{B}_i \to \mathcal{B}_k)$, so $v(\mathcal{B}_i) = 1$ and $v(\mathcal{B}_i \to \mathcal{B}_k) = 1$.

Thus we must have $v(\mathcal{B}_k) = 1$. Hence $\Sigma \models \mathcal{B}_k$.

Hence, by induction, $\Sigma \models \mathcal{B}_k$ for $1 \leqslant k \leqslant n$, and in particular $\Sigma \models \mathcal{B}_n$, i.e. $\Sigma \models \mathcal{A}$. □

Theorem 2.15. *Every theorem of L is a tautology.*

Proof. Let $\mathcal{A}$ be a theorem of L. Then $\varnothing \vdash_L \mathcal{A}$, so $\varnothing \models \mathcal{A}$. Let v be a truth assignment. Then, vacuously, $v(\mathcal{B}) = 1$ for every $\mathcal{B} \in \varnothing$, so $v(\mathcal{A}) = 1$. Thus $\mathcal{A}$ is a tautology, as claimed. □

Exercises

11. For $\Sigma \subseteq \text{form}(L)$, let $\mathcal{E}(\Sigma) = \{\, \mathcal{A} \in \text{form}(L) \mid \Sigma \models \mathcal{A} \,\}$. Prove the following properties of $\mathcal{E}$:
    (a) $\Sigma \subseteq \mathcal{E}(\Sigma)$;
    (b) if $\Sigma \subseteq \Pi$ then $\mathcal{E}(\Sigma) \subseteq \mathcal{E}(\Pi)$; and
    (c) $\mathcal{E}(\mathcal{E}(\Sigma)) = \mathcal{E}(\Sigma)$.

12. For $\Sigma \subseteq \text{form}(L)$, let $\text{Con}(\Sigma) = \{\, \mathcal{A} \in \text{form}(L) \mid \Sigma \vdash_L \mathcal{A} \,\}$. Prove the following properties of Con:

    (a) $\Sigma \subseteq \text{Con}(\Sigma)$; [Note: this is not a trick question, just very easy.]

    (b) if $\Sigma \subseteq \Pi$ then $\text{Con}(\Sigma) \subseteq \text{Con}(\Pi)$; and

    (c) $\text{Con}(\text{Con}(\Sigma)) = \text{Con}(\Sigma)$.

2.5 Adequacy of L

Our next goal is to prove that the formal system L is adequate, in other words that if $\Sigma \cup \{\mathcal{A}\} \subseteq \text{form}(L)$ and $\Sigma \models \mathcal{A}$ then $\Sigma \vdash_L \mathcal{A}$. To do this we will introduce two new properties that a set of wffs of L may or may not have: *completeness* and *satisfiability*.

Definition 2.16. *Let $\Sigma \subseteq \text{form}(L)$. We say that Σ is* consistent *if there is no wff $\mathcal{A}$ such that $\Sigma \vdash_L \mathcal{A}$ and $\Sigma \vdash_L \neg\mathcal{A}$. We say that Σ is* complete *if for every $\mathcal{A} \in \text{form}(L)$ we have either $\mathcal{A} \in \Sigma$ or $\neg\mathcal{A} \in \Sigma$.*

Definition 2.17. *A set $\Sigma \subseteq \text{form}(L)$ is* satisfied *by a truth assignment $v : \text{form}(L) \to \{0,1\}$ if $v(\mathcal{B}) = 1$ for all $\mathcal{B} \in \Sigma$. Σ is* satisfiable *if there is some truth assignment v which satisfies Σ.*

We have two closely related notions here, a syntactic notion of consistency and a semantic notion of satisfiability. In fact, we will end up showing that they are equivalent: Σ is consistent if and only if it is satisfiable.

- The plan is to show that if Σ is both complete and consistent then the function $f_\Sigma : \text{form}(L) \to \{0,1\}$ defined by
$$f_\Sigma(\mathcal{A}) = \begin{cases} 1 & \text{if } \mathcal{A} \in \Sigma \\ 0 & \text{if } \mathcal{A} \notin \Sigma \end{cases}$$
is actually a truth assignment (so Σ is satisfiable because f_Σ satisfies Σ).
- We will then show that if $\Sigma \nvdash_L \mathcal{A}$ then there is a complete consistent set Π with $\Sigma \cup \{\neg\mathcal{A}\} \subseteq \Pi$.
- But then f_Π is a truth assignment with $f_\Pi(\mathcal{B}) = 1$ for all $\mathcal{B} \in \Sigma$ but $f_\Pi(\mathcal{A}) = 0$, which shows that $\Sigma \not\models \mathcal{A}$.
- Hence, by contraposition, if $\Sigma \models \mathcal{A}$ then $\Sigma \vdash_L \mathcal{A}$.

Lemma 2.18. *$\varnothing$ is consistent.*

Proof. Suppose not, i.e. suppose that there is some $\mathcal{A}$ such that $\vdash_L \mathcal{A}$ and $\vdash_L \neg\mathcal{A}$. Then, by the Soundness Theorem for L, both $\mathcal{A}$ and $\neg\mathcal{A}$ would have to be tautologies, and this is impossible. □

In Exercise 2.7 we will see that there is a complete Σ which is not consistent. Also $\varnothing$ is consistent but certainly not complete. So the two notions are independent.

Lemma 2.19. *Let $\Sigma \subseteq \mathrm{form}(L)$. If Σ is not consistent then $\Sigma \vdash_L \mathcal{A}$ for every $\mathcal{A} \in \mathrm{form}(L)$.*

Proof. Suppose that Σ is inconsistent. Then there is some $\mathcal{B}$ such that $\Sigma \vdash_L \mathcal{B}$ and $\Sigma \vdash_L \neg\mathcal{B}$.

Since $\Sigma \vdash_L \mathcal{B}$ and $(\mathcal{B} \to (\neg\mathcal{A} \to \mathcal{B}))$ is an axiom, $\Sigma \vdash_L (\neg\mathcal{A} \to \mathcal{B})$.

Similarly, since $\Sigma \vdash_L \neg\mathcal{B}$ we have $\Sigma \vdash_L (\neg\mathcal{A} \to \neg\mathcal{B})$.

Now

$$((\neg\mathcal{A} \to \neg\mathcal{B}) \to ((\neg\mathcal{A} \to \mathcal{B}) \to \mathcal{A}))$$

is an axiom, so by modus ponens $\Sigma \vdash_L ((\neg\mathcal{A} \to \mathcal{B}) \to \mathcal{A})$, and by modus ponens again we have $\Sigma \vdash_L \mathcal{A}$, as required. □

Note this is the argument used for Example 2.6.

Lemma 2.20. *Let $\Sigma \cup \{\mathcal{A}\} \subseteq \mathrm{form}(L)$. If $\Sigma \nvdash_L \mathcal{A}$ then $\Sigma \cup \{\neg\mathcal{A}\}$ is consistent.*

Proof. We prove the contrapositive: if $\Sigma \cup \{\neg\mathcal{A}\}$ is inconsistent then $\Sigma \vdash_L \mathcal{A}$.

So suppose that $\Sigma \cup \{\neg\mathcal{A}\}$ is inconsistent. Then there is some $\mathcal{B}$ such that $\Sigma \cup \{\neg\mathcal{A}\} \vdash_L \mathcal{B}$ and $\Sigma \cup \{\neg\mathcal{A}\} \vdash_L \neg\mathcal{B}$. By the Deduction Theorem, we have $\Sigma \vdash_L (\neg\mathcal{A} \to \mathcal{B})$ and $\Sigma \vdash_L (\neg\mathcal{A} \to \neg\mathcal{B})$. We also have

$$\Sigma \vdash_L ((\neg\mathcal{A} \to \neg\mathcal{B}) \to ((\neg\mathcal{A} \to \mathcal{B}) \to \mathcal{A})),$$

since the latter is an axiom. So, by modus ponens we have $\Sigma \vdash_L ((\neg\mathcal{A} \to \mathcal{B}) \to \mathcal{A})$, and by modus ponens again we have $\Sigma \vdash_L \mathcal{A}$, as required. □

Lemma 2.21. *Let $\Sigma \subseteq \mathrm{form}(L)$ be complete and consistent. For any $\mathcal{A} \in \mathrm{form}(L)$, we have $\Sigma \vdash_L \mathcal{A}$ iff $\mathcal{A} \in \Sigma$.*

Proof. Suppose first that $\mathcal{A} \in \Sigma$. Then certainly we have $\Sigma \vdash_L \mathcal{A}$.

Conversely, suppose that $\Sigma \vdash_L \mathcal{A}$. We must show that $\mathcal{A} \in \Sigma$, so suppose for a contradiction that $\mathcal{A} \notin \Sigma$. Then, by completeness we have $\neg\mathcal{A} \in \Sigma$, so $\Sigma \vdash_L \neg\mathcal{A}$, and thus Σ is inconsistent, contradicting our assumption about Σ. So we must have $\mathcal{A} \in \Sigma$ as required. □

Theorem 2.22. *If $\Sigma \subseteq \mathrm{form}(L)$ is complete and consistent then the function $f_\Sigma : \mathrm{form}(L) \to \{0, 1\}$ defined by*

$$f_\Sigma(\mathcal{A}) = \begin{cases} 1 & \text{if } \mathcal{A} \in \Sigma \\ 0 & \text{if } \mathcal{A} \notin \Sigma \end{cases}$$

is a truth assignment.

Proof. We need to check that, for arbitrary wff $\mathcal{A}, \mathcal{B}$,

- if $f_\Sigma(\mathcal{A}) = 1$ then $f_\Sigma(\neg\mathcal{A}) = 0$;
- if $f_\Sigma(\mathcal{A}) = 0$ then $f_\Sigma(\neg\mathcal{A}) = 1$;
- if $f_\Sigma(\mathcal{A}) = 1$ and $f_\Sigma(\mathcal{B}) = 0$ then $f_\Sigma((\mathcal{A} \to \mathcal{B})) = 0$; and
- if $f_\Sigma(\mathcal{A}) = 0$ or $f_\Sigma(\mathcal{B}) = 1$ then $f_\Sigma((\mathcal{A} \to \mathcal{B})) = 1$.

Suppose that $f_\Sigma(\mathcal{A}) = 1$. Then $\mathcal{A} \in \Sigma$, so $\Sigma \vdash_L \mathcal{A}$, so $\Sigma \nvdash_L \neg\mathcal{A}$ (since Σ is consistent), so $\neg\mathcal{A} \notin \Sigma$, so $f_\Sigma(\neg\mathcal{A}) = 0$.

Suppose instead that $f_\Sigma(\mathcal{A}) = 0$. Then $\mathcal{A} \notin \Sigma$ so $\neg\mathcal{A} \in \Sigma$ (since Σ is complete) so $f_\Sigma(\neg\mathcal{A}) = 1$.

Next suppose $f_\Sigma(\mathcal{A}) = 1$ and $f_\Sigma(\mathcal{B}) = 0$. Suppose, for a contradiction, that $f_\Sigma((\mathcal{A} \to \mathcal{B})) \neq 0$, i.e. that $f_\Sigma((\mathcal{A} \to \mathcal{B})) = 1$. Then we have $\mathcal{A} \in \Sigma$ and $(\mathcal{A} \to \mathcal{B}) \in \Sigma$, so $\Sigma \vdash_L \mathcal{A}$ and $\Sigma \vdash_L (\mathcal{A} \to \mathcal{B})$. But then, by modus ponens, $\Sigma \vdash_L \mathcal{B}$, and so $\mathcal{B} \in \Sigma$, so $f_\Sigma(\mathcal{B}) = 1$, a contradiction. Thus $f_\Sigma((\mathcal{A} \to \mathcal{B})) = 0$, as required.

Suppose instead that $f_\Sigma(\mathcal{A}) = 0$. Then $\mathcal{A} \notin \Sigma$, so $\neg\mathcal{A} \in \Sigma$ (by completeness), so $\Sigma \vdash_L \neg\mathcal{A}$. We also know that

$$\vdash_L (\neg\mathcal{A} \to (\mathcal{A} \to \mathcal{B})),$$

so

$$\Sigma \vdash_L (\neg\mathcal{A} \to (\mathcal{A} \to \mathcal{B})),$$

so by modus ponens we have $\Sigma \vdash_L (\mathcal{A} \to \mathcal{B})$.

Thus $(\mathcal{A} \to \mathcal{B}) \in \Sigma$, so $f_\Sigma((\mathcal{A} \to \mathcal{B})) = 1$ as required.

Finally, suppose $f_\Sigma(\mathcal{B}) = 1$. Then $\mathcal{B} \in \Sigma$ so $\Sigma \vdash_L \mathcal{B}$. We also have $\Sigma \vdash_L (\mathcal{B} \to (\mathcal{A} \to \mathcal{B}))$ since this is an axiom, so by modus ponens we have $\Sigma \vdash_L (\mathcal{A} \to \mathcal{B})$, and thus $f_\Sigma((\mathcal{A} \to \mathcal{B})) = 1$, as required. □

Exercises

13. Let $\mathcal{A}, \mathcal{B}, \mathcal{C} \in \text{form}(L)$.

    (a) Complete the following derivation showing that $\{(\mathcal{A} \to (\mathcal{B} \to \mathcal{C})), \mathcal{B}, \mathcal{A}\} \vdash_L \mathcal{C}$:

1.	Hyp
2.	Hyp
3.	2, 1, MP
4.	Hyp
5.	4, 3, MP

(b) Deduce that $\vdash_L ((\mathcal{A} \to (\mathcal{B} \to \mathcal{C})) \to (\mathcal{B} \to (\mathcal{A} \to \mathcal{C})))$.

14. Let $\mathcal{A}$ and $\mathcal{B}$ be wffs of L.

(a) Find a derivation showing that $\{(\neg\mathcal{A} \to \neg\mathcal{B}), \mathcal{B}\} \vdash_L \mathcal{A}$.

(b) Use the Deduction Theorem to deduce that $\vdash_L ((\neg\mathcal{A} \to \neg\mathcal{B}) \to (\mathcal{B} \to \mathcal{A}))$.

(c) Find a derivation showing that $\{(\neg\mathcal{A} \to \neg\mathcal{B})\} \vdash_L (\mathcal{B} \to \mathcal{A})$.

2.6 Countable sets

Our proof that every consistent set can be extended to a complete consistent set depends on the fact that the set form(L) is countable. In this section we will give a brief overview of countable sets.

Definition 2.23. *A set X is* countable *if $X = \varnothing$ or we can enumerate X as $\{ x_n \mid n \in \mathbb{N} \}$ (possibly with repetitions). In other words, X is countable if $X = \varnothing$ or there is an onto map $g : \mathbb{N} \mapsto X$ (given by $n \mapsto x_n$).*

The idea is that countable sets are "small": they may be infinite, but they are "only just" infinite. This being so the following result is obvious:

Proposition 2.24. *If X is countable and $Y \subseteq X$ then Y is countable.*

Proof. If $Y = \varnothing$ then Y is countable, so we assume $Y \neq \varnothing$ (and therefore certainly $X \neq \varnothing$). Then we can enumerate X as $\{ x_n \mid n \in \mathbb{N} \}$. Since $Y \neq \varnothing$, there is at least one $y \in Y$. Put

$$y_n = \begin{cases} x_n & \text{if } x_n \in Y \\ y & \text{if } x_n \notin Y \end{cases}$$

Then $Y = \{ y_n \mid n \in \mathbb{N} \}$, so Y is countable, as required. □

Proposition 2.25. *If X and Y are sets with X countable and there is a bijection from X to Y then Y is also countable.*

Proof. Again we assume X is non-empty, so we can enumerate it as $\{ x_n \mid n \in \mathbb{N} \}$. Let $f : X \to Y$ be a bijection. Then we can enumerate Y as $\{ f(x_n) \mid n \in \mathbb{N} \}$, so Y is countable as required. □

Proposition 2.26. *If X and Y are countable then so is $X \cup Y$.*

Proof. We assume X and Y are both non-empty (else the result is trivial). So we can enumerate X as $\{ x_n \mid n \in \mathbb{N} \}$ and Y as $\{ y_n \mid n \in \mathbb{N} \}$. So then we have $X \cup Y = \{ z_n \mid n \in \mathbb{N} \}$, where

$$z_n = \begin{cases} x_{n/2} & \text{if } n \text{ is even} \\ y_{(n-1)/2} & \text{if } n \text{ is odd} \end{cases}$$

□

Lemma 2.27. *The function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ given by $f(m,k) = 2^m(2k+1) - 1$ is a bijection.*

Proof. We show first that f is 1–1. So suppose $f(m_1, k_1) = f(m_2, k_2)$. Without loss of generality (WLOG), $m_1 \leqslant m_2$. Now we have

$$2^{m_1}(2k_1 + 1) - 1 = 2^{m_2}(2k_2 + 1) - 1$$

so

$$2k_1 + 1 = 2^{m_2 - m_1}(2k_2 + 1)$$

Now the left hand side is odd, and the right hand side is even unless $m_2 - m_1 = 0$. So we have $m_2 - m_1 = 0$, so $m_1 = m_2$. Thus we have $2k_1 + 1 = 2k_2 + 1$, so $k_1 = k_2$ also, and thus $(m_1, k_1) = (m_2, k_2)$.

Next we must show that f is onto. So let $n \in \mathbb{N}$. Then $n + 1 > 0$, so $n + 1$ can be expressed in the form $2^m(2k + 1)$ (by repeatedly dividing 2 into $n + 1$, m times, until the result is an odd number). But then $f(m, k) = (n + 1) - 1 = n$. □

Proposition 2.28. *Let $\{ A_n \mid n \in \mathbb{N} \}$ be a countable family of countable non-empty sets. Then $\bigcup_{n\in\mathbb{N}} A_n$ is countable.*

Proof. Let $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ be the bijection from Lemma 2.27. For each m we can enumerate A_m as $\{ x_{m,k} \mid k \in \mathbb{N} \}$. Put

$$B = \{ z_n \mid n \in \mathbb{N} \},$$

where if $m, k \in \mathbb{N}$ with $f(m, k) = n$ then $z_n = x_{m,k}$. Note that, for each n there is a unique pair (m, k) with $f(m, k) = n$ (since f is a bijection), so z_n is well defined. Note that B is countable.

If $b \in B$ then $b = z_n = x_{m,k}$ for some $m, k \in \mathbb{N}$. But then $b \in A_m$, so $b \in \bigcup_{n\in\mathbb{N}} A_n$.

Conversely if $a \in \bigcup_{n\in\mathbb{N}} A_n$ then $a \in A_m$ for some m, so $a = x_{m,k}$ for some k, and then $a = z_{f(m,k)} \in B$.

Thus we have $B = \bigcup_{n\in\mathbb{N}} A_n$. So $\bigcup_{n\in\mathbb{N}} A_n$ is countable. □

Lemma 2.29. *If A and B are countable sets then $A \times B$ is countable.*

Proof. If either A or B is empty then $A \times B = \varnothing$ so $A \times B$ is countable. So we assume that A and B are nonempty and can therefore be enumerated as $\{ a_n \mid n \in \mathbb{N} \}$ and $\{ b_n \mid n \in \mathbb{N} \}$ respectively. For each $n \in \mathbb{N}$ let

$$A_n = \{a_n\} \times B = \{ (a_n, b_m) \mid m \in \mathbb{N} \}.$$

Then each A_n is countable so $\bigcup_{n\in\mathbb{N}} A_n$ is countable. But $\bigcup_{n\in\mathbb{N}} A_n = A \times B$, so $A \times B$ is countable as required. □

Proposition 2.30. *Let X be a countable set of symbols. Then the set X^* of all finite non-empty strings of symbols from X is countable.*

Proof. We have $X^* = \bigcup_{n\in\mathbb{N}} X_n$, where X_n is the set of all strings of length $n+1$. So we only need to show that each X_n is countable.

We do this by induction on n: X_0 is just the set of strings of length 1, so there is a bijection from X to X_0. So suppose that X_n is countable, and we wish to deduce that X_{n+1} is countable. Each string of length $n+2$ can be obtained by taking a string from X_n and adding another symbol from X at the end. Thus there is a bijection between X_{n+1} and $X_n \times X$. Since X_n and X are both countable, so is $X_n \times X$, and therefore so is X_{n+1}. □

Theorem 2.31. *The set* form(L) *is countable.*

Proof. Let X be the set of symbols of L. Then X is countable, so X^* is countable. Thus, since form$(L) \subseteq X^*$, form(L) is countable. □

Here is a sketch of an alternative proof of the previous theorem. We use the first exercise below, Ex. 2.6.

- Let $S = \{\rightarrow, \neg, (,), p_0, p_1, \ldots\}$ be the symbol set for L. Choose a bijection $g : S \rightarrow \mathbb{N} - \{0\}$.
- Given a string $w = s_0 s_1 \ldots s_{n-1}$ over S, let

$$G(w) = \prod_{i=0}^{n-1} \alpha_i^{g(s_i)}$$

  where α_i is the i-th prime number. Then G is 1-1 since factoring into prime numbers is unique.
- the restriction of G to form(L) is a 1-1 map form(L) $\rightarrow \mathbb{N}$. So form(L) is countable.

So what is an example of a non-countable set? The set of real numbers is not countable. The easiest example, however is $\mathcal{P}(\mathbb{N})$, the power set of $\mathbb{N}$. If $g : \mathbb{N} \mapsto \mathcal{P}(\mathbb{N})$, one considers the set $R = \{n : n \notin g(n)\}$. If $R = g(n)$, then $n \in R$ implies $n \notin R$, and $n \notin R$ implies $n \in R$. This contradiction (which is the same as the one in Russell's paradox) show that R is not in the range of g. Thus g is not onto.

Exercises

15. The following are equivalent for any set X.

(i) There is a 1–1 function $f : X \to \mathbb{N}$

(ii) X is empty or there is an onto function $g : \mathbb{N} \to X$.

16. Let $\mathcal{F}$ be the set of all finite subsets of $\mathbb{N}$. Show that $\mathcal{F}$ is countable. [Hint: express $\mathcal{F}$ as $\bigcup_{n\in\mathbb{N}} A_n$ where each A_n is countable.]

17. Show that the sets $\mathbb{Z}$ of integers and $\mathbb{Q}$ of rational numbers are both countable.

18. Let I be an uncountable set, and for each $i \in I$ let $U_i \subseteq \mathbb{R}$ be a set of real numbers with the property that there is at least one $q \in \mathbb{Q} \cap U_i$. Show that there is some $q \in \mathbb{Q}$ with $q \in U_i$ for uncountably many $i \in I$. [Hint: express I as $\bigcup_{n\in\mathbb{N}} A_n$ and observe that if all the sets A_n were countable then I would have to be countable.]

2.7 The Adequacy Theorem for L

We are now in a position to pull the ingredients together to prove the adequacy theorem for L. We begin with two more purely syntactical Lemmas concerning consistency.

Lemma 2.32. *Let $\Sigma \subseteq \text{form}(L)$ be a consistent set, and let $\mathcal{A} \in \text{form}(L)$. Then $\Sigma \cup \{\mathcal{A}\}$ is consistent or $\Sigma \cup \{\neg\mathcal{A}\}$ is consistent.*

Proof. We have already shown in Lemma 2.19 that if $\Sigma \nvdash_L \mathcal{A}$ then $\Sigma \cup \{\neg\mathcal{A}\}$ is consistent. So suppose that $\Sigma \vdash_L \mathcal{A}$. We must show that $\Sigma \cup \{\mathcal{A}\}$ is consistent.

Suppose, for a contradiction, that it is inconsistent, in other words that there is some $\mathcal{B}$ such that $\Sigma \cup \{\mathcal{A}\} \vdash_L \mathcal{B}$ and $\Sigma \cup \{\mathcal{A}\} \vdash_L \neg\mathcal{B}$.

Then, by the Deduction Theorem, $\Sigma \vdash_L (\mathcal{A} \to \mathcal{B})$ and $\Sigma \vdash_L (\mathcal{A} \to \neg\mathcal{B})$.

But we also have $\Sigma \vdash_L \mathcal{A}$ so by modus ponens we have $\Sigma \vdash_L \mathcal{B}$ and $\Sigma \vdash_L \neg\mathcal{B}$, contradicting the assumption that Σ is consistent.

Hence there is no such $\mathcal{B}$, in other words $\Sigma \cup \{\mathcal{A}\}$ is consistent, as required. □

Lemma 2.33. *Let $\Sigma \subseteq \text{form}(L)$ be an inconsistent set. Then there is a finite $\Gamma \subseteq \Sigma$ such that Γ is inconsistent.*

Proof. Since Σ is inconsistent, there is some $\mathcal{A}$ such that $\Sigma \vdash_L \mathcal{A}$ and $\Sigma \vdash_L \neg\mathcal{A}$.

Let $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m$ be a derivation of $\mathcal{A}$ from Σ

and let $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n$ be a derivation of $\neg\mathcal{A}$ from Σ.

Put

$$\Gamma = \Sigma \cap \big(\{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m\} \cup \{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n\}\big).$$

Then Γ is a finite subset of Σ.

Consider the sequence $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m$. Every term in this sequence is an axiom, in Σ or follows from two earlier terms by modus ponens. But any of the terms which are in Σ are in Γ, so every term in the sequence is an axiom, is in Γ or follows from two earlier terms by modus ponens. In other words, this is a derivation of $\mathcal{A}$ from Γ. Similarly, $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n$ is a derivation of $\neg\mathcal{A}$ from Γ. So Γ is a finite inconsistent subset of Σ, as required. □

Theorem 2.34. *Let $\Sigma \subseteq \mathrm{form}(L)$ be a consistent set. Then there is a complete consistent set Π with $\Sigma \subseteq \Pi$.*

Proof. We know that $\mathrm{form}(L)$ is countable, so we can enumerate it as $\{\, \mathcal{A}_n \mid n \in \mathbb{N} \,\}$.

We construct a sequence $(\Sigma_n)_{n\in\mathbb{N}}$ of subsets of $\mathrm{form}(L)$ as follows: $\Sigma_0 = \Sigma$, and if we have defined Σ_n then

$$\Sigma_{n+1} = \begin{cases} \Sigma_n \cup \{\mathcal{A}_n\} & \text{if } \Sigma \cup \{\mathcal{A}_n\} \text{ is consistent} \\ \Sigma_n \cup \{\neg\mathcal{A}_n\} & \text{otherwise} \end{cases}$$

Note that if $\Sigma_n \cup \{\mathcal{A}_n\}$ is inconsistent then $\Sigma_n \cup \{\neg\mathcal{A}_n\}$ is consistent, by Lemma 2.32. Hence we can prove by induction that each Σ_n is consistent.

Put $\Pi = \bigcup_{n\in\mathbb{N}}$. We claim that Π is consistent. By the previous lemma, it is enough to show that each finite subset of Π is consistent, so let $\Gamma \subseteq \Pi$ be finite, say

$$\Gamma = \{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k\}.$$

For each i, $\mathcal{A}_i \in \Sigma_{n_i}$ for some $n_i \in \mathbb{N}$. Put

$$m = \max\{n_1, n_2, \ldots, n_k\}.$$

Then we have $n_i \leqslant m$ for each i, and $\mathcal{A}_i \in \Sigma_{n_i} \subseteq \Sigma_m$, so we have $\Gamma \subseteq \Sigma_m$. Then if Γ were inconsistent, Σ_m would also be inconsistent. So Γ is consistent, as required.

Finally we must check that Π is complete. So let $\mathcal{A} \in \mathrm{form}(L)$. Then $\mathcal{A} = \mathcal{A}_n$ for some n. So we have either

$$\mathcal{A}_n \in \Sigma_{n+1} \text{ or } \neg\mathcal{A}_n \in \Sigma_{n+1},$$

and $\Sigma_{n+1} \subseteq \Pi$, so $\mathcal{A} \in \Pi$ or $\neg\mathcal{A} \in \Pi$, as required. □

Theorem 2.35. *Let Σ be a set of wffs of L. Then Σ is satisfiable if and only if Σ is consistent.*

Proof. Suppose first that Σ is satisfiable. Let v be a truth assignment which satisfies Σ. Suppose, for a contradiction, that Σ is inconsistent, in other words there is some $\mathcal{B}$ with $\Sigma \vdash_L \mathcal{B}$ and

$\Sigma \vdash_L \neg\mathcal{B}$. Then, by the Soundness Theorem, $\Sigma \models \mathcal{B}$ and $\Sigma \models \neg\mathcal{B}$, so since v satisfies Σ we must have $v(\mathcal{B}) = 1$ and $v(\neg\mathcal{B}) = 1$. But this is impossible, so there is no such $\mathcal{B}$, i.e. Σ is consistent.

Conversely, suppose that Σ is consistent. Then there is some complete consistent Π with $\Sigma \subseteq \Pi$. Then, by Theorem 2.22, the function $f_\Pi : \text{form}(L) \to \{0, 1\}$ defined by

$$f_\Pi(\mathcal{B}) = \begin{cases} 1 & \text{if } \mathcal{B} \in \Pi \\ 0 & \text{if } \mathcal{B} \notin \Pi \end{cases}$$

is a truth assignment which satisfies Π. In particular, it satisfies Σ, so Σ is satisfiable. □

The following Theorem, together with the Soundness theorem for L, establishes the desired equivalence between syntax and semantics:

$$\Sigma \vdash_L \mathcal{A} \text{ if and only if } \Sigma \models \mathcal{A}.$$

Theorem 2.36 (The Adequacy Theorem for L)**.** *Let $\Sigma \cup \{\mathcal{A}\} \subseteq \text{form}(L)$. If $\Sigma \models \mathcal{A}$ then $\Sigma \vdash_L \mathcal{A}$.*

Proof. We prove the contrapositive: if $\Sigma \nvdash_L \mathcal{A}$ then $\Sigma \not\models \mathcal{A}$. So suppose $\Sigma \nvdash_L \mathcal{A}$. Then $\Sigma \cup \{\neg\mathcal{A}\}$ is consistent, so it is satisfiable. Let v be a truth assignment which satisfies $\Sigma \cup \{\neg\mathcal{A}\}$. Then $v(\mathcal{B}) = 1$ for all $\mathcal{B} \in \Sigma$ but $v(\mathcal{A}) = 0$, so $\Sigma \not\models \mathcal{A}$, as required. □

Corollary 2.37. *Every tautology is a theorem of L.*

Proof. If $\mathcal{A}$ is a tautology then $\varnothing \models \mathcal{A}$, so $\varnothing \vdash_L \mathcal{A}$, in other words $\mathcal{A}$ is a theorem of L. □

Corollary 2.38 (The Compactness Theorem for L)**.** *Let $\Sigma \subseteq \text{form}(L)$. If every finite subset of Σ is satisfiable then Σ is satisfiable.*

Proof. Suppose that every finite subset of Σ is satisfiable. Then every finite subset of Σ is consistent, so Σ is consistent, so Σ is satisfiable. □

Exercises

19. Let $\Sigma \cup \{\mathcal{A}, \mathcal{B}\} \subseteq \text{form}(L)$. Show that if $\Sigma \cup \{\mathcal{A}\}$ is satisfiable and $\Sigma \cup \{\mathcal{B}\}$ is not satisfiable then $\Sigma \vdash_L (\mathcal{B} \to \mathcal{A})$.

20. Let $\Sigma \cup \{\mathcal{A}, \mathcal{B}\} \subseteq \text{form}(L)$. Show that either $\Sigma \cup \{(\mathcal{A} \to \mathcal{B})\}$ is satisfiable or $\Sigma \vdash_L \neg\mathcal{B}$.

21. (a) If $\mathcal{A}$ is a tautology and $\Sigma \subseteq$ form(L) is a set such that $\Sigma \vdash_L \neg\mathcal{A}$, then Σ is inconsistent.
    (b) Give an example of a set $\Sigma \subseteq$ form(L) which is complete but inconsistent.

Chapter 3

Informal Predicate Logic

First a little history.

Aristotle (384-322 BC). He tried to set up a system for valid reasoning, using syllogisms.

> All men are mortal.
> Socrates is a man.
> So Socrates is mortal.

Logic was, till the end of the 19th century, a discipline of philosophy. Nowadays there is mathematical logic and philosophical logic. The syllogisms are still around in philosophy. A formal counterpart appears in predicate logic, as we will see later.

Boole 1854 Laws of Thought. In fact, he introduced a forerunner of statement logic (also called propositional logic). He was the first to have a sort of formalized system.

Frege 1870 Begriffsschrift ("conceptscript").

First system also including things like "for all x". He introduced naive set theory, trying to found mathematics on logic, but unfortunately naive set theory is not consistent, because one can form the set $\{x : x \notin x\}$.

Russell, 1910 Principia Mathematica.

A second attempt to found mathematics on logic. Huge formal system (3 fat volumes). Avoids Frege' s contradiction. One is not allowed to form the set $\{x : x \notin x\}$.

Gödel, 1933 Incompleteness theorems. Not everything true is provable in a formal system.

Nowadays, mathematical logic is mostly a mathematical discipline, like algebra or topology.

- Set theory: still concerned about foundations. Also creates interesting objects, like ordinals
- Model theory: abstract version of algeba
- Computability theory: computability in principle
- Proof Theory: study syntactical proof systems

Now we will introduce predicate logic. (We make the step from Boole to Frege and Russell.)

The statement logic we learned about in the previous part is fine as far as it goes. However, it is not rich enough to describe ordinary reasoning in mathematics.

For example, how would we express a statement such as

> "For every natural number k, if k is even then k^2 is even"?

We could denote the statement "k is even" by A_k. Then we wish to say something like

$$(A_0 \to A_0) \wedge (A_1 \to A_1) \wedge (A_2 \to A_4) \\ \wedge (A_3 \to A_9) \wedge \cdots \wedge (A_k \to A_{k^2}) \wedge \ldots$$

However, this infinite string of symbols is not allowed in our system. So we use the *universal quantifier* $\forall$ (read as "for all") to represent the above as $\forall k\,(A_k \to A_{k^2})$.

Unlike the situation with statement variables p_k where the subscript k was just an index to tell the variables apart, in this case the statement really does depend on k. We will adjust our notation shortly to make this clearer.

Just as in statement logic, where we were interested in statement forms, we prefer to consider **predicate forms** rather than specific predicates. We use predicate symbols to represent a particular predicate, constant symbols to represent particular constants, and function symbols to represent particular functions, and we build these into formulas using similar rules to those for statement forms.

In this chapter we will first study these formulas from a semantic point of view—we will give meaning to the symbols, and consider what kind of formula in predicate logic corresponds to the tautologies of statement logic. (Later we will consider a syntactic approach, with formal rules for deriving theorems.) Whereas in statement logic we dealt with statements that could be true of false, in predicate logic, the situation is a little more complicated. Once we have established what constitutes a predicate form, we will discuss how to interpret a predicate form.

To get an idea of what is to come:

$$\forall k\,(Ak \to Afk)$$

is a predicate form. If we interpret Ak as "k is even" and fk as the function $k \mapsto k^2$, then the predicate form is interpreted as "for all k, if k is even then k^2 is even". Although it is still not clear what we mean by "for all k". We also need to state the values k is allowed have. For example, if k must be a natural number, then we can interpret the predicate form by the statement:

"For every natural number k, if k is even then k^2 is even."

3.1 First order languages

The symbols we use to build formulas in predicate logic are as follows:

variables	x, y, z, ...
constant symbols	a, b, c, ...
predicate symbols	$A^1, A^2, \ldots, B^1, B^2, \ldots, C^1, C^2, \ldots$
function symbols	$f^1, f^2, \ldots, g^1, g^2, \ldots, h^1, h^2, \ldots$
punctuation	$(,)$
connectives	$\neg$, $\wedge$, $\vee$, $\to$, $\leftrightarrow$
quantifiers	$\forall$, $\exists$.

The superscript associated with each predicate symbol and function symbol is called the **arity**. Thus when the arity of A^1 is 1 (in which case we call it a *unary* predicate), then we can form A^1x but not A^1xy. On the other hand when the arity of A^2 is 2 (in which case we say that A^2 is a *binary* predicate) then we can form A^2xy but not A^2x or A^2xyz.

So, for example, we could have the predicate form:

$$\forall x\,(B^2ax \to B^2af^1x), \tag{3.1}$$

an interpretation of B^2xy as "x divides y", of f^1x as $x \mapsto x^2$ and the constant symbol c as 2. What is the interpretation of the predicate form (3.1)?

In other situations we may want to use a 3-ary predicate $Cxyz$, and so on. We now give the formal definitions of a predicate form and a first order language.

Definition 3.1. *A* **first-order language** *is a 3-tuple $\mathcal{L} = (C, P, F)$, where*

- *C is a (possibly empty) set of constant symbols*
- *P is a non-empty set of predicate symbols*

- *F is a (possibly empty) set of function symbols*
- *C, P and F are disjoint.*

If $A^n \in P$, we say that A^n is an n-ary predicate symbol, and if $f^n \in F$, we say that f is an n-ary function symbol.

Example 3.2. $\mathcal{L}_N = (C, P, F)$, where

$$C = \{c\} \qquad P = \{E^2, B^2\} \qquad F = \{f^2, g^2, h^1\}$$

is a first-order language, and we can write formulas like:

(N1) $\forall x \, \forall y \, B^2 f^2 xy g^2 xy$.
(N2) $\forall x \, E^2 g^2 x h^1 x f^2 g^2 xxx$.
(N3) $B^2 g^2 h^1 h^1 c x f^2 h^1 h^1 c x$.

This is not very interesting in itself. When we create a language, we usually do so with some mathematical structure in mind. For example $\mathcal{L}_N$ has all the symbols required to make statements about the natural numbers. To see this, interpret

- c as the number 0,
- f^2 as addition,
- g^2 as multiplication,
- h^1 the successor function, that is,

$$n \mapsto n + 1$$

- E^2 equality and
- B^2 "less than or equal to".

Example 3.3. What do the statements N1-N3 above represent? Write your answer below.

So we can think of $\mathcal{L}_N$ as representing the natural numbers. In Definition 3.6 below we will formalise this notion; we will give a precise definition of an *interpretation* of a language.

A note about arity: When we write g^2xy, for example, it is clear that the arity of g^2 is 2 because there are 2 variables x and y, so we could just write gxy. From now on when the arity is clear we will usually leave out the superscripts and just write gxy. However, the arity is not clear if we write $gfxyz$. There are 3 possibilities: g^3f^1xyz, g^2f^2xyz or g^1f^3xyz. So we really need to know the arity of f and g if we want to leave off the superscripts. In cases like this we will always specify the arity and only leave out the superscripts once the arity is known.

Definition 3.4 (Terms)**.** *Let $\mathcal{L} = (C, P, F)$ be a first-order language. The* **terms** *of $\mathcal{L}$ are all strings of symbols which can be built up using the following rules:*

1. *every variable is a term;*
2. *every constant symbol in C is a term;*
3. *if $f \in F$ is an n-ary function symbol and $t_1, t_2, \ldots t_n$ are terms then $ft_1t_2 \ldots t_n$ is a term.*

We denote the set of terms of $\mathcal{L}$ by $\text{term}(\mathcal{L})$.

Definition 3.5 (Predicate forms)**.** *Let $\mathcal{L} = (C, P, F)$ be a first-order language. An* **atomic formula** *of $\mathcal{L}$ is a string of symbols of the form $At_1t_2, \ldots t_n$, where A is an n-ary predicate symbol in P and $t_1, t_2, \ldots, t_n \in \text{term}(\mathcal{L})$.*

The **predicate forms** *of $\mathcal{L}$ are all strings of symbols which can be built up using the following rules:*

1. *every atomic formula is a predicate form.*
2. *if $\mathcal{A}$ and $\mathcal{B}$ are predicate forms then so are $\neg\mathcal{A}$, $(\mathcal{A} \wedge \mathcal{B})$, $(\mathcal{A} \vee \mathcal{B})$, $(\mathcal{A} \to \mathcal{B})$, $(\mathcal{A} \leftrightarrow \mathcal{B})$, $\forall x\, \mathcal{A}$ and $\exists x\, \mathcal{A}$.*

In this definition we have introduced the *existential quantifier* $\exists$ (read as "there exists").

The applications we have in mind for the formal language we are developing are mathematical. However, the formal language certainly reflects aspects of our natural language. Consider the real life example where $P = \{M^1, W^1, L^2\}$. Interpretate Mx as "x is a man", Wx as "x is a woman" and Lxy as "x likes y". What is the intended meaning of the predicate form:

$$\forall x\, (Mx \to \exists y\, (Wy \wedge (Lxy \wedge \neg Lyx))).$$

We see from this example that nouns often correspond to unary predicates and verbs to binary ones. How about function symbols? For example, include $F = \{g^1\}$ in the language, where the intended meaning of gx is "the mother of x". What is the meaning of the following?

$$\forall y\,(Wy \to \exists x\ (Mx \wedge (Lyx \wedge \neg Lgxy))).$$

Of course natural language is much richer (and also much vaguer) than this.

Exercises

1. Let $\mathcal{L}_G = (C, P, F)$, where

$$C = \{c\} \qquad P = \{E^2\} \qquad F = \{f^2, g^1\}.$$

   (a) If we let

   - c represent the identity 1.
   - f represent the group multiplication.
   - g represent the inversion operation.
   - E represent $=$.

   how would we express the following statements in $\mathcal{L}_G$?

   (i) "for all $x, y, z \in G$, $x(yz) = (xy)z$."

   (ii) "for all $x \in G$, $xx^{-1} = 1$."

   (iii) "there exist $x, y \in G$ such that $xy \neq yx$."

   (b) What kind of mathematical structure have we described?

2. Let $\mathcal{L}_{ord} = (\varnothing, P, \varnothing)$ where $P = \{E^2, R^2\}$. What kind of mathematical structure does the following system of predicate forms describe?

   LO1 $\forall x\, Rxx$

   LO2 $\forall x\, \forall y\, ((Rxy \wedge Ryx) \to Exy))$

   LO3 $\forall x\, \forall y\, \forall z\, ((Rxy \wedge Ryz) \to Rxz))$

   LO4 $\forall x\, \forall y\, (Rxy \vee Ryx)$

3.2 Interpretations, satisfaction and truth

Interpretations

Definition 3.6. *Let $\mathcal{L} = (C, P, F)$ be a first-order language. An* **interpretation** *of $\mathcal{L}$ is a quadruple $\mathcal{I} = (D, \mathbf{C}, \mathbf{P}, \mathbf{F})$ such that*

1. *D is a nonempty set (called the* **domain** *of $\mathcal{I}$);*

2. *for each constant symbol $c \in C$ there is some element $c^{\mathcal{I}}$ of D in* $\mathbf{C}$*;*

3. *for each n-ary predicate symbol $A \in P$ there is an n-ary relation $A^{\mathcal{I}}$ on D (i.e. a subset of D^n) in* $\mathbf{P}$*; and*

4. *for each n-ary function symbol $f \in F$ there is a function $f^{\mathcal{I}} : D^n \to D$ in* $\mathbf{F}$*.*

Keep in mind that it depends on the particular interpretation $\mathcal{I}$ what the function $f^{\mathcal{I}}$ is.

The idea of the above definition is that the domain is the set of possible values that the variables can take, the relations, and functions are the "meanings" of the predicate symbols, and the values of the constant symbols are elements of the domain (e.g. $c_{\mathcal{I}}$ is the value of c).

For example, we might have

- $D = \mathbb{N}$,
- $E^{\mathcal{I}}$ the relation of equality (i.e. $E^{\mathcal{I}}(m, n)$ if and only if $m = n$),
- $h^{\mathcal{I}}(n) = n + 1$ and
- $c^{\mathcal{I}} = 0$.

Then the predicate form

$$\forall x \, \forall y \, (Exy \to Ehxhy)$$

is interpreted by the statement "for all natural numbers m and n, if $m = n$ then $m + 1 = n + 1$", which is a true statement. In other words, the predicate form is true in that interpretation.

Example 3.7. Let $\mathfrak{L}_N$ be the language of the natural numbers, described in Example 3.3. The **standard interpretation** of $\mathfrak{L}_N$, denoted $\mathcal{N}$, is given by

$$\mathcal{N} = (\mathbb{N}, \{c^{\mathcal{N}}\}, \{E^{\mathcal{N}}, B^{\mathcal{N}}\}, \{f^{\mathcal{N}}, g^{\mathcal{N}}, h^{\mathcal{N}}\}),$$

where

- $c^{\mathcal{N}} = 0$
- $E^{\mathcal{N}}$ is the relation of equality
- $B^{\mathcal{N}}$ is the relation "less than or equal to"
- $f^{\mathcal{N}}$ is the addition operation
- $g^{\mathcal{N}}$ is the multiplication operation
- $h^{\mathcal{N}}$ is the successor operation $n \mapsto n + 1$.

Exercises

3. Translate the following into $\mathcal{L}_N$:
   (a) m is even; [hint: express this as "there is an n such that $m = 2n$"],
   (b) for all natural numbers n, $n(n+1)$ is even.
4. Translate the following predicate form from $\mathcal{L}_N$ into an assertion about the natural numbers:

$$\forall x \, \forall y \, (Bxy \to (Exy \vee Bhxy)).$$

5. Translate the following assertions about the natural numbers into $\mathcal{L}_N$:
   (a) There is a natural number m such that for all natural numbers n, $m \cdot n = n$.
   (b) For all natural numbers n, $n \cdot 0 = 0$.
   (c) For all natural numbers m and n, $m \leqslant n$ if and only if there is a natural number k with $m + k = n$.
6. Translate the following predicate forms from $\mathcal{L}_N$ into assertions about the natural numbers:
   (a) $\forall x \, \forall y \, (Bxy \vee Bgyygxx)$.
   (b) $\forall x \, \forall y \, Egfxyfxyfgxxfghhcgxygyy$.

Satisfaction

Even when we know what the symbols mean, we can't always say that a formula is true or false until we know the values of the variables. So we introduce the notions of *$\mathcal{I}$-assignments* and *satisfaction*.

To understand the idea, consider the formula $n + 3 = 2n$. It does not make sense to say that this is true or that this is false until we decide what n is. If we assign the variable n the value 3 then the predicate form will be satisfied. If we assign the variable n any other value, then the predicate form will not be satisfied.

Definition 3.8. *Let $\mathcal{I} = (D, \mathbf{C}, \mathbf{P}, \mathbf{F})$ be an interpretation of a first-order language $\mathcal{L} = (C, P, F)$. An* **$\mathcal{I}$-assignment** *is a function v from the set of variables to D. Given an $\mathcal{I}$-assignment v, we define the function $\tilde{v} : \mathrm{term}(\mathcal{L}) \to D$ as follows:*

- *if x is a variable then $\tilde{v}(x) = v(x)$;*
- *if $c \in C$ then $\tilde{v}(c) = c^{\mathcal{I}}$; and*
- *if f is an n-ary function symbol in F and $t_1, t_2, \ldots, t_n$ are terms then*

$$\tilde{v}(ft_1t_2 \ldots t_n) = f^{\mathcal{I}}(\tilde{v}(t_1), \tilde{v}(t_2), \ldots, \tilde{v}(t_n))$$

An $\mathcal{I}$-assignment is a way of assigning "values" to each of the terms. In other words, for each term t, we can talk about the element $\tilde{v}(t)$ of D.

Once we have assigned values to the variables in this way, it makes sense to say whether or not a predicate form is satisfied. We abbreviate the statement "$\mathcal{I}$ **satisfies** $\mathcal{A}$ **under the** $\mathcal{I}$**-assignment** v" (or, more briefly, "$\mathcal{I}$ **and** v **satisfy** $\mathcal{A}$") by

$$\mathcal{I} \models_v \mathcal{A}.$$

Before defining $\mathcal{I} \models_v \mathcal{A}$ precisely, we introduce some special notation to change the value of an $\mathcal{I}$-assignment at a particular variable.

Definition 3.9. *Let $\mathcal{I}$ be an interpretation with domain D, let v be an $\mathcal{I}$-assignment, let x be a variable and let $d \in D$. Then $v\frac{d}{x}$ is the $\mathcal{I}$-assignment with*

$$v\tfrac{d}{x}(y) = \begin{cases} d & \text{if } y = x \\ v(y) & \text{if } y \neq x \end{cases}$$

for every variable y.

Exercises

7. Let $\mathcal{N}$ be the standard interpretation of $\mathcal{L}_N$, and let v be an $\mathcal{N}$-assignment with $v(x) = 2$, $v(y) = 5$ and $v(z) = 0$. Calculate the following:
   (a) $\tilde{v}(fxy)$.
   (b) $\tilde{v}(gfxcgyhz)$.
   (c) $\widetilde{v\frac{3}{x}}(fxy)$.
   (d) $\widetilde{v\frac{3}{x}}(gfxhyfcx)$.
   (e) $\widetilde{v\frac{3}{x}\frac{6}{y}}(fxy)$.

Definition 3.10. *Let v be an $\mathcal{I}$-assignment, and let $\mathcal{A}$ and $\mathcal{B}$ be predicate forms. We define the notion of* **satisfaction** *as follows.*

1. *If $\mathcal{A}$ is the atomic formula $At_1t_2\dots t_n$, then $\mathcal{I} \models_v \mathcal{A}$ iff $A^{\mathcal{I}}(\tilde{v}(t_1), \tilde{v}(t_2), \dots, \tilde{v}(t_n))$ holds (i.e. iff $(\tilde{v}(t_1), \tilde{v}(t_2), \dots, \tilde{v}(t_n)) \in A^{\mathcal{I}}$).*
2. *$\mathcal{I} \models_v \neg\mathcal{A}$ iff $\mathcal{I} \not\models_v \mathcal{A}$.*
3. *$\mathcal{I} \models_v \mathcal{A} \wedge \mathcal{B}$ iff $\mathcal{I} \models_v \mathcal{A}$ and $\mathcal{I} \models_v \mathcal{B}$.*
4. *$\mathcal{I} \models_v \mathcal{A} \vee \mathcal{B}$ iff $\mathcal{I} \models_v \mathcal{A}$ or $\mathcal{I} \models_v \mathcal{B}$.*
5. *$\mathcal{I} \models_v \mathcal{A} \to \mathcal{B}$ iff $\mathcal{I} \not\models_v \mathcal{A}$ or $\mathcal{I} \models_v B$.*
6. *$\mathcal{I} \models_v \mathcal{A} \leftrightarrow \mathcal{B}$ iff either $\mathcal{I} \models_v \mathcal{A}$ and $\mathcal{I} \models_v \mathcal{B}$ or $\mathcal{I} \not\models_v \mathcal{A}$ and $\mathcal{I} \not\models_v \mathcal{B}$.*
7. *$\mathcal{I} \models_v \forall x\,\mathcal{A}$ iff $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$ for every $d \in D$.*
8. *$\mathcal{I} \models_v \exists x\,\mathcal{A}$ iff there is some $d \in D$ with $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$.*

Example 3.11. Take the language to be $\mathcal{L}_N$, and $\mathcal{N}$ to be the interpretation described in Example 3.7. Let $\mathcal{A}$ be the predicate form

$$Efxyx$$

Let v be an $\mathcal{I}$-assignment. Then

$$\tilde{v}(fxy) = f^{\mathcal{I}}(\tilde{v}(x), \tilde{v}(y)) = \tilde{v}(x) + \tilde{v}(y) = v(x) + v(y).$$

Also $E^{\mathcal{I}}(\tilde{v}(s), \tilde{v}(t))$ holds if and only if $\tilde{v}(s) = \tilde{v}(t)$. Thus

$$\mathcal{N} \models_v \mathcal{A} \text{ if and only if } v(x) + v(y) = v(x).$$

- Let v be an $\mathcal{N}$-assignment with $v(x) = 3$ and $v(y) = 2$. Then $v(x) + v(y) = 5 \neq v(x)$, so $\mathcal{N} \not\models_v \mathcal{A}$.
- On the other hand, if u is an $\mathcal{N}$-assignment with

  $$u(x) = 3 \text{ and } u(y) = 0$$

  then $u(x) + u(y) = 3 + 0 = 3 = u(x)$, so $\mathcal{N} \models_u \mathcal{A}$.

- Let $n \in \mathbb{N}$. Then $u\frac{n}{x}(x) = n$ and $u\frac{n}{x}(y) = u(y) = 0$, so

  $$u\tfrac{n}{x}(x) + u\tfrac{n}{x}(y) = n + 0 = n = u\tfrac{n}{x}(x).$$

  Hence $\mathcal{N} \models_{u\frac{n}{x}} \mathcal{A}$. Since this holds for any $n \in \mathbb{N}$, $\mathcal{N} \models_u \forall x\,\mathcal{A}$.
  (This says that in $\mathcal{N}$, " for all n, $n + 0 = n$.")

- We have
$$v\tfrac{0}{y}(x) + v\tfrac{0}{y}(y) = v(x) + 0 = v(x) = v\tfrac{0}{y}(x),$$
so $\mathcal{N} \models_{v\frac{0}{y}} \mathcal{A}$, so $\mathcal{N} \models_v \exists y\, \mathcal{A}$.
(This says that in $\mathcal{N}$, " there exists an m such that $3 + m = 3$.")

Truth

Definition 3.12. *A predicate form $\mathcal{A}$ is* **true** *in an interpretation $\mathcal{I}$ if,*

$$\textit{for every } \mathcal{I}\textit{-assignment } v,\ \mathcal{I} \models_v \mathcal{A}.$$

It is **false** *if*

$$\textit{for every } \mathcal{I}\textit{-assignment } v,\ \mathcal{I} \not\models_v \mathcal{A}.$$

If $\mathcal{A}$ is true in $\mathcal{I}$, we write $\mathcal{I} \models \mathcal{A}$.

Notice that, just as not every statement has to be either a tautology or a contradiction, not every predicate form has to be either true or false in a given interpretation—there may be some $\mathcal{I}$-assignments which satisfy it and others which do not.

For example, we saw in Example 3.11 that $\mathcal{N} \models_v Efxyx$ for some v and not for others. So this predicate form is neither true nor false in this interpretation.

Example 3.13. The predicate form $Efxyfyx$ is true in $\mathcal{N}$.

Proof. Let v be an $\mathcal{N}$-assignment. We have
$$\tilde{v}(fxy) = f^{\mathcal{N}}(\tilde{v}(x), \tilde{v}(y)) = \tilde{v}(x) + \tilde{v}(y) = v(x) + v(y).$$
Similarly $\tilde{v}(fyx) = v(y) + v(x)$. Thus
$$\mathcal{N} \models_v Efxyfyx \text{ iff } E^{\mathcal{N}}(v(x) + v(y), v(y) + v(x)) \text{ holds.}$$
In other words, it holds iff
$$v(x) + v(y) = v(y) + v(x).$$
Since $m + n = n + m$ for all $m, n \in \mathbb{N}$, this is true for every v, in other words the predicate form is true in $\mathcal{N}$. □

Example 3.14. For a real-life analogy, think of interpretations as countries, and $\mathcal{I}$-assignments map variables to people who live in that country. We use the first order language where $P = \{\text{Politician}^1, \text{SaysTruth}^1\}$

Let $\mathcal{A}$ be the predicate form

$$\text{Politician}(x) \to \text{SaysTruth}(x).$$

First consider the case where $\mathcal{I}$ is the United States. For the $\mathcal{I}$-assignment v where $v(x)=$ "Jimmy Carter", we have $I \models_v \mathcal{A}$, while for the $\mathcal{I}$-assignment u where $u(x)=$"Richard Nixon", $I \not\models_v \mathcal{A}$. So $\mathcal{A}$ is not true in the US. On the other hand, if $\mathcal{I}=$ New Zealand, then $\mathcal{A}$ is true in $\mathcal{I}$.

Exercises

8. Let $\mathcal{A}$ be the predicate form

$$\forall z\,((Bxz \wedge Bzy) \to (Exz \vee Ezy)).$$

   (a) Let v be an $\mathcal{N}$-assignment with $v(x) = 1$ and $v(y) = 2$. Show that $\mathcal{N} \models_v \mathcal{A}$.

   (b) Let u be an $\mathcal{N}$-assignment with $u(x) = 1$ and $u(y) = 4$. Show that $\mathcal{N} \not\models_u \mathcal{A}$.

   (c) Explain why $\mathcal{N} \models \forall x\,\exists y\,\mathcal{A}$.

3.3 Logical validity and logical implication

We now move on to the predicate logic version of tautologies.

Definition 3.15. *A predicate form $\mathcal{A}$ of a first-order language $\mathcal{L}$ is* **logically valid** *if, for every interpretation $\mathcal{I}$ of $\mathcal{L}$, $\mathcal{I} \models \mathcal{A}$.*

For example, we will show that for any predicate form $\mathcal{A}$, the following predicate forms are logically valid:

- $(\forall x\, \mathcal{A} \to \exists x\, \mathcal{A})$
- $(\exists x\, \forall y\, \mathcal{A} \to \forall y\, \exists x\, \mathcal{A})$

However the predicate form

$$(\forall y\, \exists x\, A(x, y) \to \exists x\, \forall y\, A(x, y))$$

is not logically valid.

Proposition 3.16. *For any $\mathcal{A}$, the predicate form*

$$(\forall x\, \mathcal{A} \to \exists x\, \mathcal{A})$$

is logically valid.

Proof. Let $\mathcal{I}$ be an interpretation and let v be an $\mathcal{I}$-assignment.

If $\mathcal{I} \not\models_v \forall x\, \mathcal{A}$ then $\mathcal{I} \models_v (\forall x\, \mathcal{A} \to \exists x\, \mathcal{A})$.

So suppose that $\mathcal{I} \models_v \forall x\, \mathcal{A}$. Pick some d in the domain of $\mathcal{I}$ (recall that we insist the domain of an interpretation is a *non-empty* set).

Then, since $\mathcal{I} \models_v \forall x\, \mathcal{A}$, $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$.

Thus $\mathcal{I} \models_v \exists x\, \mathcal{A}$, so $\mathcal{I} \models_v (\forall x\, \mathcal{A} \to \exists x\, \mathcal{A})$, as required.

□

Proposition 3.17. *For any $\mathcal{A}$, the wff*

$$(\exists x\, \forall y\, \mathcal{A} \to \forall y\, \exists x\, \mathcal{A})$$

is logically valid.

Proof. Let $\mathcal{I}$ be an interpretation with domain D, and let v be an $\mathcal{I}$-assignment.

Case 1. If $\mathcal{I} \not\models_v \exists x \, \forall y \, \mathcal{A}$ then

$$\mathcal{I} \models_v (\exists x \, \forall y \, \mathcal{A} \to \forall y \, \exists x \, \mathcal{A}).$$

Case 2. Now suppose that $\mathcal{I} \models_v \exists x \, \forall y \, \mathcal{A}$: we will show that $\mathcal{I} \models_v \forall y \, \exists x \, \mathcal{A}$. To this end, let $d \in D$: we will show that $\mathcal{I} \models_{v\frac{d}{y}} \exists x \, \mathcal{A}$.

- Since $\mathcal{I} \models_v \exists x \, \forall y \, \mathcal{A}$, there is some $e \in D$ with

  $$\mathcal{I} \models_{v\frac{e}{x}} \forall y \, \mathcal{A}.$$

- So for any $d' \in D$, $v\frac{e}{x}\frac{d'}{y}$ satisfies $\mathcal{A}$. In particular, $v\frac{e}{x}\frac{d}{y}$ satisfies $\mathcal{A}$.
- But $v\frac{e}{x}\frac{d}{y} = v\frac{d}{y}\frac{e}{x}$, so there is some e such that $\mathcal{I} \models_{v\frac{d}{y}\frac{e}{x}} \mathcal{A}$, so $\mathcal{I} \models_{v\frac{d}{y}} \exists x \, \mathcal{A}$.
- Since $d \in D$ was chosen arbitrarily,

  $$\mathcal{I} \models_v \forall y \, \exists x \, \mathcal{A},$$

  and so

  $$\mathcal{I} \models_v (\exists x \, \forall y \, \mathcal{A} \to \forall y \, \exists x \, \mathcal{A}),$$

  as required.

□

To understand this, consider $\mathcal{I} = (D, \varnothing, P, \varnothing)$, where D is the set of people in New Zealand and $P = \{\text{knows}^2\}$. Let $\mathcal{A}$ be the predicate form

$$\text{knows}yx.$$

Pick e to be Peter Jackson. Of course, everybody in New Zealand knows Peter Jackson.

Proposition 3.18. *The predicate form*

$$(\forall y \, \exists x \ Axy \to \exists x \ \forall y \, Axy)$$

is not logically valid.

Proof. Let $\mathcal{I}$ be the interpretation with domain $\mathbb{N}$ in which $A^{\mathcal{I}}mn$ holds iff $m > n$. We will give an $\mathcal{I}$-assignment v which does not satisfy the predicate form.

Let v be the $\mathcal{I}$-assignment with $v(z) = 0$ for all variables z. We will show that v satisfies $\forall y\, \exists x\ Axy$ but does not satisfy $\exists x\, \forall y\, Axy$.

Let $n \in \mathbb{N}$. We wish to show that $v\frac{n}{y}$ satisfies $\exists x\, Axy$. Put $u = v\frac{n}{y}\frac{(n+1)}{x}$.

Then $u(y) = n$ and $u(x) = n + 1$, so $u(x) > u(y)$, in other words $A^{\mathcal{I}}(\tilde{u}(x), \tilde{u}(y))$ holds.

So we have $\mathcal{I} \models_u Axy$, i.e. $\mathcal{I} \models_{v\frac{n}{y}\frac{(n+1)}{x}} Axy$.

Thus $\mathcal{I} \models_{v\frac{n}{y}} \exists x\, Axy$. Since this holds for any n, $\mathcal{I} \models_v \forall y\, \exists x\, Axy$.

On the other hand, let $m \in \mathbb{N}$. We wish to show that $\mathcal{I} \not\models_{v\frac{m}{x}} \forall y\, Axy$.

Put $w = v\frac{m}{x}\frac{m}{y}$. Then $w(x) = w(y) = m$, so $\mathcal{I} \not\models_w Axy$, i.e. $\mathcal{I} \not\models_{v\frac{m}{x}\frac{m}{y}} Axy$, so we do not have $\mathcal{I} \models_{v\frac{m}{x}} \forall y\, Axy$.

Thus there is no m for which $\mathcal{I} \models_{v\frac{m}{x}} \forall y\, Axy$, so $\mathcal{I} \not\models_v \exists x\, \forall y\, Axy$. □

Definition 3.19. *Let $\mathcal{A}$ and $\mathcal{B}$ be predicate forms. We say that $\mathcal{A}$* **logically implies** *$\mathcal{B}$ (written $\mathcal{A} \Rightarrow \mathcal{B}$) if $(\mathcal{A} \to \mathcal{B})$ is logically valid, in other words if for every interpretation $\mathcal{I}$ and every $\mathcal{I}$-assignment v, if v satisfies $\mathcal{A}$ then v also satisfies $\mathcal{B}$.*

Example 3.20. From the examples we have seen above, we know that

1. $\forall x\, \mathcal{A} \Rightarrow \exists x\, \mathcal{A}$.
2. $\exists x\, \forall y\, \mathcal{A} \Rightarrow \forall y\, \exists x\, \mathcal{A}$.
3. $\forall y\, \exists x\, \mathcal{A} \not\Rightarrow \exists x\, \forall y\, \mathcal{A}$.

Exercises

9. Let A be a binary predicate symbol.
   (a) Show that $\forall x\, Axy \Rightarrow Ayy$.
   (b) Show that $\exists x\, Axy \not\Rightarrow Ayy$.
10. Let A be a binary predicate symbol.
   (a) Show that $\forall x\, \exists y\, Axy \Rightarrow \exists y\, Azy$.
   (b) Show that $\forall x\, \exists y\, Axy \not\Rightarrow \exists y\, Ayy$.

3.4 Free and bound variables

Looking back at the examples of predicate forms in Example 3.3, we can see two different types of predicate forms. The first two, "for all natural numbers m and n, $m + n \leqslant mn$" and "for every natural number n, $n(n + 1) = n^2 + n$" are statements which are either true or false. On the other hand, it doesn't make sense to say that "$2n \leqslant 2 + n$" is either true or false—it depends on what number n is.

The difference between the "n" in the first two predicate forms and in the third predicate form is that, in the first two, they were "bound" by the quantifiers "for all..." or "for every...", whereas in the last one there was no quantifier.

Definition 3.21. *Let $\mathcal{A}$ be a predicate form of a first-order language $\mathcal{L}$, containing the quantifier Qx (where Q is either $\forall$ or $\exists$). The* **scope** *of the quantifier is the predicate form $\mathcal{B}$ such that $Q\,x\,\mathcal{B}$ is a substring of $\mathcal{A}$.*

For example, in the predicate form

$$\forall x\,(\forall y\,\neg Axy \to Bxy)$$

the scope of the quantifier $\forall x$ is

$$(\forall y\,\neg Axy \to Bxy),$$

whereas the scope of the quantifier $\forall y$ is $\neg Axy$.

Definition 3.22. *An occurrence of a variable x in a predicate form $\mathcal{A}$ of a first-order language is* **bound** *if it is either the variable immediately following a quantifier symbol, or it is in the scope of some quantifier involving x. Otherwise it is* **free**.

In the above example, all three occurrences of x are bound, as are the first two occurrences of y. The last occurrence of y is free.

Exercises

11. Identify the scope of each of the quantifiers in the following predicate forms, and indicate whether each occurence of a variable is bound or free. A is a binary predicate symbol and f is a binary function symbol.
    (a) $\forall x\,(Axfxy \to \exists y\,Axy)$.
    (b) $(\forall x\,Axfxy \to \exists y\,Axy)$.
    (c) $(\forall y\,Axfxy \to \exists x\,Axy)$.
12. Identify the scope of each of the quantifiers in the following predicate forms, and indicate whether each occurence of a variable is bound or free. A is a binary predicate symbol and f is a binary function symbol.

(a) $\forall x\,(\exists y\,Axy \to Afxzy)$.

(b) $\forall x\,\exists y\,(Axy \to Afxzy)$.

(c) $(\forall x\,Axy \to \exists y\,Afxzy)$.

(d) $\forall x\,(\exists y\,Axy \to \forall y\,Afxzy)$.

In deciding whether or not an $\mathcal{I}$-assignment v satisfies a predicate form $\mathcal{A}$, we only need to know the values of $v(x)$ for the variables x which occur free in $\mathcal{A}$. We will prove this now.

Proposition 3.23. *Let t be a term, and let v and w be $\mathcal{I}$-assignments such that $v(x) = w(x)$ for every variable x which occurs in t. Then $\tilde{v}(t) = \tilde{w}(t)$.*

Proof. We use induction on the number of function symbols in t.

Base step: if there are no function symbols in t then t is either a constant symbol or a variable. If t is the constant symbol c then we have $\tilde{v}(c) = c^{\mathcal{I}} = \tilde{w}(c)$, and if t is the variable x then

$$\tilde{v}(x) = v(x) = w(x) = \tilde{w}(x)$$

by hypothesis.

Inductive step: Assume t has at least one function symbol, and the result holds for all terms with fewer function symbols than t. Let t be the term $ft_1t_2\dots t_n$. Then, by inductive hypothesis, $\tilde{v}(t_i) = \tilde{w}(t_i)$ for $i = 1, 2, \dots, n$, so

$$\begin{aligned}\tilde{v}(t) &= \tilde{v}(ft_1t_2\dots t_n)\\ &= f^{\mathcal{I}}(\tilde{v}(t_1), \tilde{v}(t_2), \dots, \tilde{v}(t_n))\\ &= f^{\mathcal{I}}(\tilde{w}(t_1), \tilde{w}(t_2), \dots, \tilde{w}(t_n))\\ &= \tilde{w}(ft_1t_2\dots t_n)\\ &= \tilde{w}(t)\end{aligned}$$

This completes the induction.

□

Proposition 3.24. *Let $\mathcal{A}$ be a predicate form and let v and w be $\mathcal{I}$-assignments such that $v(x) = w(x)$ for every variable which occurs free in $\mathcal{A}$. Then*

$$\mathcal{I} \models_v \mathcal{A} \qquad \textit{iff} \qquad \mathcal{I} \models_w \mathcal{A}$$

Proof. We use induction on the complexity of $\mathcal{A}$.

Base: Suppose $\mathcal{A}$ is the atomic formula

$$At_1t_2\dots t_n.$$

Every variable occurring in each t_i is free, so by Proposition 3.23, $\tilde{v}(t_i) = \tilde{w}(t_i)$ for each i. Thus we have

$$\begin{aligned}
\mathcal{I} \models_v \mathcal{A} \quad &\text{iff} \quad \mathcal{I} \models_v At_1t_2\dots t_n \\
&\text{iff} \quad A^{\mathcal{I}}(\tilde{v}(t_1), \tilde{v}(t_2), \dots, \tilde{v}(t_n)) \\
&\text{iff} \quad A^{\mathcal{I}}(\tilde{w}(t_1), \tilde{w}(t_2), \dots, \tilde{w}(t_n)) \\
&\text{iff} \quad \mathcal{I} \models_w At_1t_2\dots t_n \\
&\text{iff} \quad \mathcal{I} \models_w \mathcal{A}
\end{aligned}$$

Inductive step: Assume that $\mathcal{A}$ contains at least one connective or quantifier, and that the result holds for all shorter predicate forms. There are seven cases to consider, depending on whether $\mathcal{A}$ is of the form $\neg\mathcal{B}$, $(\mathcal{B} \wedge \mathcal{C})$, $(\mathcal{B} \vee \mathcal{C})$, $(\mathcal{B} \to \mathcal{C})$, $(\mathcal{B} \leftrightarrow \mathcal{C})$, $\forall x\,\mathcal{B}$ or $\exists x\,\mathcal{B}$. We will omit most of these cases, leaving them as exercises.

Case 1 If $\mathcal{A}$ is of the form $\neg\mathcal{B}$, then

$$\mathcal{I} \models_v \mathcal{A} \quad \text{iff} \quad \mathcal{I} \not\models_v \mathcal{B} \quad \text{iff} \quad \mathcal{I} \not\models_w \mathcal{B} \quad \text{iff} \quad \mathcal{I} \models_w \mathcal{A}$$

Cases 2 and 3 $\mathcal{A}$ has the form $(\mathcal{B} \wedge \mathcal{C})$ or $(\mathcal{B} \vee \mathcal{C})$—exercise.

Case 4 If $\mathcal{A}$ is of the form $(\mathcal{B} \to \mathcal{C})$ then

$$\begin{aligned}
\mathcal{I} \models_v \mathcal{A} \quad &\text{iff} \quad \mathcal{I} \not\models_v \mathcal{B} \text{ or } \mathcal{I} \models_v \mathcal{C} \\
&\text{iff} \quad \mathcal{I} \not\models_w \mathcal{B} \text{ or } \mathcal{I} \models_w \mathcal{C} \\
&\text{iff} \quad \mathcal{I} \models_w \mathcal{A}
\end{aligned}$$

Case 5 $\mathcal{A}$ has the form $(\mathcal{B} \leftrightarrow \mathcal{C})$—exercise.

Case 6 Suppose $\mathcal{A}$ is of the form $\forall x\,\mathcal{B}$. Suppose $\mathcal{I} \models_v \mathcal{A}$. We need to show that $\mathcal{I} \models_w \mathcal{A}$, i.e. that $\mathcal{I} \models_w \forall x\,\mathcal{B}$.

- So let $d \in D$ (where D is the domain of $\mathcal{I}$). Consider $v\frac{d}{x}$ and $w\frac{d}{x}$: for every free variable y of $\mathcal{A}$ we have
$$v\tfrac{d}{x}(y) = v(y) = w(y) = w\tfrac{d}{x}(y).$$
- The only other free variable of $\mathcal{B}$ is (possibly) x, and we have $v\frac{d}{x}(x) = d = w\frac{d}{x}(x)$. So we have $v\frac{d}{x}(y) = w\frac{d}{x}(y)$ for every y which is free in $\mathcal{B}$.
- So by inductive hypothesis we have $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{B}$ iff $\mathcal{I} \models_{w\frac{d}{x}} \mathcal{B}$. Now, since $\mathcal{I} \models_v \forall x\,\mathcal{B}$, we have $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{B}$, so $\mathcal{I} \models_{w\frac{d}{x}} \mathcal{B}$.
- Since $d \in D$ was arbitrary, $\mathcal{I} \models_w \forall x\,\mathcal{B}$, i.e. $\mathcal{I} \models_w \mathcal{A}$.

Similarly, if $\mathcal{I} \models_w \mathcal{A}$ then $\mathcal{I} \models_v \mathcal{A}$.

Case 7 $\mathcal{A}$ is $\exists x\,\mathcal{B}$—exercise.

This completes the induction. □

Definition 3.25. *A predicate form is* **closed** *if it contains no free variables.*

Unlike the situation for arbitrary predicate forms, a closed predicate form must be either true or false in a given interpretation.

Theorem 3.26. *Let $\mathcal{A}$ be a closed predicate form of a first-order language $\mathcal{L}$, and let $\mathcal{I}$ be an interpretation of $\mathcal{L}$. Then either $\mathcal{I} \models \mathcal{A}$ or $\mathcal{I} \models \neg\mathcal{A}$.*

Proof. Let v and w be $\mathcal{I}$-assignments. Then, vacuously, $v(x) = w(x)$ for every free variable of $\mathcal{A}$, because $\mathcal{A}$ doesn't have any free variables. Thus $\mathcal{I} \models_v \mathcal{A}$ if and only if $\mathcal{I} \models_w \mathcal{A}$.

Hence we either have $\mathcal{I} \models_v \mathcal{A}$ for every $\mathcal{I}$-assignment v, in which case $\mathcal{I} \models \mathcal{A}$, or $\mathcal{I} \not\models_v \mathcal{A}$ for every $\mathcal{I}$-assignment v, in which case $\mathcal{I} \models \neg\mathcal{A}$. □

3.5 Logical equivalence

Definition 3.27. *Two predicate forms $\mathcal{A}$ and $\mathcal{B}$ are* **logically equivalent**, *written $\mathcal{A} \Leftrightarrow \mathcal{B}$, if both $\mathcal{A} \Rightarrow \mathcal{B}$ and $\mathcal{B} \Rightarrow \mathcal{A}$. Thus, $\mathcal{A} \Leftrightarrow \mathcal{B}$ means that for every interpretation $\mathcal{I}$ and every $\mathcal{I}$-assignment v , $\mathcal{I} \models_v \mathcal{A}$ iff $\mathcal{I} \models_v \mathcal{B}$.*

Proposition 3.28. *For any predicate form $\mathcal{A}$,*

$$\neg\forall x\, \mathcal{A} \Leftrightarrow \exists x\ \neg\mathcal{A}.$$

Proof. Let $\mathcal{I}$ be an interpretation with domain D and let v be an $\mathcal{I}$-assignment.

Suppose first that $\mathcal{I} \models_v \neg\forall x\, \mathcal{A}$. Then $\mathcal{I} \not\models_v \forall x\, \mathcal{A}$, so there is at least one $d \in D$ with $\mathcal{I} \not\models_{v\frac{d}{x}} \mathcal{A}$. But then there is at least one $d \in D$ with $\mathcal{I} \models_{v\frac{d}{x}} \neg\mathcal{A}$, so $\mathcal{I} \models_v \exists x\, \neg\mathcal{A}$.

Conversely, suppose that $\mathcal{I} \models_v \exists x\, \neg\mathcal{A}$. Then there is some $d \in D$ with $\mathcal{I} \models_{v\frac{d}{x}} \neg\mathcal{A}$, so there is at least one $d \in D$ for which $\mathcal{I} \not\models_{v\frac{d}{x}} \mathcal{A}$, so $\mathcal{I} \not\models_v \forall x\, \mathcal{A}$, so $\mathcal{I} \models_v \neg\forall x\, \mathcal{A}$. □

Proposition 3.29. *Let $\mathcal{A}$ and $\mathcal{B}$ be predicate forms and let x be a variable which does not occur free in $\mathcal{B}$. Then*

$$\forall x\, (\mathcal{A} \to \mathcal{B}) \Leftrightarrow (\exists x\, \mathcal{A} \to \mathcal{B}).$$

Proof. Let $\mathcal{I}$ be an interpretation with domain D and let v be an $\mathcal{I}$-assignment. Suppose first that $\mathcal{I} \models_v \forall x\, (\mathcal{A} \to \mathcal{B})$. We wish to show that $\mathcal{I} \models_v (\exists x\, \mathcal{A} \to \mathcal{B})$. So suppose that $\mathcal{I} \models_v \exists x\, \mathcal{A}$: we must show that $\mathcal{I} \models_v \mathcal{B}$. Since $\mathcal{I} \models_v \exists x\, \mathcal{A}$, there is some $d \in D$ such that $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$. Since $\mathcal{I} \models_v \forall x\, (\mathcal{A} \to \mathcal{B})$ we also have $\mathcal{I} \models_{v\frac{d}{x}} (\mathcal{A} \to \mathcal{B})$, so $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{B}$. Now x does not occur free in $\mathcal{B}$, $v\frac{d}{x}(y) = v(y)$ for all y which occur free in $\mathcal{B}$, so we also have $\mathcal{I} \models_v \mathcal{B}$, as required.

Conversely, suppose that $\mathcal{I} \models_v (\exists x\, \mathcal{A} \to \mathcal{B})$. We must show that $\mathcal{I} \models_v \forall x\, (\mathcal{A} \to \mathcal{B})$. We consider two cases: either $\mathcal{I} \not\models_v \exists x\, \mathcal{A}$ or $\mathcal{I} \models_v \mathcal{B}$.

Case 1: $\mathcal{I} \not\models_v \exists x\, \mathcal{A}$. Then for every $d \in D$ we have $\mathcal{I} \not\models_{v\frac{d}{x}} \mathcal{A}$, so $\mathcal{I} \models_{v\frac{d}{x}} (\mathcal{A} \to \mathcal{B})$. Since this holds for all d, $\mathcal{I} \models_v \forall x\, (\mathcal{A} \to \mathcal{B})$.

Case 2: $\mathcal{I} \models_v \mathcal{B}$. For every $d \in D$ we have $v\frac{d}{x}(y) = v(y)$ for every y which is free in $\mathcal{B}$, so $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{B}$ and therefore $\mathcal{I} \models_{v\frac{d}{x}} (\mathcal{A} \to \mathcal{B})$. Since this holds for every d, $\mathcal{I} \models_v \forall x\, (\mathcal{A} \to \mathcal{B})$.

□

Theorem 3.30. *Let $\mathcal{A}$ and $\mathcal{B}$ be predicate forms with $\mathcal{A} \Leftrightarrow \mathcal{B}$, and let $\mathcal{C}$ be a predicate form which contains (one or more instances of) $\mathcal{A}$. Let $\mathcal{D}$ be a predicate form obtained from $\mathcal{C}$ by replacing some instances of $\mathcal{A}$ with $\mathcal{B}$. Then $\mathcal{C} \Leftrightarrow \mathcal{D}$.*

Proof. We apply induction on the complexity of $\mathcal{C}$.

Base: the simplest possible $\mathcal{C}$ is $\mathcal{C} = \mathcal{A}$. In this case we have $\mathcal{D} = \mathcal{B}$ so $\mathcal{C} \Leftrightarrow \mathcal{D}$ by hypothesis.

Inductive step: Suppose that $\mathcal{C}$ is more complex than $\mathcal{A}$ and that the result holds for all predicate forms less complex than $\mathcal{C}$. We should consider seven cases depending on whether $\mathcal{C}$ has the form $\neg\mathcal{F}$, $(\mathcal{F} \wedge \mathcal{G})$, $(\mathcal{F} \vee \mathcal{G})$, $(\mathcal{F} \to \mathcal{G})$, $(\mathcal{F} \leftrightarrow \mathcal{G})$, $\forall x\, \mathcal{F}$ or $\exists x\, \mathcal{F}$. We will omit most of these: the first five are all very similar to one another and the last two are very similar to one another, so we will only include one from each of these two groups.

Case 2: $\mathcal{C}$ has the form $(\mathcal{F} \wedge \mathcal{G})$. Then $\mathcal{D}$ is $(\mathcal{F}' \wedge \mathcal{G}')$, where either $\mathcal{F} = \mathcal{F}'$ or $\mathcal{F}'$ is obtained from $\mathcal{F}$ by replacing one or more instances of $\mathcal{A}$ with $\mathcal{B}$. Similarly $\mathcal{G}' = \mathcal{G}$ or $\mathcal{G}'$ is obtained from $\mathcal{G}$ by replacing one or more instances of $\mathcal{A}$ with $\mathcal{B}$. Either way, by inductive hypothesis we have $\mathcal{F} \Leftrightarrow \mathcal{F}'$ and $\mathcal{G} \Leftrightarrow \mathcal{G}'$. So for any interpretation $\mathcal{I}$ and any $\mathcal{I}$-assignment v ,

$$\begin{aligned} \mathcal{I} \models_v \mathcal{C} &\text{ iff } (\mathcal{I} \models_v \mathcal{F} \text{ and } \mathcal{I} \models_v \mathcal{G}) \\ &\text{iff } (\mathcal{I} \models_v \mathcal{F}' \text{ and } \mathcal{I} \models_v \mathcal{G}') \\ &\text{iff } \mathcal{I} \models_v \mathcal{D} \end{aligned}$$

Case 6: $\mathcal{C}$ has the form $\forall x\, \mathcal{F}$. Then $\mathcal{D}$ is $\forall x\, \mathcal{G}$ where $\mathcal{G}$ is obtained from $\mathcal{F}$ by replacing some or all instances of $\mathcal{A}$ with $\mathcal{B}$. By inductive hypothesis, $\mathcal{F} \Leftrightarrow \mathcal{G}$. So if $\mathcal{I}$ is an interpretation with domain D and v is an $\mathcal{I}$-assignment then

$$\begin{aligned} \mathcal{I} \models_v \mathcal{C} &\text{ iff } \text{ for all } d \in D, \mathcal{I} \models_{v\frac{d}{x}} \mathcal{F} \\ &\text{iff } \text{ for all } d \in D, \mathcal{I} \models_{v\frac{d}{x}} \mathcal{G} \\ &\text{iff } \mathcal{I} \models_v \mathcal{D} \end{aligned}$$

□

Corollary 3.31. *For any predicate form $\mathcal{A}$ and any variable x, $\neg\exists x\,\mathcal{A} \Leftrightarrow \forall x\,\neg\mathcal{A}$.*

Proof. By Proposition 3.28, applied to $\neg\mathcal{A}$, we have $\neg\forall x\,\neg\mathcal{A} \Leftrightarrow \exists x\,\neg\neg\mathcal{A}$. Therefore,

$$\begin{aligned}\neg\exists x\,\mathcal{A} &\Leftrightarrow \neg\exists x\,\neg\neg\mathcal{A}\\ &\Leftrightarrow \neg\neg\forall x\,\neg\mathcal{A}\\ &\Leftrightarrow \forall x\,\neg\mathcal{A}\end{aligned}$$

□

Exercises

13. Show that $\forall x\,(\mathcal{A} \to \mathcal{B}) \Rightarrow (\exists x\,\mathcal{A} \to \exists x\,\mathcal{B})$.
14. Show that $\forall x\,\exists y\,(Ax \vee By) \Leftrightarrow \exists y\,\forall x\,(Ax \vee By)$.

3.6 Changing bound variables

A predicate form like

$$(\forall x\,Axy \vee Bxy)$$

is perfectly legitimate according to our rules. However it is confusing because x occurs free in Bxy and also occurs bound in $\forall x\,Axy$. It would be less confusing to use the predicate form

$$(\forall z\,Azy \vee Bxy)$$

so that the free variables are not also used as bound variables. We will indicate that the predicate form we get by changing the bound variable is logically equivalent to the original predicate form, provided we use a completely new variable (obviously changing x into y in our example above would be liable to change the meaning).

As well as avoiding possible confusion, the ability to change bound variables is important in changing predicate forms to logically equivalent ones in a standard form, which we will be doing later. We have a rule which says that $(\exists x\,\mathcal{A} \to \mathcal{B}) \Leftrightarrow \forall x\,(\mathcal{A} \to \mathcal{B})$, provided x does not occur free in $\mathcal{B}$. This is all very well, but what do we do if x *does* occur free in $\mathcal{B}$?

The answer is that we replace x with a different variable which does not occur free in $\mathcal{B}$.

We write $\mathcal{A}\left(\frac{z}{x}\right)$ for the predicate form obtained by replacing every **free** instance of x with z.

For instance, if $\mathcal{A}$ is $(\forall x\,Axy \to Bxy)$, then $\mathcal{A}\left(\frac{z}{x}\right)$ is $(\forall x\,Axy \to Bzy)$. Only the second, free, instance got replaced.

Then if we want to apply the rule above, we would have to do it in two steps as

$$(\exists x\, \mathcal{A} \to \mathcal{B}) \Leftrightarrow (\exists z\, \mathcal{A}(\tfrac{z}{x}) \to \mathcal{B})$$
$$\Leftrightarrow \forall z\, (\mathcal{A}(\tfrac{z}{x}) \to \mathcal{B}).$$

Theorem 3.32. *Let $\mathcal{A}$ be a predicate form in which x occurs free, and let z be a variable which does not occur, free or bound, in $\mathcal{A}$. Then*

(i) $\forall x\, \mathcal{A} \Leftrightarrow \forall z\, \mathcal{A}\left(\frac{z}{x}\right)$, *and*

(ii) $\exists x\, \mathcal{A} \Leftrightarrow \exists z\, \mathcal{A}\left(\frac{z}{x}\right)$.

Proof is at the end of this Section (but not required). You only have to prove the first, as the second can be derived from it (exercise).

Example 3.33. By the theorem, applied to $\mathcal{A}$ being Axy,

$$\exists x\, Axy \Leftrightarrow \exists z\, Azy.$$

Then by Theorem 3.30

$$(\exists x\, Axy \to Bxy) \Leftrightarrow (\exists z\, Azy \to Bxy)$$

Now that we have changed the bound variable x to z, which doesn't occur free in Bxy any more, we can use Proposition 3.28: the latter is equivalent to $\forall z\, (Azy \to Bxy)$

We now give the proof of Theorem 3.32. First we need a lemma taking care of what happens to terms when we replace x with z.

Lemma 3.34. *Let x and z be variables. For every term t, let t' be the term we get by replacing every x in t with z. Let v be an $\mathcal{I}$-assignment. Then, for every term t, $\tilde{v}(t') = \widetilde{v\frac{d}{x}}(t)$, where $d = v(z)$.*

Proof. * We use induction on the complexity of t.

Base: t contains no function symbols. There are three cases to consider: t is a constant, $t = y$ for some variable $y \neq x$, and $t = x$.

Case 1: $t = c$ for some constant symbol c. Then $t' = c$ and $\tilde{v}(t') = c^{\mathcal{I}} = \widetilde{v\frac{d}{x}}(c) = \widetilde{v\frac{d}{x}}(t)$.

Case 2: $t = y$ for some variable $y \neq x$. Then $t' = t$ and $\tilde{v}(t') = v(y) = v\frac{d}{x}(y) = \widetilde{v\frac{d}{x}}(t)$.

Case 3: $t = x$. Then $t' = z$ and $\tilde{v}(t') = v(z) = d = \widetilde{v\frac{d}{x}}(x) = \widetilde{v\frac{d}{x}}(t)$.

Inductive step: Suppose t is the term $ft_1t_2\dots t_n$, and that the result holds for all simpler terms than t. Then $t' = ft'_1t'_2\dots t'_n$, and by inductive hypothesis we have $\tilde{v}(t'_i) = \widetilde{v\frac{d}{x}}(t_i)$ for each i, so

$$\begin{aligned}\tilde{v}(t') &= \tilde{v}(ft'_1t'_2\dots t'_n)\\ &= f^{\mathcal{I}}(\tilde{v}(t'_1), \tilde{v}(t'_2), \dots, \tilde{v}(t'_n))\\ &= f^{\mathcal{I}}(\widetilde{v\tfrac{d}{x}}(t_1), \widetilde{v\tfrac{d}{x}}(t_2), \dots, \widetilde{v\tfrac{d}{x}}(t_n))\\ &= \widetilde{v\tfrac{d}{x}}(ft_1t_2\dots t_n)\\ &= \widetilde{v\tfrac{d}{x}}(t).\end{aligned}$$

□

Next we need the following, a special case of the Substitution Lemma below.

Lemma 3.35. *Let $\mathcal{A}$ be a predicate form, let z be a variable which does not occur (free or bound) in $\mathcal{A}$. Let $\mathcal{I}$ be an interpretation and let v be an $\mathcal{I}$-assignment. Then $\mathcal{I} \models_v \mathcal{A}(\frac{z}{x})$ iff $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$, where $d = v(z)$.*

Proof. * We use induction on the complexity of $\mathcal{A}$.

Base: $\mathcal{A}$ is the atomic formula $At_1t_2\dots t_n$. Then $\mathcal{A}(\frac{z}{x})$ is $At'_1t'_2\dots t'_n$, where t'_i is the result of replacing each x in t_i with z. By the previous lemma, $\tilde{v}(t'_i) = \widetilde{v\frac{d}{x}}(t_i)$ for each i, so

$$\begin{aligned}&\mathcal{I} \models_v At'_1t'_2\dots t'_n\\ &\text{iff } A^{\mathcal{I}}(\tilde{v}(t'_1), \tilde{v}(t'_2), \dots, \tilde{v}(t'_n)) \text{ holds}\\ &\text{iff } A^{\mathcal{I}}(\widetilde{v\tfrac{d}{x}}(t_1), \widetilde{v\tfrac{d}{x}}(t_2), \dots, \widetilde{v\tfrac{d}{x}}(t_n)) \text{ holds}\\ &\text{iff } \mathcal{I} \models_{v\frac{d}{x}} At_1t_2\dots t_n\end{aligned}$$

Inductive step: Suppose $\mathcal{A}$ is not an atomic formula and the result holds for all simpler predicate forms. There are seven cases to consider, depending on whether $\mathcal{A}$ has the form $\neg\mathcal{B}$, $(\mathcal{B} \wedge \mathcal{C})$, $(\mathcal{B} \vee \mathcal{C})$, $(\mathcal{B} \to \mathcal{C})$, $(\mathcal{B} \leftrightarrow \mathcal{C})$, $\forall x\,\mathcal{B}$ or $\exists x\,\mathcal{B}$. As before we omit all but two of these cases.

Case 4: $\mathcal{A}$ is $(\mathcal{B} \to \mathcal{C})$. Then $\mathcal{A}(\frac{z}{x})$ is $(\mathcal{B}(\frac{z}{x}) \to \mathcal{C}(\frac{z}{x}))$, and by inductive hypothesis $\mathcal{I} \models_v \mathcal{B}(\frac{z}{x})$ iff $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{B}$ and $\mathcal{I} \models_v \mathcal{C}(\frac{z}{x})$ iff $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{C}$. So we have

$$\begin{aligned}\mathcal{I} \models_v \mathcal{A}\left(\frac{z}{x}\right) &\text{ iff } \mathcal{I} \not\models_v \mathcal{B}\left(\frac{z}{x}\right) \text{ or } \mathcal{I} \models_v \mathcal{C}\left(\frac{z}{x}\right)\\ &\text{iff } \mathcal{I} \not\models_{v\frac{d}{x}} \mathcal{B} \text{ or } \mathcal{I} \models_{v\frac{d}{x}} \mathcal{C}\\ &\text{iff } \mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}\end{aligned}$$

Case 7: $\mathcal{A}$ is $\exists y\,\mathcal{B}$ for some y. We need to split this into two subcases, depending on whether or not x is free in $\exists y\,\mathcal{B}$:

Case 7a: $\mathcal{A}$ is $\exists y\,\mathcal{B}$ for some y and x does not occur free in $\mathcal{A}$. In this case we have $\mathcal{A}(\frac{z}{x}) = \mathcal{A}$ and $v(y') = v\frac{d}{x}(y')$ for every y' which occurs free in $\mathcal{A}$, so $\mathcal{I} \models_v \mathcal{A}(\frac{z}{x})$ iff $\mathcal{I} \models_v \mathcal{A}$ iff $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$.

Case 7b: $\mathcal{A}$ is $\exists y\,\mathcal{B}$ for some y and x occurs free in $\mathcal{A}$. This tells us that $y \neq x$ (and we also know that $y \neq z$ since z does not occur in $\mathcal{A}$. Suppose $\mathcal{I} \models_v \mathcal{A}(\frac{z}{x})$, i.e. $\mathcal{I} \models_v \exists y\,\mathcal{B}(\frac{z}{x})$. Then there is some e such that $\mathcal{I} \models_{v\frac{e}{y}} \mathcal{B}(\frac{z}{x})$. Applying the inductive hypothesis to the predicate form $\mathcal{B}$ and the $\mathcal{I}$-assignment $u = v\frac{e}{y}$, we have $\mathcal{I} \models_u \mathcal{B}(\frac{z}{x})$ iff $\mathcal{I} \models_{u\frac{d'}{z}}$, where

$$d' = u(z) = v\tfrac{e}{y}(z) = v(z) = d.$$

Thus, since $\mathcal{I} \models_{v\frac{e}{y}} \mathcal{B}(\frac{z}{x})$, $\mathcal{I} \models_{v\frac{e}{y}\frac{d}{x}} \mathcal{B}$. Since $v\frac{e}{y}\frac{d}{x} = v\frac{d}{x}\frac{e}{y}$, we have $\mathcal{I} \models_{v\frac{d}{x}\frac{e}{y}} \mathcal{B}$, so $\mathcal{I} \models_{v\frac{d}{x}} \exists y\,\mathcal{B}$. Thus $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$, as required.

Similarly if $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$ then $\mathcal{I} \models_v \mathcal{A}(\frac{z}{x})$.

□

We are now in the position to prove Theorem 3.32.

Proof. (i) Let $\mathcal{I}$ be an interpretation with domain D, and let v be a $\mathcal{I}$-assignment. Suppose $\mathcal{I} \models_v \forall x\,\mathcal{A}$. [We will show that $\mathcal{I} \models_v \forall z\,\mathcal{A}(\frac{z}{x})$.] Let $d \in D$. [We must show that $\mathcal{I} \models_{v\frac{d}{z}} \mathcal{A}(\frac{z}{x})$.] By the previous lemma, $\mathcal{I} \models_{v\frac{d}{z}} \mathcal{A}(\frac{z}{x})$ iff $\mathcal{I} \models_{v\frac{d}{z}\frac{e}{x}} \mathcal{A}$, where $e = v\frac{d}{z}(z) = d$. Since $\mathcal{I} \models_v \forall x\,\mathcal{A}$, $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$. Since $v\frac{d}{x}(y) = v\frac{d}{z}\frac{d}{x}(y)$ for every y which occurs free in $\mathcal{A}$, we have $\mathcal{I} \models_{v\frac{d}{z}\frac{d}{x}} \mathcal{A}$. So $\mathcal{I} \models_{v\frac{d}{z}} \mathcal{A}(\frac{z}{x})$. Since this holds for all $d \in D$, $\mathcal{I} \models_v \forall z\,\mathcal{A}(\frac{z}{x})$.

Conversely, suppose that $\mathcal{I} \models_v \forall z\,\mathcal{A}(\frac{z}{x})$. [We will show that $\mathcal{I} \models_v \forall x\,\mathcal{A}$.] Let $d \in D$. [We will show that $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$.] Since $\mathcal{I} \models_v \forall z\,\mathcal{A}(\frac{z}{x})$, $\mathcal{I} \models_{v\frac{d}{z}} \mathcal{A}(\frac{z}{x})$, and by the previous lemma $\mathcal{I} \models_{v\frac{d}{z}\frac{e}{x}} \mathcal{A}$, where $e = \widetilde{v\frac{d}{z}}(z) = d$. So $\mathcal{I} \models_{v\frac{d}{z}\frac{d}{x}} \mathcal{A}$, and since $v\frac{d}{x}(y) = v\frac{d}{z}\frac{d}{x}(y)$ for all y which occur free in $\mathcal{A}$ we have $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$ as required.

(ii) Exercise.

□

3.7 Substitutions

Recall that if $\mathcal{A}$ is a predicate form, then $\mathcal{A}(\frac{z}{x})$ is the predicate form obtained from $\mathcal{A}$ by replacing every **free** occurrence of x with z. In general, if t is any term, then we denote by $\mathcal{A}(\frac{t}{x})$ the predicate form we get by replacing every free occurrence of x in $\mathcal{A}$ with t.

For example, let $\mathcal{A}$ denote the predicate form $(\forall x)Lxy$.

Then $\mathcal{A}(\frac{z}{y})$ is $\forall x\, Lxz$, $\mathcal{A}\left(\frac{gy}{y}\right)$ is $\forall x\, Lxgy$.

Intuitively, $\mathcal{A}$, $\mathcal{A}(\frac{w}{y})$, $\mathcal{A}\left(\frac{gy}{y}\right)$ and so on all express the same property of objects represented by y, w, gy and so on.

On the other hand, $\mathcal{A}(\frac{x}{y})$ is $\forall x\, Lxx$, which is different from all of the others—it does not have any free variables. In substituting x for y, we get "tangled up" by the quantifier $\forall x$. This is the idea behind the following definition.

Definition 3.36. *Let $\mathcal{A}$ be a predicate form, and t be a term. Then t is* **free to substitute for** y **in** $\mathcal{A}$*, or, in brief,* **free for** y **in** $\mathcal{A}$ *if y does not occur free in $\mathcal{A}$ within the scope of any quantifier $Q\,x$ such that the variable x occurs in t.*

Example 3.37. Let $\mathcal{A}$ be the predicate form

$$(\forall x\, Axy \to \forall z\, (Bx \vee Ayz))$$

Perform each of the following substitutions, and state whether or not each is a free substitution.

1. x for y
2. y for x
3. x for x
4. x for z
5. z for x
6. fxz for y

Exercises

15. Let $\mathcal{A}$ be the predicate form

$$(\forall x\, (Axy \to \exists z\, Afxyz \vee Bx),$$

where A is a binary predicate, f is a binary function and B is a unary predicate.

(a) Find the scope of the quantifiers and indicate which occurences of variables are free.

(b) Perform the following substitutions and indicate which of these are free substitutions.

i. x for y.

ii. y for x.

iii. z for x.

iv. fxy for y.

16. Consider the following four predicate forms from Exercise 12.

$$\forall x\,(\exists y\,Axy \to Afxzy)$$
$$\forall x\,\exists y\,(Axy \to Afxzy)$$
$$(\forall x\,Axy \to \exists y\,Afxzy)$$
$$\forall x\,(\exists y\,Axy \to \forall y\,Afxzy)$$

(a) Perform the substitution x for y in each of these four predicate forms.

(b) Complete the following table indicating which substitutions are free and which are not free in each of these four predicate forms.

	(1)	(2)	(3)	(4)
x for y	$\times$	$\surd$	$\times$	$\surd$
z for x				
y for z				
x for z				
fxy for x				
fyy for y				

Proposition 3.38.

- *Let t be a term and x a variable.*
- *For any term s let s' be the term obtained from s by replacing each x with t (often denoted by $s(\frac{t}{x})$).*
- *Let v be an $\mathcal{I}$-assignment and put $d = \tilde{v}(t)$.*

Then for every s we have

$$\tilde{v}(s') = \widetilde{v\tfrac{d}{x}}(s).$$

Proof. This is an exercise at the end of this Section. □

Proposition 3.39 (Substitution Lemma)**.**

- *Let $\mathcal{A}$ be a predicate form, and let t be a term which is free for x in $\mathcal{A}$.*
- *Let v be an $\mathcal{I}$-assignment in an interpretation $\mathcal{I}$.*

- *Put* $d = \tilde{v}(t)$.

Then

$$\mathcal{I} \models_v \mathcal{A}\left(\frac{t}{x}\right) \qquad \textit{iff} \qquad \mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}.$$

For a real life analogy, consider the language with a binary predicate symbol Lyx ("y likes x") and a unary function symbol f (we intend to interpret fx as the father of x). We use the interpretation I where the domain is the people in the US. Consider an $\mathcal{I}$-assignment ν where $\nu(x)$ =George W. Bush and $\nu(y)$ =Hilary Clinton. Let the predicate form $\mathcal{A}$ be Lyx and let the term t be fx. Now analyze the meaning of both sides and realize they are equivalent.

Proof. * We use induction on the complexity of $\mathcal{A}$.

Base: $\mathcal{A}$ is the atomic formula $As_1s_2\dots s_n$. For each i let s_i' be the term obtained from s_i by replacing each x with t. Then, by Proposition 3.38, $\tilde{v}(s_i') = \widetilde{v\frac{d}{x}}(s_i)$ for each i. We also have $\mathcal{A}(\frac{t}{x}) = As_1's_2'\dots s_n'$. So we have

$$\begin{aligned}
\mathcal{I} \models_v \mathcal{A}(\frac{t}{x}) &\text{ iff } \mathcal{I} \models_v As_1's_2'\dots s_n' \\
&\text{iff } A^{\mathcal{I}}(\tilde{v}(s_1'), \tilde{v}(s_2'), \dots, \tilde{\tilde{v}}(s_n')) \\
&\text{iff } A^{\mathcal{I}}(\widetilde{v\tfrac{d}{x}}(s_1), \widetilde{v\tfrac{d}{x}}(s_2), \dots, \widetilde{v\tfrac{d}{x}}(s_n)) \\
&\text{iff } \mathcal{I} \models_{v\frac{d}{x}} As_1s_2\dots s_n \\
&\text{iff } \mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}
\end{aligned}$$

Inductive step: Suppose that $\mathcal{A}$ is not an atomic formula and the result holds for all simpler predicate forms. We have seven cases to consider, depending on whether $\mathcal{A}$ has the form $\neg\mathcal{B}$, $(\mathcal{B} \wedge \mathcal{C})$, $(\mathcal{B} \vee \mathcal{C})$, $(\mathcal{B} \to \mathcal{C})$, $(\mathcal{B} \leftrightarrow \mathcal{C})$, $\forall y\,\mathcal{B}$ or $\exists y\,\mathcal{B}$. In the latter two we should consider separately the cases where x does or does not occur free in $\mathcal{A}$. As usual we will omit most of these cases.

Case 2: $\mathcal{A}$ has the form $(\mathcal{B} \wedge \mathcal{C})$. Then $\mathcal{A}(\frac{t}{x})$ is $(\mathcal{B}(\frac{t}{x}) \wedge \mathcal{C}(\frac{t}{x}))$, and by inductive hypothesis $\mathcal{I} \models_v \mathcal{B}\left(\frac{t}{x}\right)$ iff $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{B}$ and $\mathcal{I} \models_v \mathcal{C}(\frac{t}{x})$ iff $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{C}$, so

$$\begin{aligned}
\mathcal{I} \models_v \mathcal{A}\left(\frac{t}{x}\right) &\text{ iff } \mathcal{I} \models_v \mathcal{B}\left(\frac{t}{x}\right) \text{ and } \mathcal{I} \models_v \mathcal{C}\left(\frac{t}{x}\right) \\
&\text{iff } \mathcal{I} \models_{v\frac{d}{x}} \mathcal{B} \text{ and } \mathcal{I} \models_{v\frac{d}{x}} \mathcal{C} \\
&\text{iff } \mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}
\end{aligned}$$

Case 6(a): $\mathcal{A}$ has the form $\forall y\,\mathcal{B}$, and x does not occur free in $\mathcal{A}$. Then $\mathcal{A}(\frac{t}{x})$ is $\mathcal{A}$, and also $v(z) = v\frac{d}{x}(z)$ for all z which occur free in $\mathcal{A}$, so, by Proposition 3.24, $\mathcal{I} \models_v \mathcal{A}(\frac{t}{x})$ iff $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$.

Case 6(b): $\mathcal{A}$ has the form $\forall y\,\mathcal{B}$ for some y, and x occurs free in $\mathcal{A}$. Then we must have $y \neq x$, and also $\mathcal{A}(\frac{t}{x})$ is $\forall y\,\mathcal{B}(\frac{t}{x})$. We will show that for every $e \in D$,

$$\mathcal{I} \models_{v\frac{e}{y}} \mathcal{B}(\tfrac{t}{x}) \text{ iff } \mathcal{I} \models_{v\frac{d}{x}\frac{e}{y}} \mathcal{B}.$$

This follows from the inductive hypothesis, provided that we check that $d = \widetilde{v\frac{e}{y}}(t)$ [recall that $d = \tilde{v}(t)$].

Well, we know that t is free for x in $\mathcal{A}$, and we know that x occurs free in the scope of $\forall y$, so y does not occur in t. Thus $v(z) = v\frac{e}{y}(z)$ for every z which occurs in t, so

$$\widetilde{v\tfrac{e}{y}}(t) = \tilde{v}(t) = d,$$

as required. So by inductive hypothesis (applied to $\mathcal{B}$ with the assignment $v\frac{e}{y}$) we do have $\mathcal{I} \models_{v\frac{e}{y}} \mathcal{B}(\frac{t}{x})$ iff $\mathcal{I} \models_{v\frac{e}{y}\frac{d}{x}} \mathcal{B}$ for every $e \in D$. So

$$\begin{aligned}
\mathcal{I} \models_v \mathcal{A}\left(\frac{t}{x}\right) &\text{ iff } \mathcal{I} \models_v \forall y\,\mathcal{B}\left(\frac{t}{x}\right) \\
&\text{iff for all } e \in D,\ \mathcal{I} \models_{v\frac{e}{y}} \mathcal{B}\left(\frac{t}{x}\right) \\
&\text{iff for all } e \in D,\ \mathcal{I} \models_{v\frac{e}{y}\frac{d}{x}} \mathcal{B} \\
&\text{iff for all } e \in D,\ \mathcal{I} \models_{v\frac{d}{x}\frac{e}{y}} \mathcal{B} \\
&\text{iff } \mathcal{I} \models_{v\frac{d}{x}} \forall y\,\mathcal{B} \\
&\text{iff } \mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}
\end{aligned}$$

□

Proposition 3.40. *Let $\mathcal{A}$ be a wff, and let t be a term which is free for x in $\mathcal{A}$. Then*

$$\forall x\,\mathcal{A} \Rightarrow \mathcal{A}\left(\frac{t}{x}\right).$$

Proof.

- Let $\mathcal{I}$ be an interpretation and let v be an $\mathcal{I}$-assignment. Suppose that $\mathcal{I} \models_v \forall x\,\mathcal{A}$. We must show that $\mathcal{I} \models_v \mathcal{A}(\frac{t}{x})$.
- Put $d = v(t)$. Then, by the Substitution Lemma 3.39, $\mathcal{I} \models_v \mathcal{A}(\frac{t}{x})$ iff $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$.
- Since $\mathcal{I} \models_v \forall x\,\mathcal{A}$, $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$, so $\mathcal{I} \models_v \mathcal{A}(\frac{t}{x})$ as required.

□

Exercises

17* Let t be a term and x a variable. For any term s let s' be the term obtained from s by replacing each x with t. Let v be an $\mathcal{I}$-assignment and put $d = v(t)$. Prove that for every s we have $\tilde{v}(s') = \widetilde{v\frac{d}{x}}(s)$.

3.8 Prenex normal form

In Chapter 1 we saw that every statement form is logically equivalent to a statement form in disjunctive normal form. We now show that every predicate form is logically equivalent to one in prenex normal form. As before, it is desirable to have a normal form as it can make the formula easier to read. Prenex normal form also gives a measure of complexity of a predicate form, determined by the number of quantifiers.

Definition 3.41. *A predicate form $\mathcal{A}$ is in* **prenex normal form** *if it is in the form*

$$Q_1\, x_1\, Q_2\, x_2 \ldots Q_n\, x_n\, \mathcal{B}$$

where each Q_i is either $\forall$ or $\exists$ and $\mathcal{B}$ is a predicate form containing no quantifiers. We allow the possibility that $n = 0$, in other words if $\mathcal{A}$ has no quantifiers in it at all then it is in prenex normal form.

Just as we saw that every statement form is logically equivalent to a statement form in disjunctive normal form, we will show that every predicate form is logically equivalent to one which is in prenex normal form.

Theorem 3.42. *Let $\mathcal{A}$ and $\mathcal{B}$ be wffs and let x be a variable which does not occur free in $\mathcal{B}$. Then we have*

$$(\forall x\, \mathcal{A} \wedge \mathcal{B}) \Leftrightarrow \forall x\, (\mathcal{A} \wedge \mathcal{B}) \tag{3.2}$$
$$(\exists x\, \mathcal{A} \wedge \mathcal{B}) \Leftrightarrow \exists x\, (\mathcal{A} \wedge \mathcal{B}) \tag{3.3}$$
$$(\mathcal{B} \wedge \forall x\, \mathcal{A}) \Leftrightarrow \forall x\, (\mathcal{B} \wedge \mathcal{A}) \tag{3.4}$$
$$(\mathcal{B} \wedge \exists x\, \mathcal{A}) \Leftrightarrow \exists x\, (\mathcal{B} \wedge \mathcal{A}) \tag{3.5}$$
$$(\forall x\, \mathcal{A} \vee \mathcal{B}) \Leftrightarrow \forall x\, (\mathcal{A} \vee \mathcal{B}) \tag{3.6}$$
$$(\exists x\, \mathcal{A} \vee \mathcal{B}) \Leftrightarrow \exists x\, (\mathcal{A} \vee \mathcal{B}) \tag{3.7}$$
$$(\mathcal{B} \vee \forall x\, \mathcal{A}) \Leftrightarrow \forall x\, (\mathcal{B} \vee \mathcal{A}) \tag{3.8}$$
$$(\mathcal{B} \vee \exists x\, \mathcal{A}) \Leftrightarrow \exists x\, (\mathcal{B} \vee \mathcal{A}) \tag{3.9}$$
$$(\forall x\, \mathcal{A} \to \mathcal{B}) \Leftrightarrow \exists x\, (\mathcal{A} \to \mathcal{B}) \tag{3.10}$$
$$(\exists x\, \mathcal{A} \to \mathcal{B}) \Leftrightarrow \forall x\, (\mathcal{A} \to \mathcal{B}) \tag{3.11}$$
$$(\mathcal{B} \to \forall x\, \mathcal{A}) \Leftrightarrow \forall x\, (\mathcal{B} \to \mathcal{A}) \tag{3.12}$$
$$(\mathcal{B} \to \exists x\, \mathcal{A}) \Leftrightarrow \exists x\, (\mathcal{B} \to \mathcal{A}) \tag{3.13}$$

Proof. We will only check the first, leave the second as an exercise, and show how to deduce the others from these two.

Let $\mathcal{I}$ be an interpretation with domain D and let v be a $\mathcal{I}$-assignment. Suppose that $\mathcal{I} \models_v (\forall x\, \mathcal{A} \wedge \mathcal{B})$. We want to show that $\mathcal{I} \models_v \forall x\, (\mathcal{A} \wedge \mathcal{B})$. So let $d \in D$: we must show that $\mathcal{I} \models_{v\frac{d}{x}} (\mathcal{A} \wedge \mathcal{B})$. Since we know that $\mathcal{I} \models_v \forall x\, \mathcal{A}$ and $\mathcal{I} \models_v \mathcal{B}$, we know that $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$. Also, since $v\frac{d}{x}(y) = v(y)$ for every y which occurs free in $\mathcal{B}$, $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{B}$. Hence $\mathcal{I} \models_{v\frac{d}{x}} (\mathcal{A} \wedge \mathcal{B})$, as required.

Conversely, suppose $\mathcal{I} \models_v \forall x\, (\mathcal{A} \wedge \mathcal{B})$: we must show that $\mathcal{I} \models_v (\forall x\, \mathcal{A} \wedge \mathcal{B})$. For every $d \in D$ we have $\mathcal{I} \models_{v\frac{d}{x}} (\mathcal{A} \wedge \mathcal{B})$, so $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$. Since this holds for all d, $\mathcal{I} \models_v \forall x\, \mathcal{A}$. Also, putting $e = v(x)$ we have $v = v\frac{e}{x}$, and we know that $\mathcal{I} \models_{v\frac{e}{x}} (\mathcal{A} \wedge \mathcal{B})$, so $\mathcal{I} \models_{v\frac{e}{x}} \mathcal{B}$, i.e. $\mathcal{I} \models_v \mathcal{B}$. So $\mathcal{I} \models_v (\forall x\, \mathcal{A} \wedge \mathcal{B})$, as required.

We leave as an exercise the proof that $((\exists x)\mathcal{A} \wedge \mathcal{B}) \Leftrightarrow \exists x\, (\mathcal{A} \wedge \mathcal{B})$.

The others all follow from these two and the rules we already know for logical equivalences. For example, let us do the nineth. we have $(p \to q) \Leftrightarrow \neg(p \wedge \neg q)$ so we have

$$\begin{aligned}(\forall x\, \mathcal{A} \to \mathcal{B}) &\Leftrightarrow \neg(\forall x\, \mathcal{A} \wedge \neg\mathcal{B}) \\ &\Leftrightarrow \neg\forall x\, (\mathcal{A} \wedge \neg\mathcal{B}) \\ &\Leftrightarrow \exists x\, \neg(\mathcal{A} \wedge \neg\mathcal{B}) \\ &\Leftrightarrow \exists x\, (\mathcal{A} \to \mathcal{B})\end{aligned}$$

□

Example 3.43. Find a predicate form in prenex normal form which is logically equivalent to

$$(\exists x\, Ax \to \exists y \forall z\, Byz).$$

Solution. We have

$$\begin{aligned}&(\exists x\, Ax \to \exists y \forall z\, Byz) \\ \Leftrightarrow\ &\forall x\, (Ax \to \exists y \forall z\, Byz) \\ \Leftrightarrow\ &\forall x \exists y\, (Ax \to \forall z\, Byz) \\ \Leftrightarrow\ &\forall x \exists y \forall z\, (Ax \to Byz)\end{aligned}$$

□

In this case all the variables were distinct so we could simply apply the equivalences from Theorem 3.42. If this is not the case, we have to change bound variables first so that we can move quantifiers past the free variables.

Example 3.44. Find a predicate form in prenex normal form which is logically equivalent to

$$(\exists x\, Ax \to \forall x \forall y\, Bxy).$$

Solution. We have

$$\begin{aligned}
&(\exists x\, Ax \to \forall x \forall y\, Bxy) \\
\Leftrightarrow\ &\forall x\, (Ax \to \forall x \forall y\, Bxy) \\
\Leftrightarrow\ &\forall x\, (Ax \to \forall z \forall y\, Bzy) \\
\Leftrightarrow\ &\forall x \forall z\, (Ax \to \forall y\, Bzy) \\
\Leftrightarrow\ &\forall x \forall z \forall y\, (Ax \to Bzy)
\end{aligned}$$

□

Using this technique it is easy to see that we can rewrite every predicate form in prenex normal form. To prove this carefully, we would use induction on the complexity of the predicate form.

Exercises

17. Find a predicate form in prenex normal form which is logically equivalent to

$$((\forall y\, Axy \to Bx) \vee (Bx \to \exists y\, Axy))$$

18. Find a predicate form in prenex normal form which is logically equivalent to

$$((Ax \to \forall y\, Bxy) \vee \forall x\, (Bxy \wedge Ay)).$$

Chapter 4

Formal Predicate Logic

We have seen a semantic approach to predicate logic, giving meaning through interpretations and assignments. We will now consider a syntactic approach, giving a Post Production System. Essentially, we are extending the system L we used for statement logic. But an important difference to L is that the wff's are not statement forms any longer, but rather predicate forms in a first order language $\mathcal{L} = (C, P, F)$. In particular, our system now depends on the choice of the symbols given by C, P and F (often called "non-logical" symbols).

Once again, we will restrict ourselves to a more limited range of "logical" symbols to make the system simpler to describe and analyse. We have already seen that all the connectives can be represented using just $\neg$ and $\rightarrow$. Similarly, since $\exists x\mathcal{A} \Leftrightarrow \neg\forall x\,\neg\mathcal{A}$, we can eliminate all existential quantifiers: every predicate form is logically equivalent to one which uses only the connectives $\neg$ and $\rightarrow$ and only the quantifier $\forall$.

Again, the goal is to show that the system we describe is sound and adequate: the theorems are precisely the predicate forms that are logically valid. A proof of adequacy is too tricky for this course: we will defer that proof to the course MATHS 713. You can also read the proof in *Logic for Mathematicians*, Section 4.4.

4.1 The formal system $K_{\mathcal{L}}$

Let $\mathcal{L} = (C, P, F)$ be a first-order language. We define the Post production system $K_{\mathcal{L}}$ as follows.

The wffs of $K_{\mathcal{L}}$ are precisely those predicate forms which can be built using the symbols of $\mathcal{L}$ and the connectives $\neg$ and $\rightarrow$ and the quantifier $\forall$. We denote the set of wffs of $K_{\mathcal{L}}$ by form$(K_{\mathcal{L}})$.

For any wffs $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ of $\mathcal{L}$, any variable x and any term t, the following are axioms.

(K1) $(\mathcal{A} \to (\mathcal{B} \to \mathcal{A}))$

(K2) $((\mathcal{A} \to (\mathcal{B} \to \mathcal{C})) \to ((\mathcal{A} \to \mathcal{B}) \to (\mathcal{A} \to \mathcal{C})))$

(K3) $((\neg\mathcal{A} \to \neg\mathcal{B}) \to ((\neg\mathcal{A} \to \mathcal{B}) \to \mathcal{A}))$

(K4) $(\forall x\, \mathcal{A} \to \mathcal{A}\left(\frac{t}{x}\right))$, provided t is free for x in $\mathcal{A}$

(K5) $(\forall x\, (\mathcal{A} \to \mathcal{B}) \to (\mathcal{A} \to \forall x\, \mathcal{B}))$, provided x is not a free variable of $\mathcal{A}$

There are two rules of inference in $K_{\mathcal{L}}$.

Modus Ponens: from $\mathcal{A}$ and $(\mathcal{A} \to \mathcal{B})$ infer $\mathcal{B}$.

Generalisation: from $\mathcal{A}$ infer $\forall x\, \mathcal{A}$, for any variable x.

- The axiom schemes (K1)-(K3) and the rule Modus Ponens are as in the system L. However, keep in mind that the wff are now predicate forms rather than statement forms.
- The new axiom schemes (K4), (K5) and the new rule, generalization, tell us how to deal with the quantifiers.
- Note that from (K4) we have the axiom $(\forall x\, \mathcal{A} \to \mathcal{A}\left(\frac{x}{x}\right))$, since x is always free to substitute for x, which in fact is just $(\forall x\, \mathcal{A} \to \mathcal{A})$.

Let us consider some deriviations in $K_{\mathcal{L}}$.

Example 4.1. Let A be a unary predicate symbol in $\mathcal{L}$. Then

$$\vdash_{K_{\mathcal{L}}} (\forall x\, Ax \to \forall y\, Ay)$$

Proof. We have the following derivation:

1.	$(\forall x\, Ax \to Ay)$	K4
2.	$\forall y\, (\forall x\, Ax \to Ay)$	1, Gen
3.	$(\forall y\, (\forall x\, Ax \to Ay) \to (\forall x\, Ax \to \forall y\, Ay))$	K5
4.	$(\forall x\, Ax \to \forall y\, Ay)$	2, 3, MP

□

Example 4.2. Let A and B be unary predicate symbols in $\mathcal{L}$. Then

$$\{Ax\} \vdash_{K_{\mathcal{L}}} (By \to \forall x\, Ax)$$

Proof. We have the following derivation from $\{Ax\}$:

1.	Ax	Hyp
2.	$(Ax \to (By \to Ax))$	K1
3.	$(By \to Ax)$	1, 2, MP
4.	$\forall x\,(By \to Ax)$	3, Gen
5.	$(\forall x\,(By \to Ax) \to (By \to \forall x\,Ax))$	K5
6.	$(By \to \forall x\,Ax)$	4, 5, MP

□

Example. Let Man and $Mortal$ be unary predicate symbols in $\mathcal{L}$ and let s be a constant symbol in $\mathcal{L}$ (standing for Socrates). We show that Socrates is mortal by combining MP and the axiom scheme K4.

$$\{\forall x\,(Man\,x \to Mortal\,x), Man\,s\} \vdash_{K_\mathcal{L}} Mortal\,s$$

Proof. We have the following derivation:

1.	$\forall x\,(Man\,x \to Mortal\,x)$	Hyp
2.		K4 with s as term t
3.		1.2. MP
4.	$Man\,s$	Hyp
5.	$Mortal\,s$	3.4. MP

□

Definition 4.3. *A wff $\mathcal{A}$ of $\mathcal{L}$ is a* **tautology instance** *if there is some (statement form) tautology $\mathcal{B}$, involving statement variables $p_1, p_2, \ldots, p_n$, and some wffs $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ of $\mathcal{L}$ such that $\mathcal{A}$ is obtained from $\mathcal{B}$ by replacing each occurrence of p_i by $\mathcal{A}_i$.*

For example, $(p \to (q \to q))$ is a tautology, so replacing p with $\forall x\,Axy$ and q with $Bfxyz$ we get the tautology instance

$$(\forall x\,Axy \to (Bfxyz \to Bfxyz)).$$

Notice that the rules of satisfaction match the truth table rules for the connectives, and therefore every tautology instance is logically valid.

Notice also that all the axioms and the rule of inference of L apply in $K_\mathcal{L}$. We know that every tautology is a theorem of L. Therefore, every tautology instance is a theorem of $K_\mathcal{L}$. When we

give derivations in $K_{\mathcal{L}}$ we will allow ourselves the abbreviation of simply justifying a line as being a tautology instance, without actually finding a derivation strictly with the system.

Example 4.4. Show that $\{Axy\} \vdash_{K_{\mathcal{L}}} (\neg Ayy \to Bz)$.

Solution. We have the derivation

1.		Hyp
2.		1, Gen
3.		K4
4.		2, 3, MP
5.	$(Ayy \to (\neg Ayy \to Bz))$	instance of tautology
6.		4, 5, MP

□

4.2 The soundness theorem for $K_{\mathcal{L}}$

In this section we show that every theorem of $K_{\mathcal{L}}$ is logically valid.

Lemma 4.5. *Every axiom of $K_{\mathcal{L}}$ is logically valid.*

Proof. All axioms of type K1–K3 are tautology instances, hence logically valid.

We saw in the previous Chapter, Proposition 3.40, that if t is free for x in $\mathcal{A}$ then

$\forall x\, \mathcal{A} \Rightarrow \mathcal{A}\left(\frac{t}{x}\right)$. Thus $(\forall x \mathcal{A} \to \mathcal{A}\left(\frac{t}{x}\right))$ is logically valid.

Similarly, by the next to last rule in Theorem 3.42 we know that if x is not free in $\mathcal{A}$ then

$$\forall x(\mathcal{A} \to \mathcal{B}) \Rightarrow (\mathcal{A} \to \forall x\, \mathcal{B}),$$

so

$$(\forall x(\mathcal{A} \to \mathcal{B}) \to (\mathcal{A} \to \forall x\, \mathcal{B}))$$

is logically valid. □

Lemma 4.6. *Let $\mathcal{I}$ be an interpretation of a first-order language $\mathfrak{L}$, and let $\mathcal{A} \in \mathrm{form}(K_{\mathfrak{L}})$. Then for any variable x we have $\mathcal{I} \models \mathcal{A}$ iff $\mathcal{I} \models \forall x\, \mathcal{A}$.*

Proof. Let D be the domain of $\mathcal{I}$.

Suppose first that $\mathcal{I} \models \mathcal{A}$ [We must show that $\mathcal{I} \models \forall x\, \mathcal{A}$]. Let v be an $\mathcal{I}$-assignment [We must show that $\mathcal{I} \models_v \forall x\, \mathcal{A}$]. Let $d \in D$ [We must show that $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$]. Then $v\frac{d}{x}$ is an $\mathcal{I}$-assignment and $\mathcal{I} \models \mathcal{A}$, so $\mathcal{I} \models_{v\frac{d}{x}} \mathcal{A}$, as required.

Conversely, suppose that $\mathcal{I} \models \forall x\, \mathcal{A}$. This means that for each $\mathcal{I}$-assignment ν and each $e \in D$, $\mathcal{I} \models_{v\frac{e}{x}} \mathcal{A}$. In particular, this holds for $e = \nu(x)$, in which case $v\frac{e}{x} = v$. Since ν was arbitrary, we obtain $\mathcal{I} \models \mathcal{A}$, as required. □

Theorem 4.7 (The Soundness Theorem)**.** *Let $\mathcal{A}$ be a theorem of $K_{\mathfrak{L}}$. Then $\mathcal{A}$ is logically valid.*

Proof. Just as with the propositional calculus, we use induction. Let $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ be a derivation. We will prove by induction on i that $\mathcal{A}_i$ is logically valid for each i.

Base step: $\mathcal{A}_1$ must be an axiom, so it is logically valid by Lemma 4.5.

Inductive step: assume that $1 < i \leqslant n$, and that $\mathcal{A}_j$ is logically valid for each $1 \leqslant j < i$. There are three cases to consider: $\mathcal{A}_i$ is an axiom, $\mathcal{A}_i$ follows from two earlier wffs in the sequence by modus ponens, or $\mathcal{A}_i$ follows from an earlier wff by generalisation.

Case 1 If $\mathcal{A}_i$ is an axiom then, as above, $\mathcal{A}_i$ is logically valid.

Case 2 Suppose $\mathcal{A}_i$ follows by modus ponens from two earlier wffs. In other words, for some $j, k < i$ we have that $\mathcal{A}_k$ is $(\mathcal{A}_j \to \mathcal{A}_i)$. By inductive hypothesis, $\mathcal{A}_j$ and $(\mathcal{A}_j \to \mathcal{A}_i)$ are both logically valid. Thus, if $\mathcal{I}$ is any interpretation and v is any $\mathcal{I}$-assignment then $\mathcal{I} \models_v \mathcal{A}_j$ and $\mathcal{I} \models_v (\mathcal{A}_j \to \mathcal{A}_i)$, so $\mathcal{I} \models_v \mathcal{A}_i$. Thus $\mathcal{A}_i$ is also logically valid.

Case 3 Suppose $\mathcal{A}_i$ follows from an earlier wff by generalisation. In other words, for some $j < i$, $\mathcal{A}_i$ is $\forall x\, \mathcal{A}_j$. By inductive hypothesis, $\mathcal{A}_j$ is logically valid. Let $\mathcal{I}$ be an interpretation. Then $\mathcal{I} \models \mathcal{A}_j$, so, by Lemma 4.6, $\mathcal{I} \models \forall x\, \mathcal{A}_j$, i.e. $\mathcal{I} \models \mathcal{A}_i$. Thus $\mathcal{A}_i$ is logically valid.

□

Notice that what we have proved is more like the corollary to the Soundness Theorem for L than the Soundness Theorem itself: in L we showed that if $\Sigma \vdash_L \mathcal{A}$ then $\Sigma \models \mathcal{A}$, and deduced that if $\vdash_L \mathcal{A}$ then $\mathcal{A}$ is a tautology. To obtain an actual analog, we first have to clarify what we mean by entailment $\Sigma \models \mathcal{A}$ in predicate logic.

Definition 4.8. *Let $\Sigma \subseteq \mathrm{form}(K_{\mathfrak{L}})$ and let $\mathcal{A} \in \mathrm{form}(K_{\mathfrak{L}})$. We say that Σ* entails *$\mathcal{A}$, written $\Sigma \models \mathcal{A}$, if for every interpretation $\mathcal{I}$,*

$$\mathcal{I} \models \mathcal{B} \text{ for every } \mathcal{B} \in \Sigma \text{ implies } \mathcal{I} \models \mathcal{A}.$$

- It is a somewhat unfortunate tradition that the same symbol "$\models$" already used here for truth, $\mathcal{I} \models \mathcal{B}$, is also used for entailment. Keep in mind that entailment is a relationship between two syntactic objects: a set Σ of predicate forms and a further predicate form $\mathcal{A}$. On the other hand, when we defined truth $\mathcal{I} \models \mathcal{B}$, the symbol expressed a relationship between an interpretation (a semantic object) and a predicate form.

- Note that $\emptyset \models \mathcal{A}$ if and only if $\mathcal{A}$ is logically valid. Thus the following result, the complete analog of the soundness theorem for L, has Theorem 4.7 as a special case where $\Sigma = \emptyset$.

- * One could also suggest another notion of entailment: $\Sigma \models^* \mathcal{A}$ if for every $\mathcal{I}$ and for every $\mathcal{I}$-assignment v, if $\mathcal{I} \models_v \mathcal{B}$ for every $\beta \in \Sigma$, then $\mathcal{I} \models_v \mathcal{A}$. [Exercise*: Show that $\models^*$ implies $\models$, and give an example to show that the two notions are in fact different.]

Theorem 4.9. *Let* $\Sigma \subseteq \text{form}(K_{\mathcal{L}})$ *and let* $\mathcal{A} \in \text{form}(K_{\mathcal{L}})$. *If* $\Sigma \vdash_{K_{\mathcal{L}}} \mathcal{A}$ *then* $\Sigma \models \mathcal{A}$.

Proof. A suitable extension of the proof of the Soundness Theorem 4.7. See Exercise 2. □

We know that if $\mathcal{A} \Rightarrow \mathcal{B}$ then $\{\mathcal{A}\} \models \mathcal{B}$, but note that in general the converse fails. $\{\mathcal{A}\} \models \mathcal{B}$ does *not* imply that $\mathcal{A} \Rightarrow \mathcal{B}$, unless both $\mathcal{A}$ and $\mathcal{B}$ are closed. See Exercise 3*.

Exercises

1. If $\mathcal{A}$ is a wff of $K_{\mathcal{L}}$ and $x_1, x_2, \ldots, x_n$ are the free variables of $\mathcal{A}$, we call the wff $\forall x_1\, \forall x_2\, \ldots \forall x_n\, \mathcal{A}$ the *universal closure* of $\mathcal{A}$, denoted by $\overline{\mathcal{A}}$.
   (a) Show that if $\mathcal{I}$ is an interpretation then $\mathcal{I} \models \mathcal{A}$ iff $\mathcal{I} \models \overline{\mathcal{A}}$.
   (b) Show that $\vdash_{K_{\mathcal{L}}} \mathcal{A}$ iff $\vdash_{K_{\mathcal{L}}} \overline{\mathcal{A}}$.
2. Let $\Sigma \subseteq \text{form}(K_{\mathcal{L}})$ and $\mathcal{A} \in \text{form}(K_{\mathcal{L}})$. Suppose that $\Sigma \vdash_{K_{\mathcal{L}}} \mathcal{A}$. Show that $\Sigma \models \mathcal{A}$.

3*. Give an example in which $\{\mathcal{A}\} \models \mathcal{B}$, but $\mathcal{A} \not\Rightarrow \mathcal{B}$

4.3 The Deduction Theorem for $K_{\mathcal{L}}$

As with the formal system L, it is difficult to find derivations in $K_{\mathcal{L}}$. To help find derivations, we have a version of the Deduction Theorem. It would be nice if this were the same as for the formal system L, in other words "if $\Sigma \cup \{\mathcal{A}\} \vdash_{K_{\mathcal{L}}} \mathcal{B}$ then $\Sigma \vdash_{K_{\mathcal{L}}} (\mathcal{A} \to \mathcal{B})$." Unfortunately, this is not true. Let us give a counterexample where $\Sigma = \emptyset$, $\mathcal{A}$ is Ax and $\mathcal{B}$ is $\forall x\, Ax$.

By a single application of generalization, we see that $\{Ax\} \vdash_{K_\mathcal{L}} \forall x\, Ax$. However, $(Ax \to \forall x\, Ax)$ is not a theorem of $K_\mathcal{L}$.

To see this, consider the interpretation $\mathcal{I}$ with domain $\mathbb{N}$ in which $A^\mathcal{I}(n)$ holds iff n is even.

- If v is an $\mathcal{I}$-interpretation with $v(x) = 2$ then $\mathcal{I} \models_v Ax$ (since $v(x) = 2$ is even).
- However, we have $I \not\models_{v\frac{3}{x}} Ax$ (since $v\frac{3}{x}(x) = 3$ is not even), so $I \not\models_v \forall x\, Ax$.
- So for this particular v we have

$$I \not\models_v (Ax \to \forall x\, Ax).$$

- Thus $(Ax \to \forall x\, Ax)$ is not logically valid, so by the Soundness Theorem for $K_\mathcal{L}$, $(Ax \to \forall x\, Ax)$ cannot be a theorem of $K_\mathcal{L}$.

The problem is that we applied generalisation to the free variable x of our hypothesis Ax. If we avoid this then everything works out nicely.

Theorem 4.10 (The Deduction Theorem)**.** *Let $\Sigma \cup \{\mathcal{A}, \mathcal{B}\}$ be a set of wffs of a first-order language $\mathcal{L}$. Suppose that $\Sigma \cup \{\mathcal{A}\} \vdash_{K_\mathcal{L}} \mathcal{B}$, and the derivation of $\mathcal{B}$ from $\Sigma \cup \{\mathcal{A}\}$ does not use an application of generalisation to a free variable of $\mathcal{A}$. Then $\Sigma \vdash_{K_\mathcal{L}} (\mathcal{A} \to \mathcal{B})$.*

The Deduction Theorem is an important property of our formal system. The proof extends the one for the Deduction Theorem in the case of L. The axiom scheme K5 was designed to make the proof work in the case of $K_\mathcal{L}$. Similarly, the scheme K2 was designed in a way to make the proof of the Deduction Theorem for L work.

Proof. Let $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n$ be derivation of $\mathcal{B}$ from $\Sigma \cup \{\mathcal{A}\}$ which does not use an application of generalisation to a free variable of $\mathcal{A}$. We will prove by induction on i that

$$\Sigma \vdash_{K_\mathcal{L}} (\mathcal{A} \to \mathcal{B}_i) \text{ for each } i.$$

Base step: we know that either $\mathcal{B}_1$ is an axiom or $\mathcal{B}_1 \in \Sigma \cup \{\mathcal{A}\}$. If $\mathcal{B}_1$ is an axiom or $\mathcal{B}_1 \in \Sigma$ then we have $\Sigma \vdash_{K_\mathcal{L}} \mathcal{B}_1$, and $(\mathcal{B}_1 \to (\mathcal{A} \to \mathcal{B}_1))$ is an instance of K1, so $\Sigma \vdash_{K_\mathcal{L}} (\mathcal{A} \to \mathcal{B}_1)$.

If $\mathcal{B}_1 = \mathcal{A}$, then we use the fact that $(\mathcal{A} \to \mathcal{A})$ is a tautology instance, and hence provable using the axiom schemes K1–K3.

Inductive step: Suppose that $1 < i \leqslant n$ and that, for every j with $1 \leqslant j < i$, $\Sigma \vdash_{K_\mathcal{L}} (\mathcal{A} \to \mathcal{B}_j)$. There are 5 cases to consider:

- $\mathcal{B}_i$ is an axiom;

- $\mathcal{B}_i \in \Sigma$;
- $\mathcal{B}_i = \mathcal{A}$;
- $\mathcal{B}_i$ follows from two earlier wffs by modus ponens; or
- $\mathcal{B}_i$ follows from an earlier wff by generalisation (this is the only case where our proof differs from the proof of the Deduction Theorem for L).

Cases 1–3 if $\mathcal{B}_i$ is an axiom or $\mathcal{B}_i \in \Sigma \cup \{\mathcal{A}\}$ then, just as in the base step, we have $\Sigma \vdash_{K_\mathcal{L}} (\mathcal{A} \to \mathcal{B}_i)$.

Case 4 Suppose that $\mathcal{B}_i$ follows from two earlier terms by modus ponens. For this to happen, the two earlier terms must have the form $\mathcal{C}$ and $(\mathcal{C} \to \mathcal{B}_i)$. In other words, there are $1 \leqslant k, j < i$ such that $\mathcal{B}_j$ is $(\mathcal{B}_k \to \mathcal{B}_i)$.

By inductive hypothesis,

$$\Sigma \vdash_{K_\mathcal{L}} (\mathcal{A} \to \mathcal{B}_k)$$

and

$$\Sigma \vdash_{K_\mathcal{L}} (\mathcal{A} \to (\mathcal{B}_k \to \mathcal{B}_i)).$$

But we also have

$$\Sigma \vdash_{K_\mathcal{L}} ((\mathcal{A} \to (\mathcal{B}_k \to \mathcal{B}_i)) \to ((\mathcal{A} \to \mathcal{B}_k) \to (\mathcal{A} \to \mathcal{B}_i))),$$

since the latter is an axiom, so by modus ponens we have

$$\Sigma \vdash_{K_\mathcal{L}} ((\mathcal{A} \to \mathcal{B}_k) \to (\mathcal{A} \to \mathcal{B}_i)),$$

and by modus ponens again we have $\Sigma \vdash_{K_\mathcal{L}} (\mathcal{A} \to \mathcal{B}_i)$.

Case 5 Suppose $\mathcal{B}_i$ follows by generalisation from an earlier wff. In other words, we must have that $\mathcal{B}_i$ is $\forall x\, \mathcal{B}_j$ for some $j < i$ and some variable x.

Since we assumed that the derivation did not use generalisation applied to a free variable of $\mathcal{A}$, x is not a free variable of $\mathcal{A}$.

By inductive hypothesis, $\Sigma \vdash_{K_\mathcal{L}} (\mathcal{A} \to \mathcal{B}_j)$, so by generalisation $\Sigma \vdash_{K_\mathcal{L}} \forall x\, (\mathcal{A} \to \mathcal{B}_j)$. Since x is not free in $\mathcal{A}$,

$$(\forall x\, (\mathcal{A} \to \mathcal{B}_j) \to (\mathcal{A} \to \forall x\, \mathcal{B}_j))$$

is an instance of K5, so by modus ponens we have

$$\Sigma \vdash_{K_\mathcal{L}} (\mathcal{A} \to \forall x\, \mathcal{B}_j),$$

i.e. $\Sigma \vdash_{K_\mathcal{L}} (\mathcal{A} \to \mathcal{B}_i)$.

This completes the induction. □

Notice that in the above proof, we can conclude more than just that $\Sigma \vdash_{K_\mathcal{L}} (\mathcal{A} \to \mathcal{B})$. We also know what variables have generalisation applied to them:

- if we apply generalisation to x in the new derivation showing that $\Sigma \vdash_{K_{\mathcal{L}}} (\mathcal{A} \to \mathcal{B})$,
- then we had already applied generalisation to x in the original derivation of $\Sigma \cup \{\mathcal{A}\} \vdash_{K_{\mathcal{L}}} \mathcal{B}$.

This will be useful in some of the following examples, where we will want to apply the Deduction Theorem twice or more.

Example 4.11. Let $\mathcal{A}$ and $\mathcal{B}$ be wffs of a first-order language $\mathcal{L}$. Then

$$\vdash_{K_{\mathcal{L}}} (\forall x\,(\mathcal{A} \to \mathcal{B}) \to (\forall x\,\mathcal{A} \to \forall x\,\mathcal{B}))$$

Proof. We will first show that

$$\{\forall x\,(\mathcal{A} \to \mathcal{B}), \forall x\,\mathcal{A}\} \vdash_{K_{\mathcal{L}}} \forall x\,\mathcal{B}.$$

1.		Hyp
2.	$(\forall x\,(\mathcal{A} \to \mathcal{B}) \to (\mathcal{A} \to \mathcal{B}))$	K4
3.		1, 2, MP
4.		Hyp
5.		K4
6.		4, 5, MP
7.		3, 6, MP
8.		7, Gen

Notice that the only application of generalisation was to the variable x, which is not free in $\forall x\,\mathcal{A}$. So, by the Deduction Theorem, $\{\forall x\,(\mathcal{A} \to \mathcal{B})\} \vdash_{K_{\mathcal{L}}} (\forall x\mathcal{A} \to \forall x\,\mathcal{B})$, and the derivation only uses generalisation applied to x, which is not free in $\forall x\,(\mathcal{A} \to \mathcal{B})$. Hence, by the Deduction Theorem again,

$$\vdash_{K_{\mathcal{L}}} (\forall x\,(\mathcal{A} \to \mathcal{B}) \to (\forall x\,\mathcal{A} \to \forall x\mathcal{B})).$$

□

Although we have eliminated $\wedge$, $\vee$, $\leftrightarrow$ and $\exists$ from our language, it is still useful to use these symbols as abbreviations for wffs. We regard $\exists x\,\mathcal{A}$ as being an abbreviation for $\neg\forall x\,\neg\mathcal{A}$, and similarly we regard $(\mathcal{A} \wedge \mathcal{B})$ as an abbreviation for $\neg(\mathcal{A} \to \neg\mathcal{B})$ and so on for the other symbols.

Example 4.12. Let $\mathcal{A}$ and $\mathcal{B}$ be wffs, and let x be a variable which does not occur free in $\mathcal{B}$. Then

$$\vdash_{K_{\mathcal{L}}} (\forall x\,(\mathcal{A} \to \mathcal{B}) \to (\exists x\,\mathcal{A} \to \mathcal{B}))$$

Proof. We first prove that

$$\{\forall x\,(\mathcal{A} \to \mathcal{B}), \neg\mathcal{B}\} \vdash_{K_{\mathcal{L}}} \forall x\,\neg\mathcal{A}$$

We have the following derivation of $\forall x\neg\mathcal{A}$ from $\{\forall x(\mathcal{A} \to \mathcal{B}), \neg\mathcal{B}\}$:

1.		Hyp
2.	$(\forall x\,(\mathcal{A} \to \mathcal{B}) \to (\mathcal{A} \to \mathcal{B}))$	K4
3.		1, 2, MP
4.		taut
5.		3, 4, MP
6.		Hyp
7.		5, 6, MP
8.	$\forall x\,\neg\mathcal{A}$	7, Gen

The only instance of generalisation in the above derivation was to x, which is not free in $\mathcal{B}$, so by the Deduction Theorem we have

$$\{\forall x\,(\mathcal{A} \to \mathcal{B})\} \vdash_{K_{\mathcal{L}}} (\neg\mathcal{B} \to \forall x\,\neg\mathcal{A}).$$

Now the proposition $((\neg p \to q) \to (\neg q \to p))$ is a tautology, so the wff

$$((\neg\mathcal{B} \to \forall x\,\neg\mathcal{A}) \to (\neg\forall x\,\neg\mathcal{A} \to \mathcal{B}))$$

is a tautology instance and hence provable in $K_{\mathcal{L}}$ (without using generalisation). Hence by Modus Ponens we have

$$\{\forall x\,(\mathcal{A} \to \mathcal{B})\} \vdash_{K_{\mathcal{L}}} (\neg\forall x\,\neg\mathcal{A} \to \mathcal{B}),$$

in other words

$$\{\forall x\,(\mathcal{A} \to \mathcal{B})\} \vdash_{K_{\mathcal{L}}} (\exists x\,\mathcal{A} \to \mathcal{B}),$$

and the only application of generalisation is to x which is not free in $\forall x\,(\mathcal{A} \to \mathcal{B})$. So, by the Deduction Theorem again, we have

$$\vdash_{K_{\mathcal{L}}} (\forall x\,(\mathcal{A} \to \mathcal{B}) \to (\exists x\,\mathcal{A} \to \mathcal{B})).$$

□

Exercises

3. Let $\mathcal{A}$ be a wff of $K_{\mathcal{L}}$.

   (a) Complete the following derivation showing that $\{\forall x \forall y\,\mathcal{A}\} \vdash_{K_{\mathcal{L}}} \forall y\,\forall x\,\mathcal{A}$:

1.	$\forall x\,\forall y\,\mathcal{A}$	Hyp
2.		K4
3.		2, 1, MP
4.		K4
5.		3, 4, MP
6.		5, Gen
7.	$\forall y\,\forall x\,\mathcal{A}$	6, Gen

   (b) Deduce that $\vdash_{K_{\mathcal{L}}} (\forall x\,\forall y\,\mathcal{A} \to \forall y\,\forall x\mathcal{A})$.

4. Let $\mathcal{A}$ and $\mathcal{B}$ be wffs of $K_{\mathcal{L}}$. Suppose x does not occur free in $\mathcal{B}$. We will show that $\vdash_{K_{\mathcal{L}}} ((\exists x\mathcal{A} \to \mathcal{B}) \to \forall x\,(\mathcal{A} \to \mathcal{B}))$. Recall that $(\exists x\mathcal{A} \to \mathcal{B})$ is an abbreviation for $(\neg\forall x\neg\mathcal{A} \to \mathcal{B})$.

   (a) Complete the following derivation showing that $\{\exists x\,\mathcal{A} \to \mathcal{B}), \neg\mathcal{B}\} \vdash_{K_{\mathcal{L}}} \neg\mathcal{A}$:

1.	$(\neg\forall x\,\neg\mathcal{A} \to \mathcal{B})$	Hyp
2.	$((\neg\forall x\,\neg\mathcal{A} \to \mathcal{B}) \to (\neg\mathcal{B} \to \forall x\neg\mathcal{A}))$	instance of tautology $((\neg p \to q) \to (\neg q \to p))$
3.		1, 2, MP
4.		Hyp
5.		4, 3, MP
6.		K4
7.	$\neg\mathcal{A}$	5, 6, MP

   (b) Deduce that $\{(\exists x\,\mathcal{A} \to \mathcal{B})\} \vdash_{K_{\mathcal{L}}} (\neg\mathcal{B} \to \neg\mathcal{A})$ and the derivation does not use generalisation.

   (c) Deduce that $\{(\exists x\,\mathcal{A} \to \mathcal{B})\} \vdash_{K_{\mathcal{L}}} \forall x\,(\mathcal{A} \to \mathcal{B})$ and the derivation only applies generalisation to x.

   (d) Deduce that $\vdash_{K_{\mathcal{L}}} ((\exists x\,\mathcal{A} \to \mathcal{B}) \to \forall x(\mathcal{A} \to \mathcal{B}))$.

5. Let A be a unary predicate symbol.

   (a) Complete the following derivation showing that $\{\forall x\,Ax\} \vdash_{K_{\mathcal{L}}} \forall y\,Ay$:

1.	$\forall x\,Ax$	Hyp
2.		K4 (NB y is free for x in $A(x)$)
3.		1, 2, MP
4.	$\forall y\,Ay$	3, Gen

   (b) Deduce that $\vdash_{K_{\mathcal{L}}} (\forall x\,Ax \to \forall y\,Ay)$.

   (c) Use the proof of the Deduction Theorem to convert the derivation in (a) into a derivation showing that $\vdash_{K_{\mathcal{L}}} (\forall x\,Ax \to \forall y\,Ay)$.

4.4 Models and consistency

We have seen that hypotheses with free variables can cause problems. For instance, the statement of the Deduction Theorem is complicated by the extra restriction that the derivation does not apply generalisation to a free variable of $\mathcal{A}$.

We can avoid this difficulty by looking at sets of *closed* wffs (wffs with no free variables). Recall that if $\mathcal{A}$ is a closed wff then for every $\mathcal{I}$ we have either $\mathcal{I} \models \mathcal{A}$ or $\mathcal{I} \models \neg\mathcal{A}$.

Definition 4.13. *Let Σ be a set of closed wffs. A* **model of** Σ *is an interpretation $\mathcal{I}$ such that $\mathcal{I} \models \mathcal{B}$ for all $\mathcal{B} \in \Sigma$.*

Example 4.14.

1. Recall Exercise 2 from Chapter 3. Let Σ be the set of predicate forms (LO1)-(LO4). Then the interpretation consisting of the rational numbers with the $\leqslant$ relation is a model of Σ.

2. Consider the language $\mathcal{L}$ where we only have the binary relation symbol E. For each n let $\mathcal{A}_n$ be the closed predicate form of $\mathcal{L}$: E is an equivalence relation, and there are at least n nonequivalent elements. Let $\Sigma = \{\mathcal{A}_n : n \in \mathbb{N}\}$. What are the models of Σ?

Definition 4.15. *Let Σ be a set of closed wffs. Then Σ is* **inconsistent** *if there is some wff $\mathcal{A}$ with $\Sigma \vdash_{K_\mathcal{L}} \mathcal{A}$ and $\Sigma \vdash_{K_\mathcal{L}} \neg\mathcal{A}$. Otherwise it is* **consistent**.

Proposition 4.16. *If Σ is a set of closed wffs and Σ has a model then Σ is consistent.*

Proof. Let $\mathcal{I}$ be a model of Σ. We know that if $\Sigma \vdash_{K_\mathcal{L}} \mathcal{A}$ then $\Sigma \models \mathcal{A}$.

Suppose Σ is inconsistent. Then there is some $\mathcal{A}$ such that

$$\Sigma \vdash_{K_\mathcal{L}} \mathcal{A} \text{ and } \Sigma \vdash_{K_\mathcal{L}} \neg\mathcal{A}.$$

But then $\Sigma \models \mathcal{A}$ and $\Sigma \models \neg\mathcal{A}$. Since $\mathcal{I}$ is a model of Σ we have $\mathcal{I} \models \mathcal{A}$ and $\mathcal{I} \models \neg\mathcal{A}$, which is impossible. □

Proposition 4.17. *Let Σ be a set of closed wffs and let $\mathcal{A}$ be a closed wff. If $\Sigma \nvdash_{K_\mathcal{L}} \mathcal{A}$ then $\Sigma \cup \{\neg\mathcal{A}\}$ is consistent.*

Proof. Suppose not, i.e. suppose that $\Sigma \cup \{\neg\mathcal{A}\} \vdash_{K_\mathcal{L}} \mathcal{B}$ and $\Sigma \cup \{\neg\mathcal{A}\} \vdash_{K_\mathcal{L}} \neg\mathcal{B}$ for some wff $\mathcal{B}$. Note that $\neg\mathcal{A}$ has no free variables, so by the Deduction Theorem we have

$$\Sigma \vdash_{K_\mathcal{L}} (\neg\mathcal{A} \to \neg\mathcal{B})$$

and $\Sigma \vdash_{K_\mathcal{L}} (\neg\mathcal{A} \to \mathcal{B})$. We also have

$$\Sigma \vdash_{K_\mathcal{L}} ((\neg\mathcal{A} \to \neg\mathcal{B}) \to ((\neg\mathcal{A} \to \mathcal{B}) \to \mathcal{A})),$$

so by modus ponens (twice) we have $\Sigma \vdash_{K_\mathcal{L}} \mathcal{A}$, a contradiction. □

Definition 4.15 says that Σ is inconsistent if there is some wff $\mathcal{A}$ with $\Sigma \vdash_{K_\mathcal{L}} \mathcal{A}$ and $\Sigma \vdash_{K_\mathcal{L}} \neg\mathcal{A}$. In fact if Σ is inconsistent then for every wff $\mathcal{A}$, $\Sigma \vdash_{K_\mathcal{L}} \mathcal{A}$ and $\Sigma \vdash_{K_\mathcal{L}} \neg\mathcal{A}$. Hence in Proposition 4.17, the hypothesis $\Sigma \nvdash_{K_\mathcal{L}} \mathcal{A}$ implies that Σ is consistent.

The major result needed to prove the adequacy of our system $K_\mathcal{L}$ is the following:

Theorem 4.18. *If Σ is a consistent set of closed wffs then Σ has a model.*

The proof of this theorem requires the construction of a model from the consistent set of closed wffs. The details of the construction are too complex for this course, but can be found in *Logic for Mathematicians*, pp92–99.

Here is some idea of the proof of Theorem 4.18. (Some serious details are omitted.) We may assume that $\mathcal{L}$ contains a binary relation E interpreted as equality, else we add it. We also extend Σ by adding axioms for equality if necessary.

Now we extend Σ twice. The first extension requires enlargement of the language, too.

Step 1: extend Σ to $\Sigma_1 \subseteq \mathcal{L}_1$ containing "examples": for each wff in $\mathcal{L}_1$ of the form $\exists x\mathcal{A}$, Σ_1 contains the implication $(\exists x\mathcal{A} \to \mathcal{A}\frac{c}{x})$, for some new constant symbol c.

Step 2: Extend Σ_1 to $\Sigma_2 \subseteq \mathcal{L}_1$ which is maximally consistent.

Now we build a model $\mathcal{I}$ of Σ_2. Let T be the set of terms in $\mathcal{L}_1$, and let $\sim$ be the equivalence relation on T given by $t \sim s$ iff $Ets \in \Sigma_2$. The domain of the model $\mathcal{I}$ is the set of equivalence classes $T/\sim$. For each $t \in T$, let $[t]$ be the equivalence class. Complete the definition of $\mathcal{I}$ by defining, say, for a binary relation symbol R and a unary function symbol f,

$$R^{\mathcal{I}}[t][s] \text{ iff } Rts \in \Sigma_2$$

$$f^{\mathcal{I}}([t]) = [s] \text{ iff} Efts \in \Sigma_2.$$

Now one can check that $\mathcal{I} \models \Sigma_2$.

Since we wish to deal with closed wffs, it is useful to be able to convert a wff to a closed wff. If $\mathcal{A}$ is a wff of $K_{\mathcal{L}}$ and $x_1, x_2, \ldots, x_n$ are the free variables of $\mathcal{A}$, we call the wff $\forall x_1 \forall x_2 \ldots \forall x_n \mathcal{A}$ the *universal closure* of $\mathcal{A}$, denoted by $\overline{\mathcal{A}}$.

We know that for any model $\mathcal{I}$ we have $\mathcal{I} \models \mathcal{A}$ iff $\mathcal{I} \models \forall x\, \mathcal{A}$. We can also easily show that $\Sigma \vdash_{K_{\mathcal{L}}} \mathcal{A}$ iff $\Sigma \vdash_{K_{\mathcal{L}}} \forall x\, \mathcal{A}$. By induction on the number of free variables we can easily show that this implies that

- $\mathcal{I} \models \mathcal{A}$ iff $\mathcal{I} \models \overline{\mathcal{A}}$, and
- $\Sigma \vdash_{K_{\mathcal{L}}} \mathcal{A}$ iff $\Sigma \vdash_{K_{\mathcal{L}}} \overline{\mathcal{A}}$.

[See Exercise 1 for more details.]

Proposition 4.19. *Let Σ be a set of closed wffs of $K_{\mathcal{L}}$ and let $\mathcal{A}$ be a wff of $K_{\mathcal{L}}$. If $\Sigma \models \mathcal{A}$ then $\Sigma \vdash_{K_{\mathcal{L}}} \mathcal{A}$.*

Proof. We prove the contrapositive.

- Suppose $\Sigma \nvdash_{K_{\mathcal{L}}} \mathcal{A}$.
- Then $\Sigma \nvdash_{K_{\mathcal{L}}} \overline{\mathcal{A}}$,
- so $\Sigma \cup \{\neg\overline{\mathcal{A}}\}$ is consistent,
- so by Theorem 4.18 it has a model.

Let $\mathcal{I}$ be such a model of $\Sigma \cup \{\neg\overline{\mathcal{A}}\}$. Then $\mathcal{I} \models \mathcal{B}$ for all $\mathcal{B} \in \Sigma$ but $I \not\models \overline{\mathcal{A}}$, so $I \not\models \mathcal{A}$, so $\Sigma \not\models \mathcal{A}$, as required. □

Theorem 4.20 (The Adequacy Theorem for $K_{\mathcal{L}}$)**.** *If $\mathcal{A}$ is logically valid then $\vdash_{K_{\mathcal{L}}} \mathcal{A}$.*

Proof. If $\mathcal{A}$ is logically valid then $\varnothing \models \mathcal{A}$, so $\varnothing \vdash_{K_{\mathcal{L}}} \mathcal{A}$ (by Proposition 4.19), i.e. $\vdash_{K_{\mathcal{L}}} \mathcal{A}$. □

Exercises

6. (a) Let Σ be a set of closed wffs and let $\mathcal{A}$ and $\mathcal{B}$ be closed wffs. Recall that $(\mathcal{A} \vee \mathcal{B})$ is an abbreviation for $(\neg\mathcal{A} \to \mathcal{B})$. Suppose that $\Sigma \cup \{(\mathcal{A} \vee \mathcal{B})\}$ is consistent. Show that either $\Sigma \cup \{\mathcal{A}\}$ is consistent or $\Sigma \cup \{\mathcal{B}\}$ is consistent.

   (b) Recall that $(\mathcal{A} \wedge \mathcal{B})$ is an abbreviation for $\neg(\mathcal{A} \to \neg\mathcal{B})$. Give an example of set set Σ of closed wffs and closed wffs $\mathcal{A}$ and $\mathcal{B}$ such that $\Sigma \cup \{\mathcal{A}\}$ is consistent and $\Sigma \cup \{\mathcal{B}\}$ is consistent but $\Sigma \cup \{(\mathcal{A} \wedge \mathcal{B})\}$ is inconsistent.

7. (a) Let Σ be a set of closed wffs and let $\mathcal{A}$ be a closed wff. Show that if Σ is consistent then either $\Sigma \cup \{\mathcal{A}\}$ is consistent or $\Sigma \cup \{\neg\mathcal{A}\}$ is consistent.

   (b) Give an example of set set Σ of closed wffs and closed wffs $\mathcal{A}$ and $\mathcal{B}$ such that $\Sigma \cup \{\mathcal{A}\}$ is consistent and $\Sigma \cup \{(\mathcal{A} \to \mathcal{B})\}$ is consistent but $\Sigma \cup \{\mathcal{B}\}$ is inconsistent. [You should give brief reasons why your example works but you need not give complete details.]

4.5 Compactness

Proposition 4.21. *Let Σ be a set of closed wffs. If every finite subset of Σ is consistent then Σ is consistent.*

Proof. We prove the contrapositive: if Σ is inconsistent then there is a finite subset of Σ that is inconsistent. So suppose that Σ is inconsistent.

- Then there is some $\mathcal{A}$ such that $\Sigma \vdash_{K_{\mathcal{L}}} \mathcal{A}$ and $\Sigma \vdash_{K_{\mathcal{L}}} \neg\mathcal{A}$.
- Let $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m$ be a derivation of $\mathcal{A}$ from Σ and let $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n$ be a derivation of $\neg\mathcal{A}$ from Σ.
- Put
$$\Gamma = \Sigma \cap \big(\{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m\} \cup \{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n\}\big).$$
- Then Γ is a finite subset of Σ.

Consider the sequence $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m$. Every term in this sequence is an axiom, in Σ, follows from two earlier terms by modus ponens or follows from an earlier term by generalisation. But any of the terms which are in Σ are in Γ, so every term in the sequence is an axiom, is in Γ, follows from two earlier terms by modus ponens or follows from an earlier term by generalisation. In other words, this is a derivation of $\mathcal{A}$ from Γ. Similarly, $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n$ is a derivation of $\neg\mathcal{A}$ from Γ. So Γ is a finite inconsistent subset of Σ, as required. □

Theorem 4.22 (The Compactness Theorem for $K_{\mathcal{L}}$)**.** *Let Σ be a set of closed wffs of $K_{\mathcal{L}}$. If every finite subset of Σ has a model then Σ has a model.*

Proof. Suppose every finite subset of Σ has a model. Then every finite subset is consistent, so Σ is consistent, so Σ has a model. □

Proposition 4.23. *Let Σ be a set of closed wffs. Suppose that, for every $n \in \mathbb{N}$, Σ has a finite model with at least n elements. Then Σ also has an infinite model.*

As a consequence, we cannot axiomatize finiteness in first-order logic: fix any first-order language $\mathcal{L}$. There is no set of closed wff Σ in $\mathcal{L}$ such that for each interpretation $\mathcal{I}$, we have $\mathcal{I} \models \Sigma$ iff the domain of $\mathcal{I}$ is finite.

Proof. We will start with a rough outline of the idea and then give the details. The idea is that we will add a collection of closed wffs to Σ to get a new set Π. The new wffs are chosen in such a way as to ensure that any model of Π must be infinite. To show that Π has a model, it is enough to show that any finite subset of Π has a model. If Γ is a finite subset of Π then it will only include finitely many of the new wffs, so we can use a sufficiently large finite model of Σ to give us a model of Γ.

The tricky bit is to add new closed wffs which do not "interfere" with the existing ones in Σ. What we will do is introduce some new symbols which were not in the language $\mathcal{L}$ and use them for the new wffs.

Let E be a new binary predicate symbol not in $\mathcal{L}$ and let $\{c_n \mid n \in \mathbb{N}\}$ be new constant symbols not in $\mathcal{L}$. Let $\widetilde{\mathcal{L}}$ be the language we get by adding these new symbols to $\mathcal{L}$. Let

$$\Pi = \Sigma \cup \{\forall x\, Exx\} \cup \{\, \neg Ec_m c_n \mid m, n \in \mathbb{N}, m < n \,\}.$$

We claim that Π has a model, which must therefore be an infinite model of Σ. To show that Π has a model it is enough to show that every finite subset of Π has a model.

Let Γ be a finite subset of Π. Then only finitely many of the new constants c_n appear in the wffs in Γ: choose N large enough so that if c_n appears in a wff in Γ then $n \leqslant N$. We assumed that Σ has a model $\mathcal{I}$ with at least N elements, so we can choose distinct elements $d_1, d_2, \ldots, d_N$ from the domain of $\mathcal{I}$. We need to extend $\mathcal{I}$ to an $\widetilde{\mathcal{L}}$-structure $\widetilde{\mathcal{I}}$ by adding interpretations $E^{\widetilde{\mathcal{I}}}$ and $c_m^{\widetilde{\mathcal{I}}}$.

- We let $E^{\widetilde{\mathcal{I}}}$ be the relation of equality: $E^{\widetilde{\mathcal{I}}}(d, d')$ holds iff $d = d'$.
- We declare that $c_m^{\widetilde{\mathcal{I}}} = d_m$ for $m \leqslant N$, and $c_m^{\widetilde{\mathcal{I}}} = d_1$ for $m > N$.

We must check that $\widetilde{\mathcal{I}} \models \mathcal{B}$ for all $\mathcal{B} \in \Gamma$.

- Since equality is reflexive we certainly have $\widetilde{\mathcal{I}} \models \forall x Exx$.
- And if $\neg Ec_m c_n$ is in Γ then $m, n \leqslant N$ so $c_m^{\widetilde{\mathcal{I}}} = d_m$, $c_n^{\widetilde{\mathcal{I}}} = d_n$, and $d_m \neq d_n$
- so $\widetilde{\mathcal{I}} \models \neg Ec_m c_n$.

Thus $\widetilde{\mathcal{I}} \models \mathcal{B}$ for all $\mathcal{B} \in \Gamma$, as required.

Now we know that every finite subset of Π has a model, so by the compactness theorem we know that Π has model, $\mathcal{J}$ say. We have $\mathcal{J} \models \mathcal{B}$ for all $\mathcal{B} \in \Pi$, so in particular $\mathcal{J} \models \mathcal{B}$ for all $\mathcal{B} \in \Sigma$, so $\mathcal{J}$ is a model of Σ. We claim that it must be infinite. To see this, consider the interpretations $c_n^{\mathcal{J}}$ of the constants c_n for $n \in \mathbb{N}$. These must all be distinct. This is because $\mathcal{J} \models \neg Ec_m c_n$ for each $m < n$, so $E^{\mathcal{J}}(c_m^{\mathcal{J}}, c_n^{\mathcal{J}})$ does not hold. Note that we do not know that $E^{\mathcal{J}}$ is the relation of equality in $\mathcal{J}$. However, we do know that $E^{\mathcal{J}}(d, d)$ holds for all d in the domain of $\mathcal{J}$, since $\mathcal{J} \models \forall x\, Exx$. Thus if $E^{\mathcal{J}}(c_m^{\mathcal{J}}, c_n^{\mathcal{J}})$ does not hold then $c_m^{\mathcal{J}} \neq c_n^{\mathcal{J}}$, as required. □

Exercises

8. Let Σ be the set of all closed wffs of $\mathcal{L}_N$ which are true in $\mathcal{N}$. For each $n \in \mathbb{N}$, let $\ddot{n}$ denote the term $\underbrace{hh\cdots h}_{n} c$, so we have $\ddot{0} = c$, $\ddot{1} = hc$, $\ddot{2} = hhc$, $\ddot{5} = hhhhhc$ and so on.

   Let $\mathcal{L}$ be the language obtained from $\mathcal{L}_N$ by adding a new constant symbol a. Let $\Pi = \Sigma \cup \{\, B\ddot{n}a \mid n \in \mathbb{N} \,\}$. Show that Π has a model.

Chapter 5

First Order Systems

What we have done so far is to set up the framework for mathematical reasoning. $K_{\mathcal{L}}$ provides the logical procedures that mathematicians use. What we will look at now is systems that actually describe mathematical objects. We obtain these systems by adding extra axioms to $K_{\mathcal{L}}$. Then any such object will be a model of the system, and we can prove theorems about the object by working in $K_{\mathcal{L}}$.

Definition 5.1. *A* first-order system *is a formal system S obtained from $K_{\mathcal{L}}$ for some first-order language $\mathcal{L}$ by adding extra axioms. The axioms of $K_{\mathcal{L}}$ are referred to as the* logical *axioms, and the extra axioms as the* proper *axioms of the system. We refer to $\mathcal{L}$ as the* language of S*, sometimes written $\mathcal{L}(S)$.*

Definition 5.2. *Suppose S is a first-order system, then* a model of S *is an interpretation of $\mathcal{L}(S)$ such that every theorem of S is true in the interpretation.*

Note Definition 4.13 for a model of Σ. If we call the collection of axioms of S, Σ, then in fact the two definitions are equivalent. Exercise*: prove this.

5.1 First order systems with equality

A fundamental relation in mathematics is the relation of equality. We would like to extend $K_{\mathcal{L}}$ to include equality. The $=$ symbol does not appear in our first-order languages. We have to treat equality like any other binary relation, and assign it to one of the binary predicate symbols. We will make the convention that the predicate symbol E will always be used to denote this relation.

We introduce the following axioms:

$$Exx \tag{E1}$$

$$(Et_k u \to Eft_1 \dots t_k, \dots t_n, ft_1 \dots u \dots t_n) \tag{E2}$$

where f is an n-ary function symbol and $t_1, \ldots, t_n$ and u are terms.

$$(Et_k u \to (At_1 \ldots t_k \ldots t_n \to At_1 \ldots u \ldots t_n)) \tag{E3}$$

where A is an n-ary predicate symbol and $t_1, \ldots, t_n$ and u are terms.

So, for example, if A is a binary predicate symbol and f is a unary function symbol, then

$$(Eyz \to (Axy \to Axz)),$$

and

$$(Exy \to Efxfy).$$

A *first-order system with equality* is a first-order system which has all instances of E1–E3 among its proper axioms.

Lemma 5.3. *Let S be a first-order system with equality. Then*

$$\vdash_S \forall x\, Exx$$
$$\vdash_S \forall x \forall y\, (Exy \to Eyx)$$
$$\vdash_S \forall x \forall y \forall z\, (Exy \to (Eyz \to Exz))$$

Proof. Exercise (see exercise 1 below). □

From this it follows that in any model, $\mathcal{M}$, of S, $E^{\mathcal{M}}$ is an equivalence relation. However, these axioms E1–E3 do not force $E^{\mathcal{M}}$ to be the actual relation of equality. [In fact it is impossible to find a collection of axioms for which the only interpretation is equality. Using E1-E3 comes close, but it is always possible to find a model in which E may be interpreted as as an equivalence relation that is not equality.]

Recall the language $\mathcal{L}_G$ of group theory, in other words the language $(\{c\}, \{E^2\}, \{f^2, g^1\})$, where c is a constant symbol (to represent the identity element), E is a binary predicate symbol (to represent equality), f is a binary function symbol (to represent the group multiplication) and g is a unary function symbol (to represent the group inverse function). Let $\mathcal{I}$ be the following interpretation:

- The domain of $\mathcal{I}$ is the set $\mathbb{Z}$ of all integers.
- The interpretation of c is the integer 0.
- The interpretation of E is the relation of congruence modulo 3, in other words $E^{\mathcal{I}}(m, n)$ holds if and only if $m \equiv n \pmod 3$.
- The interpretation of f is addition.
- The interpretation of g is the inverse function $n \mapsto -n$.

The interpretation $\mathcal{I}$ satisfies E1-E3 if and only if the following hold:

$$\begin{array}{lllll}
(x \equiv x & (\text{mod } 3)) & & & \\
(x \equiv y & (\text{mod } 3) & \to & -x \equiv -y & (\text{mod } 3)) \\
(x \equiv y & (\text{mod } 3) & \to & x + z \equiv y + z & (\text{mod } 3)) \\
(x \equiv y & (\text{mod } 3) & \to & z + x \equiv z + y & (\text{mod } 3)) \\
(x \equiv y & (\text{mod } 3) & \to & (x \equiv z \quad (\text{mod } 3) \to y \equiv z & (\text{mod } 3))) \\
(x \equiv y & (\text{mod } 3) & \to & (z \equiv x \quad (\text{mod } 3) \to z \equiv y & (\text{mod } 3)))
\end{array}$$

Since they each hold this structure is a model of E1–E3. However, in this interpretation $E^{\mathcal{I}}$ is not the relation of equality.

We can get equality if we take a *quotient* structure. We know that the relation $E^{\mathcal{I}}$, the relation of congruence mod 3, is an equivalence relation so we may consider the set $\mathbb{Z}_3$ of equivalence classes. For $n \in \mathbb{Z}$ let $[n]$ denote the equivalence class mod 3 of n, i.e.

$$\begin{aligned}
[n] &= \{\, m \in \mathbb{Z} \mid m \equiv n \pmod 3 \,\} \\
&= \{n + 3k \mid k \in \mathbb{Z}\,\}.
\end{aligned}$$

This gives us the domain of the quotient structure. Now what about the interpretations of the symbols? We will use $\ddot{s}$ to denote the interpretation of the symbol s in the new structure. We can take $\ddot{c} = [c^{\mathcal{I}}] = [0]$. What about $\ddot{f}$?

To define $\ddot{f}$, we have to say what $\ddot{f}(d, e)$ is for every d, e in the domain, in other words we have to define $\ddot{f}([m], [n])$ for $m, n \in \mathbb{Z}$. We declare that $\ddot{f}([m], [n]) = [m + n]$ for all m and n. But there is a potential problem with this definition. For example, we have $2 \equiv 5 \pmod 3$ and $6 \equiv 9 \pmod 3$, so $[2] = [5]$ and $[6] = [9]$. So if $d = [2] = [5]$ and $e = [6] = [9]$ we have different possible values for $\ddot{f}(d, e)$: do we have $\ddot{f}(d, e) = [2 + 6] = [8]$ or $\ddot{f}(d, e) = [5 + 9] = [14]$? Fortunately, $[8] = [14]$ since $8 \equiv 14 \pmod 3$, so there is not any ambiguity in this case.

It is important that addition of congruence classes in $\mathbb{Z}_k$ is well-defined, in other words whenever we have $m \equiv m' \pmod k$ and $n \equiv n' \pmod k$ we have $m + n \equiv m' + n' \pmod k$ [check that it is]. This is what allows us to define $\ddot{f}$ unambiguously, and it is precisely the axiom E2 that allows us to do this. So when we pass to the quotient structure (going from $\mathbb{Z}$ to $\mathbb{Z}_3$ in this example), the interpretations of the function symbols are well defined.

What about the predicate symbols? We would define $\ddot{A}$ in a similar way, by declaring that $\ddot{A}([m_1], [m_2], \ldots, [m_n])$ holds iff $A^{\mathcal{I}}(m_1, m_2, \ldots, m_n)$ holds. In this case the only predicate symbol is E. So we have $\ddot{E}([m], [n])$ iff $E^{\mathcal{I}}(m, n)$ i.e. iff $m \equiv n \pmod 3$, which happens iff $[m] = [n]$. So $\ddot{E}$ ends up being the relation of equality in the new quotient structure.

Exercises

1. Let S be a first-order system with equality. Prove that

(a) $\vdash_S \forall x\, Exx$.

(b) $\vdash_S \forall x \forall x\, (Exy \to Eyx)$.

(c) $\vdash_S \forall x \forall x \forall x\, (Exy \to (Eyz \to Exz))$.

2.* A first-order language $\mathcal{L}$ is *pure monadic* if it has no function symbols or constant symbols, and all the predicate symbols are unary, so $\mathcal{L} = (\varnothing, P, \varnothing, r)$. Let $\mathcal{L}$ be a pure monadic language, and let $\mathcal{I}$ be an interpretation of $\mathcal{L}$ with domain D. Define a relation $\sim$ on D by declaring that, for $d, e \in D$,

$$d \sim e \qquad \text{iff} \qquad \text{for all } A \in P,\ A^{\mathcal{I}}(d) \text{ iff } A^{\mathcal{I}}(e).$$

(a) Show that $\sim$ is an equivalence relation on D.

We denote the set of equivalence classes under $\sim$ by D^*, and for $d \in D$ we denote the equivalence class of d by $[d]$. For $A \in P$, define the relation A^* on D^* by declaring that $A^*([d])$ holds iff $A^{\mathcal{I}}(d)$ holds. Let $\mathcal{I}^*$ be the interpretation with domain D^*, and with each predicate symbol A interpreted by A^*.

(b) Show that A^* is a well-defined relation on D^*.

(c) For each $\mathcal{I}$-assignment v in $\mathcal{I}$, let v^* be the $\mathcal{I}$-assignment in $\mathcal{I}^*$ with $v^*(x) = [v(x)]$ for each variable x. Show that for every wff $\mathcal{A}$ we have $\mathcal{I} \models_v \mathcal{A}$ iff $\mathcal{I}^* \models_{v^*} \mathcal{A}$.

[Text: p 105-111]

5.2 Normal models

Definition 5.4. *Let S be a first-order system with equality. A* normal *model of S is a model of S in which E is interpreted by equality.*

We have seen an example of a first order system with equality which had a model which is not a normal model. However, in that case we were able to take a quotient to obtain a normal model. we will now see that this process works in any first order system with equality.

Theorem 5.5. *Let S be a first-order system with equality. If S has a model then it has a normal model.*

Proof *. Let $\mathcal{M}$ be a model for S with domain D, in which the predicate symbol A is interpreted by $A^{\mathcal{M}}$, the function symbol f by $f^{\mathcal{M}}$ and the constant symbol c by $c^{\mathcal{M}}$. By Lemma 5.3 we know that $E^{\mathcal{M}}$ is an equivalence relation on D: denote this equivalence relation by $\approx$, and for $d \in D$ denote the equivalence class of d under $\approx$ by $[d]$. Let D^* denote the set of equivalence classes of D under $\approx$.

We define the interpretation $\mathcal{I}$ as follows.

- The domain of $\mathcal{I}$ is D^*.
- For each constant symbol c, the interpretation of c is $c^{\mathcal{I}} = [c^{\mathcal{M}}]$.
- For each n-ary function symbol f, the interpretation of f is $f^{\mathcal{I}}$, where

$$f^{\mathcal{I}}([d_1], [d_2], \ldots, [d_n]) = [f^{\mathcal{M}}(d_1, d_2, \ldots, d_n)].$$

- For each n-ary predicate symbol A, the interpretation of A is $A^{\mathcal{I}}$, where

$$A^{\mathcal{I}}([d_1], [d_2], \ldots, [d_n]) \Leftrightarrow A^{\mathcal{M}}(d_1, d_2, \ldots, d_n).$$

We must check that these are well-defined: in other words we must check that if $d_i \approx d_i'$ for $i = 1, 2, \ldots, n$ then

$$[f^{\mathcal{M}}(d_1, d_2, \ldots, d_n)] = [f^{\mathcal{M}}(d_1', d_2', \ldots, d_n')]$$

and

$$A^{\mathcal{M}}(d_1, d_2, \ldots, d_n) \Leftrightarrow A^{\mathcal{M}}(d_1', d_2', \ldots, d_n')$$

This is easy to do: E2 ensures that $f^{\mathcal{I}}$ is well-defined and E3 ensures that $A^{\mathcal{I}}$ is well-defined.

For each $\mathcal{I}$-assignment v in $\mathcal{I}$, let v^* be the $\mathcal{I}$-assignment given by $v^*(x) = [v(x)]$ for every variable x. Notice that every $\mathcal{I}$-assignment in $\mathcal{I}$ is of the form v^* for some $\mathcal{I}$-assignment v in $\mathcal{I}$. We prove that for every wff $\mathcal{A}$,

$$I \models_v \mathcal{A} \quad \Leftrightarrow \quad I^* \models_{v^*} \mathcal{A}$$

by induction on the number of connectives and quantifiers in $\mathcal{A}$.

First, we should prove by induction on the complexity of the term that if t is any term then $\widetilde{v^*}(t) = [\widehat{v}(t)]$. We leave this as an exercise.

Base step: If $\mathcal{A}$ is the atomic wff $A(t_1, t_2, \ldots, t_n)$, then we have

$$\begin{aligned} I \models_v A(t_1, t_2, \ldots, t_n) &\Leftrightarrow A^{\mathcal{M}}(\widehat{v}(t_1), \widehat{v}(t_2), \ldots, \widehat{v}(t_n)) \\ &\Leftrightarrow A^{\mathcal{I}}([\widehat{v}(t_1)], [\widehat{v}(t_2)], \ldots, [\widehat{v}(t_n)]) \\ &\Leftrightarrow A^{\mathcal{I}}(\widetilde{v^*}(t_1), \widetilde{v^*}(t_2), \ldots, \widetilde{v^*}(t_n)) \\ &\Leftrightarrow I \models_{v^*} A(t_1, t_2, \ldots, t_n) \end{aligned}$$

Inductive Step: Suppose $\mathcal{A}$ has at least one connective or quantifier, and the result holds for all wffs with fewer connectives and quantifiers than $\mathcal{A}$. There are three cases to consider, depending on whether $\mathcal{A}$ is of the form $\neg\mathcal{B}$, $(\mathcal{B} \to \mathcal{C})$ or $\forall x\, \mathcal{B}$.

Case 1 Suppose $\mathcal{A}$ is of the form $\neg\mathcal{B}$. Then

$$I \models_v \mathcal{A} \Leftrightarrow I \not\models_v \mathcal{B} \Leftrightarrow I \not\models_{v^*} \mathcal{B} \Leftrightarrow I \models_{v^*} \mathcal{A}$$

Case 2 Suppose $\mathcal{A}$ is of the form $(\mathcal{B} \to \mathcal{C})$. Then

$$\begin{aligned} I \models_v \mathcal{A} &\Leftrightarrow I \not\models_v \mathcal{B} \text{ or } I \models_v \mathcal{C} \\ &\Leftrightarrow I \not\models_{v^*} \mathcal{B} \text{ or } I^* \models_{v^*} \mathcal{C} \\ &\Leftrightarrow I \models_{v^*} \mathcal{A} \end{aligned}$$

Case 3 Suppose $\mathcal{A}$ is of the form $\forall x\, \mathcal{B}$.

Suppose that $\mathcal{I} \models_{v^*} \mathcal{A}$, and let $d \in D$. Then $[d] \in D^*$, and $\mathcal{I} \models_{v^*} \forall x\, \mathcal{B}$, so $\mathcal{I} \models_{v^*\frac{[d]}{x}} \mathcal{B}$. But we have $v^*\frac{[d]}{x} = v\frac{d}{x}^*$, so $\mathcal{I} \models_{v\frac{d}{x}^*} \mathcal{B}$, so by inductive hypothesis $\mathcal{M} \models_{v\frac{d}{x}} \mathcal{B}$. Since this holds for all $d \in D$, $\mathcal{M} \models_v \forall x\, \mathcal{B}$, i.e. $\mathcal{M} \models_v \mathcal{A}$.

Conversely, suppose $\mathcal{M} \models_v \mathcal{A}$, and let $[d] \in D^*$. Then we have $v^*\frac{[d]}{x} = v\frac{d}{x}^*$, and $\mathcal{M} \models_{v\frac{d}{x}} \mathcal{B}$, so by inductive hypothesis we have $\mathcal{I} \models_{v^*\frac{[d]}{x}} \mathcal{B}$. Since this holds for all $[d] \in D^*$, $\mathcal{I} \models_{v^*} \forall x\, \mathcal{B}$, as required.

This completes the induction.

We have shown that for every $\mathcal{A}$ and every $\mathcal{M}$-assignment v, $\mathcal{M} \models_v \mathcal{A}$ if and only if $\mathcal{I} \models_{v^*} \mathcal{A}$. From this it follows that $\mathcal{M} \models \mathcal{A}$ if and only if $\mathcal{I} \models \mathcal{A}$. Thus $\mathcal{I}$ is also a model of the system S.

It remains only to show that $\mathcal{I}$ is a normal model, in other words that $E^{\mathcal{I}}$ is the relation of equality. Well, by definition of $E^{\mathcal{I}}$, for all $[d], [e] \in D^*$ we have that

$$\begin{aligned} E^{\mathcal{I}}([d],[e]) \quad &\text{iff} \quad E^{\mathcal{M}}(d,e) \\ &\text{iff} \quad d \approx e \\ &\text{iff} \quad [d] = [e], \end{aligned}$$

so $E^{\mathcal{I}}$ is indeed the relation of equality as required. □

In summary, if $\mathcal{I}$ is a model for S then we define a new model $\mathcal{I}^*$ based on the equivalence classes of the domain of S. That is, the domain of $\mathcal{I}^*$ is the set of equivalence classes of the domain of $\mathcal{I}$. We then show that $\mathcal{I} \models_v \mathcal{A}$ if and only if $\mathcal{I}^* \models_{v^*} \mathcal{A}$ for every wff $\mathcal{A}$, where v^* is the $\mathcal{I}$-assignment $v^*(x) = [v(x)]$. Hence $\mathcal{I}^*$ is also a model for S and we then show that $\mathcal{I}^*$ is a normal model.

From now on, we will allow ourselves to use more convenient notation, called *infix* notation. For example, if A is a binary predicate symbol we can write $x\ A\ y$ instead of Axy, and if f is a binary function symbol we can write $x\, f\, y$ instead of fxy. We will also allow ourselves to use familiar symbols such as $=$, $+$ and $-$ for predicate and function symbols. With this convention, and using $=$ instead of E, the axioms E1–E3 become

$$x = x. \tag{E1}$$

$$(t_k = u \to (ft_1 \ldots, t_k \ldots t_n = ft_1 \ldots u \ldots t_n)) \tag{E2}$$

where f is an n-ary function symbol and $t_1, \ldots, t_n$ and u are terms.

$$(t_k = u \rightarrow (At_1 \ldots t_k \ldots, t_n \rightarrow At_1 \ldots u \ldots t_n)) \qquad \text{(E3)}$$

where A is an n-ary predicate symbol and $t_1, \ldots, t_n$ and u are terms.

[Text: p 110-111]

First-order group theory

Let $\mathcal{L}_G$ be the language of group theory, as described above. Let G be the first-order system with equality which has as its proper axioms E1–E3 and the following:

G1 $Efxfyzfffxyz$

G2 $Efxcx$

G3 $Efxgxc$.

If we use infix notation, writing $=$ for E, tu or (tu) for ftu and t^{-1} for gt, and we write e instead of c, then these axioms become

G1 $x(yz) = (xy)z$

G2 $xe = x$

G3 $xx^{-1} = e$

In this abbreviated form the axioms are similar to the familiar axioms for a group. Some remarks are in order. First, notice that we allow x, y and z as free variables in these axioms. It would be possible to take the universal closures of these axioms. However, this would not make any difference to the set of theorems—we can prove the universal closures of these axioms from the axioms themselves, by generalisation. If, instead, we had taken the universal closures as axioms, then we could prove the axioms in the above form by appropriate use of axiom schema K4 and modus ponens. In practice, we find that it is more convenient to have the axioms in the above form.

Another difference between G1–G3 and the axioms of a group as they appear in most algebra textbooks is that, for example, G2 is usually expressed in the form "there exists $e \in G$ such that…". It is not necessary for us to assert the existence here—the constant symbol c must be interpreted by some element of the domain. A similar comment applies to G3.

Finally, note that many algebra textbooks give (apparently) stronger forms of G2 and G3, namely that e is a two-sided identity, and that every element has a two-sided inverse. However, it is

reasonably easy to prove that the above form is sufficient, or alternatively, to assume that e is a left identity and every element has a left inverse. It is not sufficient to assume that e is a right identity and every element has a left inverse.

The danger of allowing ourselves the abbreviations to make the group axioms more legible is that we can be misled by the "intended" meaning. It is not the case that every model of G is a group. For example, consider our earlier example of an interpretation of $\mathcal{L}_G$ (in Section 5.1). Suppose we change the interpretation of g to the function $n \mapsto 2n$. Axiom G3 now asserts that for all $n \in \mathbb{Z}$, $n + 2n \equiv 0 \pmod 3$, which is true. Thus this is a model of G. However, it is not a group, because g is not the "real" inverse of the group $(\mathbb{Z}, +)$.

On the other hand, if we restrict ourselves to *normal* models, then we get what we want: every normal model of G is a group. And if we take the quotient structure from a model of G, as defined in Theorem 5.5, then the resulting structure will be a group. Of course if we take the normal model based on this interpretation (i.e. as we did in the proof of Theorem 5.5), then we would have a group.

[Text: p 112-116]

Exercises

2. Recall that a *partial order* or a set X is a relation ρ on X which is

   Reflexive: for all $x \in X$, $x \ \rho \ x$;

   Antisymmetric: for all $x, y \in X$, if $x \ \rho \ y$ and $y \ \rho \ x$ then $x = y$; and

   Transitive: for all $x, y, z \in X$, if $x \ \rho \ y$ and $y \ \rho \ z$ then $x \ \rho \ z$.

   A *preorder* on X is a relation on X which is reflexive and transitive.

   (a) Suppose ρ is a preorder on X. We define a relation $\sim$ on X by declaring that $x \sim y$ iff both $x \ \rho \ y$ and $y \ \rho \ x$. Show that $\sim$ is an equivalence relation.

   For each $x \in X$ we denote the equivalence class of x under $\sim$ by $[x]$. let $X^* = \{\, [x] \mid x \in X \,\}$ be the set of equivalence classes. Define a relation σ on X^* by declaring that $[x] \ \sigma \ [y]$ iff $x \ \rho \ y$.

   (b) Show that σ is well-defined.

   (c) Show that σ is a partial order on X^*.

3. The language L_{PO} of partially ordered sets has two binary predicate symbols, B (for $\leqslant$) and E (for equality). Describe a first-order system with equality P, such that every normal $\mathcal{L}_{PO}$ structure, $\mathcal{M}$, is a model of P iff $B^{\mathcal{M}}$ is a partial order.

   [Remark: if we take a (not necessarily normal) model of P then $B^{\mathcal{M}}$ will be a preorder, and the quotient structure in Theorem 5.5 is the partial order constructed as in the previous question.]

5.3 The Peano Postulates

We now move on to considering a first-order system for the arithmetic of the natural numbers. There are two reasons to focus on the natural numbers (rather than, say, the real numbers needed to do calculus). The first is that they are a simpler system, and it makes sense to learn to walk before we can run. The second is that we can build the richer number systems from the natural numbers, so if we get a firm foundation for the natural numbers we will be able to build up a solid theory.

In the next section we will set the system of first order arithmetic. Before doing so we will examine the first set of axioms given for the natural numbers, the so-called *Peano postulates*. In fact they were first given by Dedekind, and they were formulated long before the introduction of formal systems.

If we allow the usual intuitive operations from set theory, such as unions and intersections, then we can characterise the natural numbers by the following statements:

P1 0 is a natural number.

P2 Every natural number n has a successor $s(n)$ which is a natural number.

P3 0 is not the successor of any natural number.

P4 If two natural numbers have the same successor, they are equal.

P5 If A is a set of natural numbers which contains 0, and which contains the successor of each of its elements, then A contains all natural numbers.

What do we mean when we say that these axioms characterise the natural numbers? The answer is that

(a) The natural numbers satisfy these axioms; and

(b) If any other structure satisfies the axioms then it must be "essentially the same" as the natural numbers (in a sense we will make precise below).

Note that we do not *prove* that the natural numbers satisfy these axioms: what we have done is describe some properties which we feel correspond to our intuition about what the natural numbers are like, and explore what the consequences of just those properties are.

We remark that one consequence is that every natural number other than 0 is a successor: let $A = \{0\} \cup \{\, s(n) \mid n \in \mathbb{N}\}$. Then we certainly have $0 \in A$, and if $n \in A$ then $s(n) \in A$, so by P5 we know that A contains all natural numbers.

By "essentially the same" we mean that there must be an isomorphism between the natural numbers and the other structure. In other words, suppose we have a set X, an element 0_X and a function s_X such that

- $0_X \in X$;
- for every $x \in X$, $s_X(x) \in X$;
- There is no $x \in X$ with $s_X(x) = 0_X$;
- If $x_1, x_2 \in X$ with $s_X(x_1) = s_X(x_2)$ then $x_1 = x_2$;
- If $A \subseteq X$ with $0_X \in A$ and $s_X(a) \in A$ for every $a \in A$ then $A = X$;

Then there is a bijection $\varphi : \mathbb{N} \to X$ such that $\varphi(0) = 0_X$ and, for every $n \in \mathbb{N}$, $\varphi(s(n)) = s_X(\varphi(n))$.

Proof of (b).* We construct the bijection φ by recursion. For these purposes let us define an *approximation* to be a function f such that

- $\mathrm{dom}(f) = \{0, 1, \ldots, k\}$ for some $k \in \mathbb{N}$;
- $f(0) = 0_X$; and
- if $0 \leqslant i < k$ then $f(s(i)) = s_X(f(i))$.

Let $\mathbf{A}$ denote the set of all approximations. Let

$$A = \{\, n \in \mathbb{N} \mid \text{ for all } f, g \in \mathbf{A}, \text{ if } n \in \mathrm{dom}(f) \cap \mathrm{dom}(g) \text{ then } f(n) = g(n) \,\}.$$

We will show that $A = \mathbb{N}$, using P5. We certainly know that if $f, g \in \mathbf{A}$ then $f(0) = 0_X = g(0)$. Thus $0 \in A$. Now suppose that $n \in A$: we must show that $s(n) \in A$. So let $f, g \in \mathbf{A}$ with $s(n) \in \mathrm{dom}(f) \cap \mathrm{dom}(g)$. Then $n \in \mathrm{dom}(f) \cap \mathrm{dom}(g)$ so, since $n \in A$, $f(n) = g(n)$. But then we must have

$$f(s(n)) = s_X(f(n)) = s_X(g(n)) = g(s(n)),$$

so $s(n) \in A$, as required. Hence, by P5, A contains all of $\mathbb{N}$.

Now we let $B = \{\, n \in \mathbb{N} \mid \text{there is some } f_n \in \mathbf{A} \text{ with } \mathrm{dom}(f_n) = \{0, 1, \ldots, n\} \,\}$. Again, we will use P5 to show that $B = \mathbb{N}$. First, to show that $0 \in B$, we note that we have $f_0 = \{(0, 0_X)\}$. Now suppose that $n \in B$. We define $f_{s(n)} = f_n \cup \{(s(n), s_X(f_n(n)))\}$. We have $\mathrm{dom}(f_{s(n)}) = \mathrm{dom}(f_n) \cup \{s(n)\} = \{0, 1, \ldots, n, s(n)\}$. We also have $f_{s(n)}(0) = f_n(0) = 0_X$ (since $f_n \in \mathbf{A}$). We must check that if $0 \leqslant i < s(n)$ then $f_{s(n)}(s(i)) = s_X(f_{s(n)}(i))$. Well, if $i < n$ we have $s(i) \in \mathrm{dom}(f_n)$ and

$$f_{s(n)}(s(i)) = f_n(s(i)) = s_X(f_n(i)) = s_X(f_{s(n)}(i)),$$

as required. On the other hand, if $i = n$ then we have explicitly defined $f_{s(n)}(s(n)) = s_X(f_n(n)) = s_X(f_{s(n)}(n))$ as required.

We can now define the function $\varphi : \mathbb{N} \to X$ by declaring that for each $n \in \mathbb{N}$, $\varphi(n) = f_n(n)$. We must show that φ is the isomorphism we require. Notice that, by construction, we certainly have $\varphi(0) = 0_X$ and $\varphi(s(n)) = s_X(\varphi(n))$ for each n. So we only need to show that φ is a bijection.

Claim: φ is onto.

For: Let $C = \mathrm{ran}(\varphi)$. We must show that $0_X \in C$ and, if $x \in C$ then $s_X(x) \in C$, from which it follows that $C = X$ as required. Well, $0_X = \varphi(0) \in C$. Suppose $x \in C$. Then $x = \varphi(n)$ for some n, so $s_X(x) = s_X(\varphi(n)) = \varphi(s(n)) \in C$, as required.

Claim 1: φ is 1–1.

For: let $D = \{\, m \in \mathbb{N} \mid$ for all $n \in \mathbb{N}$, if $\varphi(n) = \varphi(m)$ then $n = m$. As usual, we check that $0 \in D$ and if $n \in D$ then $s(n) \in D$. To show that $0 \in D$, suppose $n \in \mathbb{N}$ with $\varphi(n) = \varphi(0) = 0_X$. If $n \neq 0$ we would have $n = s(m)$ for some m, so $\varphi(n) = s_X(\varphi(m))$, which is impossible since $0_X \neq s_X(x)$ for all $x \in X$.

Now let $n \in D$. To show that $s(n) \in D$, suppose that $m \in \mathbb{N}$ with $\varphi(m) = \varphi(s(n))$. By the above, since $s(n) \neq 0$ we cannot have $m = 0$, so $m = s(k)$ for some k. But then we have $\varphi(s(k)) = \varphi(s(n))$, so $s_X(\varphi(k)) = s_X(\varphi(n))$, so $\varphi(k) = \varphi(n)$, so $k = n$, so $s(k) = s(n)$, i.e. $m = s(n)$ as required.

Thus φ is the isomorphism we claimed it to be.

The Peano postulates make no mention of addition and multiplication. However, we can prove that they allow us to define functions by *recursion*, in other words, if we want to define a function $f : \mathbb{N} \to A$, we can give a definition of the form

$$\begin{aligned} f(0) &= a_0 \\ f(s(n)) &= g(f(n)) \end{aligned}$$

where a_0 is some element of a and $g : A \to A$ is some function. This lets us define addition and multiplication recursively, as follows:

$$\begin{aligned} n + 0 &= n \\ n + s(m) &= s(n + m) \end{aligned}$$

and

$$\begin{aligned} n \cdot 0 &= 0 \\ n \cdot s(m) &= n \cdot m + n \end{aligned}$$

They also say nothing about the order $\leqslant$, but again we may define it in terms of what we have above, by declaring that $m \leqslant n$ iff there is some $k \in \mathbb{N}$ with $m + k = n$.

Exercises

3. Show that the Peano Postulates prove that for all natural numbers m, n and k, $m+(n+k) = (m+n)+k$. [Hint: consider $A_{m,n} = \{\, k \in \mathbb{N} \mid m + (n + k) = (m + n) + k \,\}$.]

4. Use the Peano Postulates to prove that for all $m, n \in \mathbb{N}$, $m + n = n + m$. [Hint: let $A = \{\, m \in \mathbb{N} \mid (\forall n \in N)(m + n = n + m) \,\}$. Show that we can apply P5 to A. To show that $0 \in A$, and to show that if $n \in A$ then $s(n) \in A$, will both require proofs using P5.]

5.4 First order arithmetic

We set up the following first-order system N, called the system of first-order arithmetic. Our language $\mathcal{L}_N$ has one constant symbol 0, two binary predicates $=$ and $\leqslant$, two binary function symbols $+$ and $\cdot$, and one unary function symbol $'$. The axioms are E1–E3, together with

N1 $x' \neq 0$

N2 $(x' = y' \to x = y)$

N3 $x + 0 = x$

N4 $x + y' = (x + y)'$

N5 $x \cdot 0 = 0$

N6 $x \cdot y' = (x \cdot y) + x$

N7 $x + y = z \to x \leqslant z$

N8 $(\mathcal{A}(\frac{0}{x}) \to (\forall x\,(\mathcal{A} \to \mathcal{A}(\frac{x'}{x})) \to \forall x\,\mathcal{A}))$, for any wff $\mathcal{A}$.

Notice that our axioms differ somewhat from the Peano postulates. First of all, note that we do not require any equivalent to P1 and P2: any interpretation of $\mathcal{L}_N$ will have to have an interpretation for 0 and for $'$. On the other hand, our axioms N3–N7 are not necessary for the Peano system—if we are allowed basic set theory, then we can define the functions $+$ and $\cdot$ and the relation $\leqslant$ using the equivalents of N3–N7.

Also, the single axiom P5 has been replaced by an infinite family N8 of axioms (one for each wff $\mathcal{A}$).

The difference between N8 and P5 is very important. We cannot express the statement P5 as a wff of $\mathcal{L}_N$. This is because the quantifier ranges over the collection of all subsets of the natural numbers ($(\forall A)$ if A is a set and $0 \in A$ then ...), whereas in $\mathcal{L}_N$ the quantifiers can only range over the natural numbers themselves. The best we can manage is the axiom schema N8, which is weaker than the full strength of P5: there are only countably many wffs of $\mathcal{L}_N$, whereas there are uncountably many subsets of the natural numbers.

In fact, the axioms N1–N8 do *not* characterise the natural numbers. We will show below that the system N has a normal model which is essentially different to the intended one.

Example*: A normal model that is not $\mathbb{N}$. For each natural number n, we define a term $\vec{n}$ as follows: $\vec{0}$ is 0 and, if we have defined $\vec{n}$, then $n \overrightarrow{+1}$ is $(\vec{n})'$. In other words, $\vec{n}$ is the n^{th} successor of 0. Thus the closed term $\vec{n}$ is supposed to be interpreted by the natural number n. We extend the language $\mathcal{L}_N$ by adding a new constant symbol d. For each natural number n, let $\mathcal{A}_n$ be the wff $\vec{n} \leqslant d$, and let S be the system obtained from N by adding all the wffs $\mathcal{A}_n$ as axioms of S.

Let Σ be the set of universal closures of the axioms of S. It is easy to check that an interpretation $\mathcal{I}$ of $\mathcal{L}_N$ is a model of S iff it is a model of Σ. So we will show that Σ has a model. Note that the wffs $\mathcal{A}_n$ are all closed wffs so they are all in Σ.

Consider any finite set $\Pi \subseteq \Sigma$. Then only finitely many of the wffs $\mathcal{A}_n$ are in Π (possibly none at all), so there is some m such that if $\mathcal{A}_n \in \Pi$ then $n \leqslant m$. To obtain a model of Π, we take the usual interpretation of $\mathcal{L}_N$, and interpret the new constant symbol d by m. As we saw in an earlier exercise, this gives us a model of Π.

Thus any finite subset of Σ has a model. But then, by the Compactness Theorem, Σ has a model, so S has a model. Therefore, by Theorem 5.5, S has a normal model, $\mathcal{M}$. In this model, $0 \leqslant d^{\mathcal{M}}$, $1 \leqslant d^{\mathcal{M}}$, $2 \leqslant d^{\mathcal{M}}$, and so on. So d is interpreted by an element which cannot be obtained from 0 by applying the successor function a finite number of times. If we ignore the interpretation of d, then this is a normal model of N in which not every element can be obtained from 0 by applying the successor function a finite number of times. So this model is essentially different from $\mathcal{N}$.

We have seen that our system N does not have $\mathcal{N}$ as its only normal model. However, it might still be reasonable to hope that N is at least strong enough to derive every true statement about the natural numbers. It turns out that this is not the case: Gödel showed that there is a true statement about the natural numbers which is not a theorem of $\mathcal{N}$. This is not because we chose the axioms badly—Gödel's proof actually showed that there cannot be a "reasonable" first-order system S with language $\mathcal{L}_N$ such that $\mathcal{N}$ is a model of S and every true statement about $\mathcal{N}$ is a theorem of S.

[Text: p 116-120]

Chapter 6

First Order Arithmetic

In this chapter we consider some properties of the formal system N for first-order arithmetic which we introduced in the previous chapter.

6.1 Basic consequences of the axioms of N

We have suggested that the system N might conceivably be strong enough to prove everything which is true about $\mathbb{N}$ (even though we know that Gödel proved that this is not the case). We will now show how to prove some of the basic properties about $\mathbb{N}$ in the system N.

We nearly always proceed by using induction: to show that $\vdash_N \forall x\, \mathcal{A}$, we show that $\vdash_N \mathcal{A}(\frac{0}{x})$, and that $\vdash_N \forall x\, (\mathcal{A} \to \mathcal{A}(\frac{x'}{x})$, and then appeal to N8 which says that

$$\left(\mathcal{A}\left(\frac{0}{x}\right) \to \left(\forall x\, \left(\mathcal{A} \to \mathcal{A}\left(\frac{x'}{x}\right)\right) \to \forall x\, \mathcal{A}\right)\right)$$

Note: Now that we are using infix notation we will allow extra brackets so that the wffs are easier to read.

Proposition 6.1. *We have* $\vdash_N \forall x\, (x = 0 \vee \exists y\, (x = y'))$.

Proof. We use induction as indicated above. Let $\mathcal{A}$ be the wff

$$(x = 0 \vee \exists y\, (x = y')).$$

Base: We have the derivation

1.	$x = x$	E1
2.	$\forall x\,(x = x)$	1, Gen
3.	$(\forall x\,(x = x) \to (0 = 0))$	K4
4.	$0 = 0$	2, 3, MP
5.	$((0 = 0) \to$ $(\neg(0 = 0) \to \exists y\,(0 = y')))$	taut. inst.
6.	$(\neg(0 = 0) \to \exists y\,(0 = y'))$	4, 5, MP

so $\vdash_N (\neg(0 = 0) \to \exists y\,(0 = y'))$, i.e. $\vdash_N \mathcal{A}(\frac{0}{x})$.

Inductive step: We have the derivation

1.	$x = x$	E1
2.	$\forall x\,(x = x)$	1, Gen
3.	$(\forall x\,(x = x) \to (x' = x'))$	K4
4.	$x' = x'$	2, 3, MP
5.	$((x' = x') \to \exists y\,(x' = y')))$	log. val.
6.	$\exists y\,(x' = y')$	4, 5, MP
7.	$(\exists y\,(x' = y') \to$ $(\neg(x' = 0) \to \exists y\,(x' = y')))$	K1
8.	$(\neg(x' = 0) \to \exists y\,(x' = y'))$	6, 7, MP

[Note that at line 5 we used the fact that a certain wff is logically valid and hence provable in $K_{\mathcal{L}_N}$. In general, if t is free for y in $\mathcal{A}$ then $(\mathcal{A}(\frac{t}{y}) \to \exists y\,\mathcal{A})$ is logically valid. In this case, x is free for y in $x' = y'$.]

Thus we have $\vdash_N \mathcal{A}(\frac{x'}{x})$. Hence since we have $\vdash_N (\mathcal{A}(\frac{x'}{x}) \to (\mathcal{A} \to \mathcal{A}(\frac{x'}{x})))$ we have $\vdash_N (\mathcal{A} \to \mathcal{A}(\frac{x'}{x}))$, and by generalisation we have $\vdash_N \forall x\,(\mathcal{A} \to \mathcal{A}(\frac{x'}{x}))$, as required.

Hence, by N8, $\vdash_N \forall x\,\mathcal{A}$, as required. □

Proposition 6.2. *We have*

$$\vdash_N \forall x\,\forall y\,\forall z\,(x + (y + z) = (x + y) + z).$$

Proof *. Let $\mathcal{A}$ be $x + (y + z) = (x + y) + z$. We will prove that $\vdash_N (\forall z)\mathcal{A}$ by induction, and then use generalisation twice to get $\forall x\,\forall y\,\forall z\,\mathcal{A}$.

Base: We have the derivation

1.	$x + 0 = x$	N3
2.	$\forall x\,(x + 0 = x)$	1, Gen
3.	$(\forall x\,(x + 0 = x) \to (y + 0 = y))$	K4
4.	$y + 0 = y$	2, 3, MP
5.	$((y + 0 = y) \to$ $(x + (y + 0) = x + y))$	E2
6.	$x + (y + 0) = x + y$	4, 5, MP
7.	$(\forall x\,(x + 0 = x) \to$ $((x + y) + 0 = x + y))$	K4
8.	$(x + y) + 0 = x + y$	2, 7, MP
9.	$(((x + y) + 0 = x + y) \to$ $((x + y = x + y) \to$ $(x + y = (x + y) + 0)))$	E3
10.	$((x + y = x + y) \to$ $(x + y = (x + y) + 0))$	8, 9, MP

11.	$x = x$	E1
12.	$\forall x\,(x = x)$	11, Gen
13.	$(\forall x\,(x = x) \to (x + y = x + y)$	K4
14.	$x + y = x + y$	12, 13, MP
15.	$x + y = (x + y) + 0$	14, 10, MP
16.	$((x + y = (x + y) + 0) \to$ $((x + (y + 0) = x + y) \to$ $(x + (y + 0) = (x + y) + 0)))$	E3
17.	$((x + (y + 0) = x + y) \to$ $(x + (y + 0) = (x + y) + 0))$	15, 16, MP
18	$x + (y + 0) = (x + y) + 0$	8, 17, MP

In other words we have $\vdash_N \mathcal{A}(\frac{0}{z})$.

Inductive Step: We need to show that $\vdash_N (\forall z)(\mathcal{A}(\frac{x}{z}) \to \mathcal{A}(\frac{z'}{z}))$. To do this we show that $\vdash_N (\mathcal{A} \to \mathcal{A}(\frac{z'}{z}))$, and then apply generalisation. The obvious way to do this would be to show that $\{\mathcal{A}\} \vdash_N \mathcal{A}(\frac{z'}{z})$, then use the deduction theorem. Of course, we must give a derivation which does not apply generalisation to z: otherwise we can give an easy derivation using the instance $((\forall z)\mathcal{A} \to \mathcal{A}(\frac{z'}{z}))$ of K4. A suitable derivation can be found, using similar techniques to the one in the base above, using N4 instead of N3. The derivation will be somewhat longer but not inherently more difficult. Here are the first few lines of the derivation:

1.	$x + y' = (x + y)'$	N4
2.	$\forall y\,(x + y' = (x + y)')$	1, Gen
3.	$(\forall y\,(x + y' = (x + y)') \rightarrow$ $(x + z' = (x + z)')$	K4
4.	$x + z' = (x + z)'$	2, 3, MP
5.	$\forall x\,(x + z' = (x + z)')$	4, Gen
6.	$(\forall x\,(x + z' = (x + z)') \rightarrow$ $(y + z' = (y + z)'))$	K4
7.	$y + z' = (y + z)'$	5, 6, MP
8.	$((y + z' = (y + z)') \rightarrow$ $(x + (y + z') = x + (y + z)'))$	E2
9.	$x + (y + z') = x + (y + z)'$	7, 8, MP
10.	$(\forall y\,(x + y' = (x + y)') \rightarrow$ $(x + (y + z)' = (x + (y + z))'))$	K4
11.	$x + (y + z)' = (x + (y + z))'$	2, 10, MP
12.	$x + (y + z) = (x + y) + z$	Hyp
13.	$((x + (y + z) = (x + y) + z) \rightarrow$ $((x + (y + z))' = ((x + y) + z)'))$	E2
14.	$(x + (y + z))' = ((x + y) + z)'$	12, 13, MP
⋮		

However, there is a minor problem, which is that the deduction theorem does not let us apply generalisation to a free variable of $\mathcal{A}$. We have applied generalisation to x and y at lines 2 and 5.

This use of generalisation (although preventing us from using our version of the Deduction Theorem) does not really cause a problem, because at this stage we have not used the hypothesis $\mathcal{A}$, so x and y are not yet free variables that we have to worry about. If $\Sigma \vdash_N \mathcal{B}$, regardless of what instances of generalisation we used, we can get $\Sigma \vdash_N (\mathcal{B} \rightarrow (\mathcal{A} \rightarrow \mathcal{B}))$ by K1, so $\Sigma \vdash_N (\mathcal{A} \rightarrow \mathcal{B})$. However, we need to find a more precise wording for the Deduction Theorem to cover this. Alternatively, we could change our wff $\mathcal{A}$ to $x_1 + (x_2 + z) = (x_1 + x_2) + z$, and make suitable changes to the derivation in the Base and the one above. A third option would be take the derivation above and use the method of the proof of the Deduction Theorem to translate it into a derivation of $(\mathcal{A} \rightarrow \mathcal{A}(\frac{z'}{z}))$. We will leave all these as exercises. □

6.2 Using adequacy of $K_{\mathcal{L}}$ to deduce theorems of N

The examples we considered in the previous section show us that it is difficult for us to work strictly within the framework of our formal system N to show that $\vdash_N \mathcal{A}$. In this section we will consider an alternative approach, which uses the Adequacy Theorem for $K_{\mathcal{L}}$. The idea is that we prove, using standard mathematical reasoning rather than reasoning within the system, that any model of the axioms of N must make $\mathcal{A}$ true. In other words, let $\Sigma = \{N1, N2, \ldots, N8\}$ and let $\mathcal{M}$ be a normal model of N, then the members of Σ are true in $\mathcal{M}$. Hence if $\mathcal{M} \models \mathcal{A}$ for every normal model $\mathcal{M}$ of N, then by the Adequacy Theorem, $\Sigma \vdash_{K_{\mathcal{L}}} \mathcal{A}$ and therefore $\overline{\Sigma} \models \mathcal{A}$, where $\overline{\Sigma}$ is the

set of universal closures of axioms of N. Hence, $\Sigma \vdash_{K_{\mathcal{L}}} \mathcal{A}$ by Proposition 4.19, and so $\vdash_N \mathcal{A}$.

One important thing to remember when we are using this approach is that we must not just consider the "intended" model $\mathcal{N}$: we must show that $\mathcal{A}$ is true in *every* model of N. However, we note that it is enough to restrict ourselves to *normal* models of N (models in which $E^{\mathcal{M}}$ is equality), see Theorem 6.3 below.

Note: Since we are using infix notation we will, for clarity, use hats to denote the interpretations of function symbols and constant symbols. So, for example, if $\mathcal{M}$ is an arbitrary normal model of N, we will use $\widehat{+}$ to denote the interpretation of $+$ in $\mathcal{M}$, $\widehat{0}$ to denote the interpretation of 0, and so on. We will use $=$ to denote the interpretation of $=$ (since $=^{\mathcal{M}}$ really is $=$). It is to avoid ambiguity that we restrict ourselves to normal models.

Theorem 6.3. *Let $\mathcal{A}$ be a wff of $K_{\mathcal{L}_N}$. Suppose that $\mathcal{A}$ is true in every normal model of N. Then $\vdash_N \mathcal{A}$.*

Proof *. Since we have $\vdash_N \mathcal{A}$ iff $\vdash_N \overline{\mathcal{A}}$, where $\overline{\mathcal{A}}$ is the universal closure of $\mathcal{A}$, we may assume WLOG that $\mathcal{A}$ is a closed wff.

Let Σ be the set of proper axioms of N and let $\overline{\Sigma} = \{\, \overline{\mathcal{C}} \mid \mathcal{C} \in \Sigma \,\}$ (Recall that $\overline{\mathcal{C}}$ is the universal closure of $\mathcal{C}$). We claim that $\overline{\Sigma} \vdash_{K_{\mathcal{L}_N}} \mathcal{A}$. To prove this we will show that $\overline{\Sigma} \models \mathcal{A}$. So let $\mathcal{I}$ be an interpretation of $\mathcal{L}_N$ such that $\mathcal{I} \models \mathcal{B}$ for each $\mathcal{B} \in \overline{\Sigma}$. For each axiom $\mathcal{C}$ of N we have $\overline{\mathcal{C}} \in \Sigma$, so $\mathcal{I} \models \overline{\mathcal{C}}$. But $\mathcal{I} \models \overline{\mathcal{C}}$ iff $\mathcal{I} \models \mathcal{C}$, so each axiom of N is true in $\mathcal{I}$. Now, any derivation in N is a derivation from Σ in $K_{\mathcal{L}_N}$, so if $\mathcal{D}$ is a theorem of N then $\Sigma \vdash_{K_{\mathcal{L}_N}} \mathcal{D}$, and so (by the Soundness Theorem) $\Sigma \models \mathcal{D}$. Thus, since $\mathcal{I} \models \mathcal{C}$ for all $\mathcal{C} \in \Sigma$, $\mathcal{I} \models \mathcal{D}$. This shows us that $\mathcal{I}$ is a model of N.

Note that we do not know that $\mathcal{I}$ is a normal model, so we cannot apply the hypothesis yet. However, we do know by Theorem 5.5 that there is a normal model $\mathcal{I}^*$ such that for every $\mathcal{B}$ we have $\mathcal{I} \models \mathcal{B}$ iff $\mathcal{I}^* \models \mathcal{B}$. Hence $\mathcal{I}^*$ is a normal model of N. So, by hypothesis, $\mathcal{I}^* \models \mathcal{A}$, and thus $\mathcal{I} \models \mathcal{A}$.

We now have $\mathcal{I} \models \mathcal{A}$ whenever $\mathcal{I} \models \mathcal{B}$ for all $\mathcal{B} \in \overline{\Sigma}$, so $\overline{\Sigma} \models \mathcal{A}$, so by the Adequacy Theorem $\overline{\Sigma} \vdash_{K_{\mathcal{L}}} \mathcal{A}$. Let $\mathcal{A}_1, \ldots, \mathcal{A}_n$ be a derivation of $\mathcal{A}$ from $\overline{\Sigma}$ in $K_{\mathcal{L}}$. It is easy to prove (by induction on $\mathcal{I}$) that $\Sigma \vdash_{K_{\mathcal{L}}} \mathcal{A}_i$ for each $\mathcal{I}$, so $\Sigma \vdash_{K_{\mathcal{L}}} \mathcal{A}$. But a derivation from Σ in $K_{\mathcal{L}}$ is just a derivation in N, so $\vdash_N \mathcal{A}$, as required. □

Example 6.4. Show that

$$\vdash_N \forall x \, \forall y \, \forall z \, (x + (y + z) = (x + y) + z).$$

Solution. Let $\mathcal{I}$ be a normal model of N with domain D. We will show that $\mathcal{I} \models \forall x \, \forall y \, \forall z \, (x + (y + z) = (x + y) + z)$. Let $\mathcal{A}$ be the wff $x + (y + z) = (x + y) + z$. We show that $\mathcal{I} \models \mathcal{A}(\frac{0}{z})$ and that $\mathcal{I} \models \forall z \, (\mathcal{A} \to \mathcal{A}(\frac{z'}{z}))$.

Base: We know that $\mathcal{I} \models x + 0 = x$, by N3, so we have $d\hat{+}\widehat{0} = d$ for all $d \in D$. So for every $d, e \in D$ we have

$$d\hat{+}(e\hat{+}\widehat{0}) = d\hat{+}e = (d\hat{+}e)\hat{+}\widehat{0}.$$

Hence $\mathcal{I} \models x + (y + 0) = (x + y) + 0$.

Inductive step: Suppose v is a valuation such that $\mathcal{I} \models_v \mathcal{A}$: we will show that $\mathcal{I} \models_v \mathcal{A}(\frac{z'}{z})$. Putting $d = v(x)$, $e = v(y)$ and $f = v(z)$ we have $d\hat{+}(e\hat{+}f) = (d\hat{+}e)\hat{+}f$. We also know that $\mathcal{I} \models x + y' = (x + y)'$, by N4, so we have $e\hat{+}f\hat{\prime} = (e + f)\hat{\prime}$, so

$$d\hat{+}(e\hat{+}f\hat{\prime}) = d\hat{+}(e\hat{+}f)\hat{\prime}$$
$$= (d\hat{+}(e\hat{+}f))\hat{\prime} = ((d\hat{+}e)\hat{+}f)\hat{\prime} = (d\hat{+}e)\hat{+}f\hat{\prime},$$

so $\mathcal{I} \models_v (x+(y+z') = (x+y)+z')$, i.e. $\mathcal{I} \models_v \mathcal{A}(\frac{z'}{z})$. Since this holds for all v, $\mathcal{I} \models (\mathcal{A} \to \mathcal{A}(\frac{z'}{z}))$, so $\mathcal{I} \models \forall z\,(\mathcal{A} \to \mathcal{A}(\frac{z'}{z}))$.

By N8 we have

$$I \models \left(\mathcal{A}\left(\frac{0}{z}\right) \to \left(\forall z\,\left(\mathcal{A} \to \mathcal{A}\left(\frac{z'}{z}\right)\right) \to \forall z\,\mathcal{A}\right)\right),$$

so we have $\mathcal{I} \models \forall z\,\mathcal{A}$. Thus we have $\mathcal{I} \models \forall y\,\forall z\,\mathcal{A}$, so $\mathcal{I} \models \forall x\,\forall y\,\forall z\,\mathcal{A}$.

Since this holds for every normal model of N, we have $\vdash_N \forall x\,\forall y\,\forall z\,\mathcal{A}$,as required. □

Example 6.5. Show that

$$\vdash_N \forall x\,\forall y\,(x + y = y + x).$$

Solution. Let $\mathcal{I}$ be a normal model of N with domain D. Let $\mathcal{A}$ be the wff $\forall y\,(x + y = y + x)$. We show that $\mathcal{I} \models \mathcal{A}(\frac{0}{x})$, and that $\mathcal{I} \models \forall x\,(\mathcal{A} \to \mathcal{A}(\frac{x'}{x}))$.

Base: We need to show that $\mathcal{I} \models \forall y\,(0 + y = y + 0)$, in other words that $\widehat{0}\hat{+}e = e\hat{+}\widehat{0}$ for all $e \in D$. We prove this by induction. The base follows from the fact that $\widehat{0}\hat{+}\widehat{0} = \widehat{0}\hat{+}\widehat{0}$. The inductive step follows from the fact that if $\widehat{0}\hat{+}e = e\hat{+}\widehat{0}$ then

$$\widehat{0}\hat{+}e\hat{\prime} = (\widehat{0}\hat{+}e)\hat{\prime} = (e\hat{+}\widehat{0})\hat{\prime} = e\hat{\prime} = e\hat{\prime}\hat{+}\widehat{0}.$$

Inductive Step: We need to show that if $d \in D$ with $d\hat{+}e = e\hat{+}d$ for all $e \in D$, then $d\hat{\prime}\hat{+}e = e\hat{+}d\hat{\prime}$ for all $e \in D$. Again, we prove this by induction. The base is the fact that $d\hat{\prime}\hat{+}\widehat{0} = \widehat{0}\hat{+}d\hat{\prime}$, which we know from the base of the main induction. The inductive step of the inner induction

follows because if $d\hat{\prime}\hat{+}e = e\hat{+}d\hat{\prime}$ then

$$\begin{aligned} d\hat{\prime}\hat{+}e\hat{\prime} &= (d\hat{\prime}\hat{+}e)\hat{\prime} \\ &= (e\hat{+}d\hat{\prime})\hat{\prime} \\ &= ((e\hat{+}d)\hat{\prime})\hat{\prime} \\ &= ((d\hat{+}e)\hat{\prime})\hat{\prime} \\ &= (d\hat{+}e\hat{\prime})\hat{\prime} \\ &= (e\hat{\prime}\hat{+}d)\hat{\prime} \\ &= e\hat{\prime}\hat{+}d\hat{\prime} \end{aligned}$$

as required.

Hence, by induction, $\mathcal{I} \models \forall x\, \mathcal{A}(x)$, as required. □

6.3 Expressibility in $\mathcal{L}_N$

We have seen examples to show that we can prove some familiar properties of $\mathbb{N}$ as theorems of N. For example, consider $\leqslant$. Recall that we have an axiom, N7, which asserts that

$$(\exists y\, (x + y = z) \leftrightarrow x \leqslant z).$$

We can readily prove from this that $\leqslant$ is a linear order, in other words we can prove

- $\forall x\, (x \leqslant x)$;
- $\forall x\, \forall y\, ((x \leqslant y \wedge y \leqslant x) \rightarrow (x = y))$;
- $\forall x\, \forall y\, \forall z\, ((x \leqslant y \wedge y \leqslant z) \rightarrow (x \leqslant z))$; and
- $\forall x\, \forall y\, (x \leqslant y \vee y \leqslant x)$.

These will require some proofs by induction along similar lines to the earlier examples. From now on we will be making a number of assertions about what can and cannot be proven in N without really justifying these assertions any more.

Now consider the "real" $\leqslant$ in $\mathbb{N}$. If we have $m, n \in \mathbb{N}$ with $m \leqslant n$, can we prove this in N? Recall the *numeral terms* $\ddot{n}$, defined by $\ddot{n} = \underbrace{h h \cdots h}_{n} 0$, so we have $\ddot{0} = c$, $\ddot{1} = hc$, $\ddot{2} = hhc$, $\ddot{5} = hhhhhc$ and so on. In the infix notation these are written as $\ddot{n} = 0\overbrace{\prime\prime\cdots\prime}^{n}$, so $\ddot{0} = 0$, $\ddot{1} = 0'$, $\ddot{5} = 0''''$ and so on.

For each m and n, we have a (closed) wff $\ddot{m} \leqslant \ddot{n}$, which might or might not be a theorem of N. In fact, we have the following:

Proposition 6.6. *Let $m, n \in \mathbb{N}$. If $m \leqslant n$ then $\vdash_N \ddot{m} \leqslant \ddot{n}$, and if $m \not\leqslant n$ then $\vdash_N \neg(\ddot{m} \leqslant \ddot{n})$.*

Similarly, consider the property of being a prime number. Recall that a natural number $n \geq 2$ is a prime if whenever $n = ab$ we have $n = a$ or $n = b$. Let $\mathsf{P}x$ be the wff

$$(\ddot{2} \leqslant x \wedge \forall y \forall z\, (x = y \cdot z \to (x = y \vee x = z))).$$

Then if n is a prime number we have $\vdash_N \mathsf{P}\ddot{n}$ and if n is not a prime number we have $\vdash_N \neg\mathsf{P}\ddot{n}$. This is an example of the general notion of *expressibility*:

Definition 6.7. *Let R be an n-ary relation on $\mathbb{N}$. Then R is* expressible in N *if there is a wff $\mathcal{A}$ with $x_1, x_2, \ldots, x_n$ as its free variables, such that for any $m_1, m_2, \ldots, m_n \in \mathbb{N}$,*

- *if $R(m_1, m_2, \ldots, m_n)$ holds then $\vdash_N \mathcal{A}(\frac{\ddot{m}_1}{x_1}, \frac{\ddot{m}_2}{x_2}, \ldots, \frac{\ddot{m}_n}{x_n})$; and*
- *if $R(m_1, m_2, \ldots, m_n)$ does not hold then $\vdash_N \neg\mathcal{A}(\frac{\ddot{m}_1}{x_1}, \frac{\ddot{m}_2}{x_2}, \ldots, \frac{\ddot{m}_n}{x_n})$.*

The examples we have considered so far were quite close to the symbols built in to our language, so it was easy to express the relations in N. What about some other properties? Consider, for example, "y is a power of x", i.e.

$$R(x, y) \qquad \text{iff} \qquad \exists n\, (y = x^n).$$

For any particular n, it is easy to express $y = x^n$: e.g. $y = x^3$ iff $y = x \cdot (x \cdot x)$. But we cannot have a wff whose length depends on the value of the n chosen in the $\exists n$.

It turns out that "y is a power of x" is expressible, but the argument requires an extra trick. We will give a different problem which can be solved using the same trick.

Consider the property "can be written as a product of primes"? we might reasonably expect that N proves that $\forall x\, \mathcal{P}x$, where $\mathcal{P}x$ is a wff which expresses that x is a product of primes, since we used proof by induction in MATHS 255 to show that every natural number can be written as a product of primes.

The problem is that, as with "a power of", we don't know in advance how many primes are needed in the expression.

We can use a trick here, which is to use the fact that the set of finite sequences of natural numbers is countable to construct a bijection between the set of natural numbers and the set of finite sequences of natural numbers.

Let S be the set of finite non-empty sequences of natural numbers.

The clever bit is to construct the bijection $\varphi : \mathbb{N} \to S$ in such a way that the relation "y is one of the terms of the sequence $\varphi(x)$" is expressible in N, by the wff $\mathcal{A}xy$ say, and "y is the product

of the terms of the sequence $\varphi(x)$" is expressible in N by the wff $\mathcal{B}xy$. Then "x is a product of primes" is expressible in N by the wff

$$\exists z\,(\mathcal{B}zx \wedge \forall y\,(\mathcal{A}zy \rightarrow \mathsf{P}y))$$

Please note that this example is only meant to illustrate some important concepts: in fact we know that every natural number greater than 1 can be written as a product of primes, so in terms of our definition we could express "can be written as a product of primes" with the wff $\ddot{2} \leqslant x$.

The relations that we have considered are expressible (though not easily) in N. It is not the case that all relations in $\mathbb{N}$ are expressible in N.

[Text p 130-132]

Chapter 7

Gödel's Incompleteness Theorem

The goal of this chapter is to give an outline of how Gödel's Incompleteness Theorem is proved. Gödel proved that the system N is incomplete, by exhibiting a closed wff $\mathcal{U}$ such that neither $\mathcal{U}$ nor $\neg\mathcal{U}$ can be a theorem of N. Of course, to prove this we need some assumption about the consistency of N: otherwise both $\mathcal{U}$ and $\neg\mathcal{U}$ would be theorems of N whatever $\mathcal{U}$ is.

It is reasonable to think that N is consistent, because it has a model: the model $\mathcal{N}$. As long as we accept our intuitive understanding of the natural numbers, then $\mathcal{N}$ is indeed a model of N. Gödel himself made a stronger assumption than simple consistency, namely that N is ω-consistent (a notion that will be defined below). Rosser later showed how to weaken the hypothesis to simple consistency.

In our outline of Gödel's proof we will use the notion of expressibility and 2 new notions: recursive functions and Gödel numbering.

Recursive functions and expressible relations

One of the two main ingredients of Gödel's proof is a characterisation of those relations which are expressible, in terms of a completely separate notion. Unofficially (but well enough for our purposes), a function $f : \mathbb{N}^k \to \mathbb{N}$ is *recursive* if we can give an algorithm which calculates $f(n_1, n_2, \ldots, n_k)$ from input $n_1, n_2, \ldots, n_k$. Gödel proved that a relation $R \subseteq \mathbb{N}^k$ is expressible if and only if the characteristic function $\chi_R : \mathbb{N}^k \to \{0, 1\}$ given by

$$\chi_R(n_1, n_2, \ldots, n_k) = \begin{cases} 1 & \text{if } R(n_1, \ldots, n_k) \text{ holds} \\ 0 & \text{otherwise} \end{cases}$$

is recursive.

Officially, the definitions is as follows:

Definition 7.1 $(*)$**.** *A function $f : \mathbb{N}^k \to \mathbb{N}$ is* recursive *if it can be built from the following three basic functions using the operations of composition, recursion and least-solution:*

- *the zero function $z(n) = 0$;*
- *the successor function $s(n) = n + 1$;*
- *the projection function $\pi_i^k(n_1, n_2, \ldots, n_k) = n_i$.*

(by "least-solution" we mean that $g : \mathbb{N}^k \to \mathbb{N}$ is obtained from $f : \mathbb{N}^{k+1} \to \mathbb{N}$ by declaring that

$$g(n_1, n_2, \ldots, n_k) = \min\{\, n \in \mathbb{N} \mid f(n_1, n_2, \ldots, n_k, n) = 0,$$

assuming that f is a function for which this set is always non-empty.)

[Text p 137-145]

Gödel Numbering

The other key ingredient in the proof is Gödel's technique for assigning to each symbol σ, string of symbols $\mathcal{A}$ or sequence of strings $S = \mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ a code number $g(\sigma)$, $g(\mathcal{A})$ or $g(S)$, in such a way that given a natural number n there is an algorithm for finding the symbol σ with $g(\sigma) = n$, or the string $\mathcal{A}$ with $g(\mathcal{A}) = n$ or the sequence S with $g(S) = n$, or asserting that no such object exists. The textbook describes one such procedure: it is rather inefficient in that it gives a code number to all strings of symbols rather than more efficient methods which would only give code numbers to wffs, but it is at least easy to describe.

We restrict ourselves to languages which only allow variables $x_1, x_2, \ldots$, constants $c_1, c_2, \ldots$, function symbols f_k^n for $n, k \geq 1$, and predicate symbols A_k^n for $n, k \geq 1$ (recall that the superscripts n give the arity as n). We assign numbers to each symbol as follows:

- $g(() = 3$
- $g()) = 5$
- $g(,) = 7$
- $g(\neg) = 9$
- $g(\to) = 11$
- $g(\forall) = 13$
- $g(x_k) = 7 + 8k$

- $g(c_k) = 9 + 8k$
- $g(f_k^n) = 11 + 8 \cdot 2^n \cdot 3^k$
- $g(A_k^n) = 13 + 8 \cdot 2^n \cdot 3^k$

Example 7.2. Find the symbols (if any) with $g(\sigma) = 587$ and with $g(\sigma) = 333$.

Solution. Since $587 > 13$, if there is such a σ it must be an x_k, c_k, f_k^n or A_k^n. We divide 8 into 587: $587 = 8 \cdot 73 + 3 = 11 + 8 \cdot 72 = 11 + 8 \cdot 8 \cdot 9 = 11 + 8 \cdot 2^3 \cdot 3^2$, so we have $g(f_2^3) = 587$.

Similarly if $333 = g(\sigma)$ then σ is x_k, f_k^n or A_k^n, so again we divide 8 into 333: $333 = 8 \cdot 41 + 5 = 13 + 8 \cdot 40 = 13 + 8 \cdot 2^3 \cdot 5$. Since this does not have the form $13 + 8 \cdot 2^n \cdot 3^k$, there is no σ with $g(\sigma) = 333$. □

Next, we extend the notion to assign a Gödel number $g(\mathcal{A})$ for a string $\mathcal{A}$ of symbols: if $\mathcal{A} = \sigma_0\sigma_1 \ldots \sigma_n$ then

$$g(\mathcal{A}) = 2^{g(\sigma_0)} \cdot 3^{g(\sigma_1)} \cdot 5^{g(\sigma_2)} \cdot \ldots \cdot p_n^{g(\sigma_n)},$$

where p_k is the k^{th} odd prime number.

Similarly, if $S = (\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_n)$ is a sequence of strings of symbols, then we define

$$g(S) = 2^{g(\mathcal{A}_0)} \cdot 3^{g(\mathcal{A}_1)} \cdot 5^{g(\mathcal{A}_2)} \cdot \ldots \cdot p_n^{g(\mathcal{A}_n)}.$$

Now, the process of figuring out what symbol, string of symbols, or sequence of strings has a given number as its Gödel number is an algorithm. Therefore, as we said earlier, it corresponds to an expressible relation. In other words there is a wff $\mathcal{A}$ such that for each n, $\vdash_N \mathcal{A}(\frac{\ddot{n}}{x})$ if there is a wff $\mathcal{B}$ with $g(\mathcal{B}) = n$, and $\vdash_N \neg\mathcal{A}(\frac{\ddot{n}}{x})$ if there is no such $\mathcal{B}$. Similarly, one can establish that each of the following relations is expressible in N:

Sub: $\mathsf{Sub}(m, n, p, q)$ holds iff there is a wff $\mathcal{B}$ and a term t such that $g(\mathcal{B}(\frac{t}{x_i})) = m$, $g(\mathcal{B}) = n$, $g(x_i) = p$ and $g(t) = q$.

Lax: $\mathsf{Lax}(n)$ holds iff there is an instance $\mathcal{B}$ of a logical axiom with $g(\mathcal{B}) = n$.

Prax: $\mathsf{Prax}(n)$ holds iff there is an instance $\mathcal{B}$ of a proper axiom with $g(\mathcal{B}) = n$.

Prf: $\mathsf{Prf}(m)$ holds iff there is a sequence $S = (\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k)$ of wffs such that $g(S) = m$ and S is a derivation in N.

Pf: $\mathsf{Pf}(m, n)$ holds iff there is a sequence $S = (\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k)$ of wffs such that $g(S) = m$ and S is a derivation in N and $g(\mathcal{A}_k) = n$.

W: $\mathsf{W}(m, n)$ holds iff there is a wff $\mathcal{A}$ with x_1 as its only free variable with $g(\mathcal{A}) = m$, and there is a derivation $S = \mathcal{B}_1, \mathcal{B}_2 \ldots, \mathcal{B}_k$ in N of the wff $\mathcal{B}_k = \mathcal{A}(\frac{\ddot{m}}{x_1})$ with $g(S) = n$.

[Text: p 146-149]

Exercises

1. Find the symbols if any with:
   (a) $g(\sigma) = 47$;
   (b) $g(\sigma) = 100$;
   (c) $g(\sigma) = 445$;
   (d) $g(\sigma) = 779$;
2. Find the code sequence for the wff
$$(A_1^1(x_1) \to A_1^2(x_1, f_3^1(x_1))).$$
3. Find the wff corresponding to the code sequence
$$2^3 \times 3^9 \times 5^{61} \times 7^3 \times 11^{15} \times 13^5 \times 17^{11} \times 19^{445} \times 23^3 \times 29^{71} \times 31^5 \times 37^5.$$

Pulling the threads together

As we mentioned before, we need to assume that N has some new consistency property.

Definition 7.3. *Let S be a first-order system with $\mathcal{L}(S) = \mathcal{L}_N$. We say that S is* ω-consistent *if, for every wff $\mathcal{A}$, if $\vdash_N \mathcal{A}(\frac{\ddot{n}}{x})$ for every $n \in \mathbb{N}$ then $\nvdash_N \neg\forall x\,\mathcal{A}$.*

In other words, if we can prove $\mathcal{A}(\frac{\ddot{n}}{x})$ for every $n \in \mathbb{N}$, we might not be able to prove $\forall x\,\mathcal{A}$, but we had better not be able to prove $\neg\forall x\,\mathcal{A}$.

We will assume that N is ω-consistent. As mentioned above, this assumption can be dropped but it makes it trickier to explain the idea.

The relation W defined above is the key to the whole thing. Since W is expressible, there is a wff $\mathcal{W}$ such that for all m, n, $\vdash_N \mathcal{W}(\frac{\ddot{m}}{x_1}, \frac{\ddot{n}}{x_2})$ if $\mathsf{W}(m, n)$ holds and $\vdash_N \neg\mathcal{W}(\frac{\ddot{m}}{x_1}, \frac{\ddot{n}}{x_2})$ if $\mathsf{W}(m, n)$ does not hold.

Let $\mathcal{B}$ be the wff $\forall x_2\,\neg\mathcal{W}$. Let $p = g(\mathcal{B})$. Let $\mathcal{U}$ be the wff $\mathcal{B}(\frac{\ddot{p}}{x_1})$. Now $\mathcal{U}$ is a closed wff: we claim that neither $\mathcal{U}$ nor $\neg\mathcal{U}$ could be a theorem of N.

Recall that $\mathcal{W}$ expresses the relation W, so $\vdash_N \mathcal{W}(\frac{\ddot{m}}{x_1}, \frac{\ddot{n}}{x_2})$ holds iff there is a wff $\mathcal{A}$ with x_1 as its only free variable, $g(\mathcal{A}) = m$ and n is the Gödel number of a proof of $\mathcal{A}(\frac{\ddot{m}}{x_1})$, and similarly $\neg\mathcal{W}(\frac{\ddot{m}}{x_1}, \frac{\ddot{n}}{x_2})$ expresses that n is *not* the Gödel number of a proof of some wff $\mathcal{A}(\frac{\ddot{m}}{x_1})$, where $g(\mathcal{A}) = m$. Thus, since we have $g(\mathcal{B}) = p$, the assertion that for all n, $\mathsf{W}(p, n)$ is false is the assertion that every n is not the Gödel number of a proof of $\mathcal{B}(\frac{\ddot{p}}{x_1})$, in other words it is the assertion that there is no proof in N of $\mathcal{B}(\frac{\ddot{p}}{x_1})$. In other words, $\mathcal{U}$ asserts that $\mathcal{U}$ is not a theorem of N.

If $\mathcal{U}$ were a theorem of N it would have to be true in $\mathbb{N}$, in other words, there would have to be no proof in N of $\mathcal{U}$, a contradiction. So $\mathcal{U}$ cannot be a theorem of N.

Since $\mathcal{U}$ is not a theorem, we know that each n is not the Gödel number of a proof of $\mathcal{B}(\frac{\ddot{p}}{x_1})$, i.e. $\mathsf{W}(p, n)$ is false for all n, so $\vdash_N \neg\mathcal{W}(\frac{\ddot{m}}{x_1}, \frac{\ddot{n}}{x_2})$ holds for all n. Hence, by ω-consistency, $\neg\forall x_2 \neg\mathcal{W}(\frac{\ddot{p}}{x_1})$ cannot be a theorem of N, i.e. $\nvdash_N \neg\mathcal{U}$.

A final comment

We have outlined Gödel's proof that N is incomplete. A natural question is: can we extend N to obtain a system that is complete? The answer is no. Any first order system that contains the system of natural numbers, whose set of proper axioms is recursive, and is consistent, is not complete.

[Text: p 150-155]

> The supreme triumph of reason is to cast doubt upon its own validity. *Miguel de Unamuno*

Chapter 8

Axiomatic Set Theory

We will describe a formal system for set theory. All kinds of mathematical objects—functions, ordered pairs, relations—can be thought of as sets. Thus set theory forms a suitable foundation for all branches of mathematics.

The Language of Set Theory

The language we will use for our formal system is called the *Language of Set Theory*, or LST. It has no constant symbols, no function symbols, and only two binary predicate symbols, $=$ and $\in$.

There are a number of other symbols we use when we are talking about sets: $\subseteq$, $\varnothing$ and $\mathbb{P}$ (power set operation) for example. We will define all these as abbreviations for statements in the language, rather than including them in the language itself. For example, we define

$$x \subseteq y$$

to be an abbreviation for $\forall z(z \in x \to z \in y)$.

Example 8.1. Show that we can express the following in LST:

- $x = \{a\}$
- $x = \{a, b\}$
- $x = \langle a, b\rangle$
- $x = \mathbb{P}(a)$

What is a set?

The notion of a set is fundamental to most of pure mathematics. However, it is not easy to give a description which is not circular (roughly speaking, a set is a collection of objects. But what is a collection? It is a set of elements...). What we will do is to describe how sets can be formed from other sets, and study properties of the collection of all sets (often referred to as the "Universe of Sets").

One answer to the question "What is a set?" would be "A set is the collection of all objects which satisfy some property, and any such collection is a set". More precisely, if $P(x)$ denotes some property of the object x, then the collection of all objects x satisfying $P(x)$ is a set, which we denote by $\{ x \mid P(x) \}$. For example, we have the sets

$$\{ x \mid x \text{ is an even integer} \}$$
$$\{ f \mid f \text{ is a continuous function from } \mathbb{R} \text{ to } \mathbb{R} \}$$

The idea that we can form sets in this way is known as the *Principle of Comprehension*. Unfortunately, it does not work: if we assume that any collection formed in this way is a set, we run into problems.

Richard's Paradox

Let $P(x)$ be the property "x is a natural number which can be defined in fewer than sixteen English words". Let

$$A = \{ x \mid P(x) \}.$$

Since there are only finitely many English words, there are only finitely many strings of fewer than sixteen English words. So there are only finitely many natural numbers which can be defined in fewer than sixteen English words, in other words A is finite. This means that there must be a least natural number n which is not in A. But then n is

"The least natural number which can not be defined in fewer than sixteen English words".

In other words, we can define n using fewer than sixteen English words, which is a contradiction.

Russell's Paradox

Let $Q(x)$ be the property "x is not an element of itself", ie "$x \notin x$". Let

$$B = \{ x \mid Q(x) \}.$$

Then, for any object x, $x \in B \Leftrightarrow x \notin x$. In particular, this applies to the object B itself, in other words $B \in B \Leftrightarrow B \notin B$. But this is impossible.

- Richard's Paradox arises because we allowed a property $P(x)$ which was too vague. We must restrict ourselves to properties which can be precisely expressed.
- Russell's Paradox is different: the property $x \notin x$ is perfectly precise. The problem is rather that the collection B is in some sense "too large" to be regarded as an object.

So our guiding principles will be that collections which are not too large and which are defined using a precise property are sets: other collections of objects need not be sets.

By a precise property, we mean any property which can be expressed in LST. We have already seen a few examples of statements which can be expressed in LST. In fact, anything you might "reasonably" want to be expressible in LST is indeed expressible in LST—for example "f is a function from A to B", "$\leqslant$ is a well-order on X", "There is a 1–1 function from X to Y", ...

What do we mean by saying that a collection is "not too large"? Basically, we have three ways of knowing a collection is not too large:

1. It is contained in some collection we already know to be a set.
2. It is built from some collection we already know to be a set, either by taking the union or the power set of that set.
3. It is obtained by taking some set A, and replacing each element of A with some precisely defined object.

For example, suppose we know that the collection $\mathbb{N}$ of natural numbers is a set. Then the collection A of all subsets of $\mathbb{N}$ with at least two elements is also a set. To show this, we must show that A is not too large, and can be defined using a formula of LST. Since $\mathbb{N}$ is a set, $\mathbb{P}(\mathbb{N})$ is a set, so since A is contained in $\mathbb{P}(\mathbb{N})$, A is not too large to be a set. Finally, "x is a subset of $\mathbb{N}$ with at least two elements" can be expressed in LST as

$$(\forall y(y \in x \to y \in \mathbb{N}) \wedge \exists y \exists z(y \in x \wedge z \in x \wedge y \neq z))$$

So A is indeed a set.

The Axiom System ZFC

ZFC is the commonly accepted system of axioms for set theory. ZFC stands for "Zermelo-Fraenkel plus Choice". These axioms describe the universe of sets.

- The first of them (Extensionality) restricts our discussion to hereditary sets (see below).
- Others (Empty set, Infinity) tell us that certain particular sets exists.
- Others (Pairing, Union, Power set, Subset, Replacement) tell us that if we are given a particular set, then we can construct other sets from it.
- Finally, we have two axioms (Foundation, Choice) which assert that, given a (suitable) set x, some set related to x must exist (without telling us how to find it).

It is important to understand that in this axiom system *every object is a set*. Whenever we see a quantifier $\forall x \ldots$ we interpret it as "for every *set* x, ...", and whenever we see $\exists x \ldots$ we interpret it as "there exists some *set* x such that ...". We need to be able to talk about the elements of sets, as well as the sets themselves. This means that we are only interested in "hereditary" sets, in other words in sets A such that every element of A is also a set, and every element of every element of A is a set, and so on.

We will describe each axiom in two ways: in English, to explain the meaning of the axiom, and in symbols, to give some practice in how to express statements in LST. In some of the more complicated examples, we will also include abbreviated statements in LST, using symbols like $\subseteq$ and $\{x\}$. In practice we almost always use these abbreviated versions. However, it is important to realise that these abbreviated versions can always be expanded to statements which are (strictly) in LST.

Extensionality

Two sets are equal if and only if they have exactly the same elements. In symbols,

$$\forall x \forall y (x = y \leftrightarrow \forall z (z \in x \leftrightarrow z \in y))$$

Empty Set

$\varnothing$ is a set.

To be more precise, we should say that there is a set which has no elements. So the official version is

$$\exists x \forall y (y \notin x)$$

Pairing

If x and y are sets then $\{x, y\}$ is a set. In symbols,

$$\forall x \forall y \exists z (z = \{x, y\})$$

or to be strictly accurate

$$\forall x \forall y \exists z \forall w (w \in z \leftrightarrow (w = x \vee w = y))$$

Proposition 8.2. *For any set x, $\{x\}$ is a set.*

Proof. Suppose x is a set. By the Axiom of Pairing, $\{x, x\}$ is a set. But, by the Axiom of Extensionality, $\{x, x\} = \{x\}$. So $\{x\}$ is a set. □

Union

If x is a set then $\bigcup x = \{\, y \mid \exists z (y \in z \wedge z \in x) \,\}$ is a set. In symbols,

$$\forall x \exists u (u = \bigcup x)$$

or to be strictly accurate

$$\forall x \exists u \forall y (y \in u \leftrightarrow \exists z (y \in z \wedge z \in x))$$

Proposition 8.3. *For any sets x and y, $x \cup y$ is a set.*

Proof. Suppose x and y are sets. Then, by the Axiom of Pairing, $\{x, y\}$ is a set. So, by the Union Axiom, $\bigcup\{x, y\}$ is a set, ie $x \cup y$ is a set. □

Power Set

If x is a set then $\mathbb{P}(x) = \{\, y \mid y \subseteq x \,\}$ is a set. In symbols,

$$\forall x \exists y (y = \mathbb{P}(x))$$

or

$$\forall x \exists y \forall z (z \in y \leftrightarrow z \subseteq x)$$

or to be strictly accurate

$$\forall x \exists y \forall z (z \in y \leftrightarrow \forall w (w \in z \rightarrow w \in x))$$

Comprehension Scheme

For a wff $\mathcal{A}$, we write $\mathcal{A}(y)$ to indicate that the variable y is free in $\mathcal{A}$.

If x is a set and $\mathcal{A}(y)$ is a formula of LST then $\{\, y \in x \mid \mathcal{A}(y) \,\}$ is a set. In symbols,

$$\forall x \exists z (z = \{\, y \in x \mid \mathcal{A}(y) \,\})$$

or to be strictly accurate

$$\forall x \exists z \forall y (y \in z \leftrightarrow (y \in x \wedge \mathcal{A}(y)))$$

The formula $\mathcal{A}(y)$ is allowed to mention other variables besides y.

Unlike the earlier axioms, this is not a single axiom but an infinite family of axioms, one for each formula $\mathcal{A}(y)$ of LST.

Replacement

If x is a set and $\mathcal{A}(z, y)$ is a formula of LST such that for each z there is at most one y satisfying $\mathcal{A}(z, y)$, then $\{\, y \mid \exists z \in x(\mathcal{A}(z, y)) \,\}$ is a set.

In symbols,

$$(\forall z \forall y_1 \forall y_2 ((\mathcal{A}(z, y_1) \wedge \mathcal{A}(z, y_2)) \rightarrow y_1 = y_2) \rightarrow \forall x \exists w \forall y (y \in w \leftrightarrow \exists z (z \in x \wedge \mathcal{A}(z, y))))$$

Again $\mathcal{A}(z, y)$ is allowed to mention other variables as well as z and y, and this gives us an infinite family of axioms.

Another way to think of this axiom is like this: suppose that A is a set, and for every $a \in A$ we can specify some unique set x_a (using a formula of LST). Then

$$\{\, x_a \mid a \in A \,\}$$

is also a set.

Infinity

There is a set x such that $\varnothing \in x$ and, for every $y \in x$, $y \cup \{y\} \in x$. In symbols,

$$\exists x (\varnothing \in x \wedge \forall y (y \in x \rightarrow y \cup \{y\} \in x))$$

or to be strictly accurate

$$\exists x((\exists y \forall z(z \notin y) \wedge y \in x) \wedge \forall z(z \in x \to \exists w(w \in x \wedge \forall t(t \in w \leftrightarrow (t \in z \vee t = z)))))$$

The idea of this axiom is to assert that there is an infinite set. We cannot easily do this directly, since we cannot easily express the notion of "infinite" in LST. So what we do is to specify the existence of a certain set which we can prove to be infinite: for any set a, let a^+ be the set $a \cup \{a\}$. Then the set x given by the Axiom of Infinity satisfies $\varnothing \in x$, $\varnothing^+ \in x$, $\varnothing^{++} \in x$, $\varnothing^{+++} \in x$ and so on. Also, the sets $\varnothing$, $\varnothing^+$, $\varnothing^{++}$, $\varnothing^{+++}$, ... are all different, as can be proved by induction.

Foundation

If x is a non-empty set then there is some $y \in x$ with $y \cap x = \varnothing$. In symbols,

$$\forall x(\exists y(y \in x) \to \exists y(y \in x \wedge \forall z(z \in y \to z \notin x)))$$

Proposition 8.4. *Let x be a set. Then $x \notin x$ and $x \neq \{x\}$.*

Proof. Suppose that x is a set and $x \in x$. Consider the set $\{x\}$ (which is a set by Proposition 8.2). Since $x \in \{x\}$, $\{x\} \neq \varnothing$. So there is some $y \in \{x\}$ with $y \cap \{x\} = \varnothing$. The only element of $\{x\}$ is x itself, so we have $x \cap \{x\} = \varnothing$. But this is impossible, since $x \in x$ and $x \in \{x\}$, so $x \in x \cap \{x\}$.

Suppose that x is a set and $x = \{x\}$. Then $x \in x$, contradicting the previous part. □

Choice

If F is a function with domain I such that $F(i) \neq \varnothing$ for every $i \in I$ then there is a function f with domain I such that $f(i) \in F(i)$ for each $i \in I$.

Equivalent forms (given the other axioms): Zorn's Lemma, wellordering Theorem.