Leaving the cave

An introduction to mathematical thinking

Samuel González-Castillo

R1-2021/10/28

Leaving the cave
An introduction to mathematical thinking

Release 1-2021/10/28

28th October 2021

To truth, reason, purity and perfection.

Contents

Quasi-preface

1 Leaving the cave. This book is the book I would have liked to read when I began my journey as a mathematics undergraduate student. It aims to provide a “baptism of fire” to the world of mathematics by first diving fairly deep into the wonders of mathematical logic and then building some basic mathematics on top of that. Do not be fooled by the seemingly innocent section titles; this is not your standard introductory book. With the exception of the preliminaries (which are fully informal), I have tried to do everything with as much rigour and formal correctness as possible.

2 How this book is organised. This book is divided into chapters which are divided into sections which are divided into blocks. These blocks are numbered within the sections, so the first block of section 2 of chapter *n* will be labelled as 2.1. References to blocks in the same chapter will use their label. Blocks in other chapters (e.g., block 3.4 in chapter I), will be referenced writing the block label after the chapter number (as in I-3.4).

At the end of each chapter, there are some exercises meant for you to work on the material and develop your skills. These exercises are ordered in increasing order of difficulty.

3 This book is a work in progress. The writing of this book is, and will always be, an ongoing effort. Nothing human is perfect and, therefore, nothing human is ever truly finished. I thus believe that, using the power of the internet to allow for continuous change, I should always be open to the possibility of enlarging and improving this book.

On future releases of this book, you should expect to see both many corrections on what has already been written and lots of new material. I have plans to include chapters on arithmetic and euclidean geometry.

While you are reading, please keep in mind that you may find some typos and errors. I have done my best to catch as many of them as I could, but, you know, I am human.

4 I want to hear from you. This book should be accessible to *anyone* who wants to read it. If you ever feel stuck with the material, please, let me know. I will be more than happy to answer any questions you may have. Moreover, your questions will help me understand which parts of the book need more attention.

I make myself no illusion. I know there is still a lot of work to be done and a big margin for improvement. Any feedback — whether positive or negative — will always be welcome. Any. Please, do not hesitate to get in touch.

5 Why is this a quasi-preface?. Prefaces are often home to acknowledgements. I certainly have people to thank for their support and help in writing this book. Nonetheless, I would rather wait until this project evolves a little bit more and is worthy of including the names of these people.

Chapter 0

Preliminaries

§1 Numbers and sets

1.1. I assume that you are familiar with the notions of addition, subtraction, multiplication and equality of integers, and also with the usual ordering of the integers and with their representation in Arabic numerals.

If you are an AI or an alien trying to learn mathematics and you need more insight into these matters, please, study them thoroughly before reading this book. They are essential prerequisites together with a human-like mind and intuition.

1.2. A *set*, vaguely defined, is a collection of mathematical objects. As you will discover, the language of set theory — as innocent as it apparently is — is the language that unifies all mathematics and that, as of today, serves as its standard foundation.

Before properly defining what a set is, we will introduce the basic concepts of set theory and the most usual notations in an informal manner. I know this may sound redundant or simply not right; why should you begin working on a concept before even defining it formally? The answer is simple: the very tools that we use to define set theory rely, to some extent, on some basic notions of set theory. As circular as it may seem, it is the way it needs to be. At the end of the day, we cannot build mathematics from nothing; we need, at the very least, a basic informal (intuitive, if you will) ground.

1.3 Your first set. Let us say that you have a finite collection of mathematical elements; let them be the natural numbers 1, 2 and 3. With them, you can create a set: an imaginary box containing the numbers 1, 2 and 3.

The way in which we represent "the set containing the numbers 1, 2 and 3" is $\{1, 2, 3\}$. The things that are "inside" a set are its *elements* and the elements of a set are said to *belong* to it. We can denote the belonging of an element a to a set A as $a \in \mathrm{A}$, so, for example, we would have $1 \in \{1, 2, 3\}$.

There exists an empty set $\varnothing$ to which no element belongs. Of course, there also exists a set $\{1, 2, 3, \dots\}$ of natural numbers and one of integers. The set of natural numbers is often represented by the symbol $\mathbb{N}$ and that of integers by $\mathbb{Z}$; nonetheless, some people use the notation N and Z. Originally, only the boldface symbols were used in print. This posed a problem when, for instance, writing on a blackboard, because there is no way to effectively write bold letters. Thus, people came up with "blackboard bold" letters $\mathbb{N}, \mathbb{Z}, \dots$ and, eventually, these new symbols made its way to print and ended up substituting the boldface letters that they were once meant to represent.

1.4 Set relations. An important feature of sets is that they can have no "repeated" elements and that the order in which their elements are presented is meaningless. In fact, set equality is *extensional*: two sets A and B are equal, denoted as $A = B$, if and only if any element belonging to A belongs to B and vice-versa. For example,

$$\{1, 1, 2\} = \{2, 1\}.$$

Now, I have a question for you: if you put an apple in a box and then put that box into another box, is it the same as if you had put the apple into just one box? Of course not! Analogously, the sets $\{1\}$ and $\{\{1\}\}$ are not equal, and neither are

$$\{1, 2, 3\}, \quad \{1, \{2, 3\}\}, \quad \{\{1\}, 2, \{3\}\}.$$

If all the elements of a set A are also elements of another set B, we say that A is *included* in B or that A is a *subset* of B, and denote it as $A \subseteq B$. If, in addition, $A \neq B$, we can also say that A is a *proper* or *strict* subset of B, and we can write $A \subset B$. For example, we have

$$\{1, 2\} \subseteq \{1, 2, 3\}, \qquad \{\{1, 2\}\} \subset \{\{1, 2\}, 3\}.$$

Notice that, given any set A, we always have $A \subseteq A$ and $\varnothing \subseteq A$;[1] additionally, if A and B are sets such that $A \subset B$, then, necessarily, $A \subseteq B$. Some people prefer using $\subset$ instead of $\subseteq$ for normal inclusion and $\subsetneq$ in lieu of our $\subset$ for strict inclusion, so keep that in mind when reading other sources.

If two sets A and B satisfy $A \subseteq B$ and $B \subseteq A$, then, trivially, $A = B$. This fact is so convenient when proving set equalities that its use has a name: proof by *double inclusion*. Conversely, of course, $A = B$ implies both $A \subseteq B$ and $A \supseteq B$.

1.5 Cardinality. A set is said to be *finite* if it has a finite number of elements, and *infinite* otherwise. The *cardinality* of a finite set A is denoted by $|A|$ or $\#A$ and is the number of elements it has. Thus, for example,

$$|\{1, \{2, 3\}\}| = 2.$$

The cardinality of an infinite set is...well, keep your infinite sets away for a moment. It is, in a way, the number of elements it has; but, as you might have expected, things get tricky when we deal with infinite stuff.

These definitions are as naive as they could be, but remember that we are doing an informal treatment of set theory just to have a basic framework in which to work.

1.6 Set operations. Now that we know everything we ever wanted to know about how to describe sets (well, kind of), let us define some tools that will allow us to construct new sets from existing ones! Let A and B be sets:

- The *union* of A and B is a set $A \cup B$ containing exclusively the elements of A and the elements of B. For example, $\{1, 3\} \cup \{1, 2\} = \{1, 2, 3\}$.
- The *intersection* of A and B is a set $A \cap B$ containing exclusively the elements that belong to both A and B. For instance, $\{1, 3\} \cap \{1, 2\} = \{1\}$.
- The *power set* of A is the set $\mathscr{P}(A)$ containing exclusively all the subsets of A. Thus, $\{1, 2\} \in \mathscr{P}(\{1, 2, 3\})$. Notice how $A \in \mathscr{P}(A)$ and $\varnothing \in \mathscr{P}(A)$.

[1] We will come back to that later.

- The *subtraction* of a A by B is the set $A \setminus B$ containing exclusively all the elements of A that do not belong to B.

The last set-construction tool that we will study is slightly more complex. Given n sets $A_1, \dots, A_n$, their *cartesian product* is a set $A_1 \times \cdots \times A_n$ consisting exclusively of all the possible ordered sequences of elements $(a_1, \dots, a_n)$ with $a_i \in A_i$ for every i between 1 and n. The elements $(a_1, \dots, a_n)$ are referred to as *n-tuples* or, in the particular case $n = 2$, as *ordered pairs*. We can consider, for example:

$$\{1,2\} \times \{1,3,4\} = \{(1,1),(1,3),(1,4),(2,1),(2,3),(2,4)\}.$$

As we will later see in our formal treatment of set theory, there are ways in which we can encode tuples as sets.

1.7 Functions. The last concept we will deal with is that of a function. Given two sets A and B, a function f from A to B is a rule that assigns to each element $a \in A$ a unique element $f(a) \in B$. This is denoted as

$$\begin{aligned} f : A &\longrightarrow B \\ a &\longmapsto f(a). \end{aligned}$$

For example, we could define a function f from $\{1,2\}$ to $\{3,4\}$ as

$$\begin{aligned} f : \{1,2\} &\longrightarrow \{3,4\} \\ 1 &\longmapsto 4 \\ 2 &\longmapsto 3. \end{aligned}$$

If we are given a function $f : A \longrightarrow B$, the set A is said to be the *domain* of f and the set B its *codomain*. All of this arrows and symbols look very fancy, but there is one thing I want you to keep in mind: notation is meant to be a tool, not a prison. There is no need to use this particular notation each time you define a function. As long as you (and your reader) know what the domain and the codomain of a function are and how it maps the elements of the domain to those of the codomain, everything is fine.

Let $f : A \longrightarrow B$ be a function. The subset of B containing the elements $b \in B$ for which there exists an $a \in A$ such that $f(a) = b$ is known as the *image* of f and is represented by $\operatorname{im} f$. The function f is said to be *injective* if, for every $b \in \operatorname{im} f$, there exists a unique $a \in A$ such that $f(a) = b$. If $B = \operatorname{im} f$, the function is said to be *surjective*. A *bijective* function is a function that is both injective and surjective. For example, the function $f : \mathbb{N} \longrightarrow \mathbb{N}$ that takes $f : n \longmapsto n + 1$ for every $n \in \mathbb{N}$ is injective but not surjective. If a function $f : A \longrightarrow B$ is injective, we define its *inverse* as the function $f^{-1} : \operatorname{im} f \longrightarrow A$ that maps every $b \in A$ to the only $a \in A$ such that $f(a) = b$.

If the domain of a function is the cartesian product of a set A with itself n times $(A \times \overset{n}{\cdots} \times A)$, the function is said to be an n-ary function taking values in A.

§2 Propositions

2.1 Propositions. The central concept in propositional logic is (you guessed it!) that of a proposition. A *proposition* is a statement that can be either true or false. If a proposition is true, we say that its *truth value* is 1; if it is false, we say that its truth value is 0.

Propositions can be modified and joined through the use of *connectives.* A connective is said to be *n*-ary if it takes *n* propositions as input to return a new one. The most basic example of a connective is the unary (1-ary) *negation* connective. As its name suggests, it takes a proposition P and transforms it into a proposition (¬P) that is true if and only if P is false. For instance, let us consider the proposition "I am human". Since that proposition is true, the proposition (¬(I am human)), which stands for "I am not human", is false.

Let $*$ be an arbitrary binary connective that, when acting on two propositions P and Q, yields a proposition $\mathrm{P} * \mathrm{Q}$. If $\mathrm{P} * \mathrm{Q}$ has the same truth value as $\mathrm{Q} * \mathrm{P}$ for any propositions P and Q, we say that $*$ is *symmetric.* Moreover, if, for any propositions P, Q and R, the truth value of $\mathrm{P} * (\mathrm{Q} * \mathrm{R})$ is the same as that of $(\mathrm{P} * \mathrm{Q}) * \mathrm{R}$, we say that $*$ is *associative.*

2.2 Propositional variables and forms. Our attention should not be focused on particular propositions, but on the way connectives act on them in an abstract way. For that purpose, we shall use *propositional variables*, which are nothing more than symbols representing arbitrary propositions. Keep in mind, however, that propositional variables are not propositions by themselves: they only "become" propositions when they have been assigned a particular proposition, this is, a particular truth value. A statement involving only propositional variables and connectives (such as $(\neg p)$) is said to be a *propositional form.* Notice that propositional variables may well be used to represent arbitrary propositional forms.

It should be clear that there is no point in talking about the truth value of a propositional form in an absolute manner: we can only talk about it under a particular assignment of truth values to its propositional variables. Nevertheless, there are two special cases that deserve some attention. If a propositional form is true for any possible assignment of truth values to its variables, it is said to be a *tautology*; if it is false for every possible assignment, it is said to be a *contradiction.*

In order to better distinguish propositions from propositional variables, we will consistently use lower-case letters to represent propositional variables and upper-case letters for propositions.

2.3 Conjunction and disjunction. Moving on to more sophisticated connectives, the *conjunction* connective "and" is a binary (2-ary) connective: if we consider the propositions "I am human" and "I like cheese", we can construct the proposition "I am human and I like cheese". As was to be expected, if p and q are propositional variables, the propositional form "p and q" (written $(p \wedge q)$) is true if and only if p and q are both true. This can be represented using what is known as a *truth table.*

p	q	$p \wedge q$
0	0	0
0	1	0
1	0	0
1	1	1

As you can see, a truth table is just a convenient way of writing down the truth value of a propositional form for each and every possible assignment of truth values to the propositional variables it consists of.

With the two connectives that we already have, we can now formulate our first contradiction! The propositional form $(p \wedge (\neg p))$ is false for any possible truth value of p.

Another example of a binary connective is the *disjunction* connective "or", which, when applied on two propositional variables p and q, transforms them into a propositional form $(p \vee q)$ that is true if and only if at least one of p and q is true. Its corresponding truth table is the following.

p	q	$p \vee q$
0	0	0
0	1	1
1	0	1
1	1	1

Using the disjunction connective, we can now write down our first tautology: $(p \vee \neg p)$. To be or not to be? That's the tautology!

2.4. When combining a group of propositions or propositional forms with connectives, it is important to use parentheses to define the order in which the connectives have to be applied. For instance, given three propositional variables p_1, p_2 and p_3, we can construct the propositional form $(p_1 \wedge (p_2 \vee p_3))$. Notice how failing to use parentheses would lead to an ambiguous expression, for $p_1 \wedge (p_2 \vee p_3)$ is not the same as $(p_1 \wedge p_2) \vee p_3$.

There are, nevertheless, a few exceptions where parentheses are unnecessary. The first of them is result of convention (such as the order of operations that we all learn in elementary school): whenever a unitary connective is acting on a propositional variable, there is no need to write a parentheses. In this way, $\neg p \wedge q$ is the same as $(\neg p) \wedge q$ and should not be mistaken with $\neg(p \wedge q)$. The second exception is more natural: whenever we have a sequence of propositional variables joined by an associative binary connective, parentheses are not necessary within that sequence because, regardless of how we wrote them, the resulting proposition would always yield the same truth values. Thus, for example, we can safely write $p_1 \wedge p_2 \wedge p_3$ instead of $p_1 \wedge (p_2 \wedge p_3)$ or $(p_1 \wedge p_2) \wedge p_3$. Lastly, there is an obvious exception: the parentheses that surround a full symbolic expression are completely dispensable in propositional logic, so $(p \wedge q)$ can perfectly be written as $p \wedge q$.

As a rule of thumb — not just for propositions, but for everything in mathematics — you can omit parentheses whenever doing so leads to no ambiguity, whenever the parentheses add no meaning or whenever a convention removes any possible ambiguity.

2.5 Conditional connective. Let us now introduce one of the most important binary connectives — and, unfortunately, one of the most problematic for newcomers, — the conditional connective. This connective takes two propositions (a *condition* P and a *consequence* Q) and produces a new proposition P → Q, which is read "if P, then Q". The truth table associated to the propositional form $p \to q$ is the following:

p	q	$p \to q$
0	0	1
0	1	1
1	0	0
1	1	1

Please, take your time to digest what this means. What we are saying is that $p \to q$ is true if and only if "if the condition p is true, so is the consequence q".

If the condition is false, we do not care about the consequence: $p \to q$ is automatically true. But if the condition is true, we need the consequence to be true in order for $p \to q$ to be true.

2.6 Example. In a attempt to make things a little bit clearer, let us consider a simple example. Both of us, at some point in our lives, have heard the phrase "if you study hard, you will pass the exam". According to the way in which we have defined the conditional operator, the truth table corresponding to all the possible scenarios is the following:

Study hard	Pass the exam	Study hard → Pass the exam
0	0	1
0	1	1
1	0	0
1	1	1

Let us then examine each case in detail. The first case should be easy: if I do not study hard and I do not pass the exam, is the phrase still true? Of course it is! I did not study hard for the exam, so there was no reason to believe I should have passed it.

Now, what about the second case? If I do not study for the exam but I manage to pass it, is it true that if I study, I will pass the exam? Yes! It is still true. What the statement "if you study, you will pass the exam" tells us is that, provided I have studied, I will pass the exam, but if I did not study, the statement says nothing about what will happen. Nevertheless, the situation would have been different had the phrase been "only if you study, will you pass the exam". Can you spot the difference? How would you express this last statement in an "if…then…" form?

The third case is easy: if I studied but I did not pass the exam, the phrase is, obviously, false. The last statement is equally trivial: if I study hard and I pass the exam, the statement "if you study hard, you will pass the exam" is, clearly, true.

2.7 Remark. There is something very significant that, at this point, should be highlighted. The fact that, for any particular propositions P and Q, the statement "if P, then Q" is true does not imply in any way the existence of a cause-effect relation between P and Q. Connectives, such as the conditional connective, combine statements to create new statements. Thus, the real meaning of a sentence is meaningless (no pun intended); this is all about whether things are true or false. For instance, the statement "if zero equals one, the Earth is flat" is perfectly valid and true (in fact, it is true regardless of your "beliefs" concerning the roundness of our planet!).

Nonetheless, it is true that conditional connectives have something to do with deductions. If we know the propositions P and P → Q to be true, we can indeed deduce

Q to be true, but, as I mentioned earlier, this does not imply the existence of any cause-effect relation between P and Q. Let us, for example, take P to be the statement "humans need water" and Q to be "the sun is a star". Is the statement P → Q true? Sure it is! Both P and Q are true, hence so must be P → Q. Then, from a purely formal point of view, we can deduce that "the sun is a star" from the fact that "humans need water" and "if humans need water, then the sun is a star". Everything we have done is completely meaningless, but, from the perspective of formal logic, it is correct. Notice, by the way, how logic did not allow us to do anything suspicious: if we were able to deduce that "the sun is a star" from "humans need water" it was because, in order to show that "if humans need water, then the sun is a star", we had to assume that "the sun is a star" in the first place.

2.8. Some people like to extend the conventions in 2.4 to give the disjunction connective precedence over the conjunction connective, and the conjunction connective precedence over the conditional one. In this way, $p \to q \wedge r \vee s$ would be interpreted as $p \to ((q \wedge r) \vee s)$. Although this conventions are widespread, we shall not use them in this book.

2.9 Biconditional connective. By this point, you should have noticed a crucial fact: the conditional connective, unlike the other ones we have studied, is not symmetric, which is to say that he truth value of $p \to q$ says nothing about that of its *converse* $q \to p$. Trust me, this is a crucial bit.

This very asymmetry leads to the definition of the symmetric *biconditional* connective, which, when applied on some variables p and q, yields a propositional form $p \leftrightarrow q$ that is defined to take the same truth value as $(p \to q) \wedge (q \to p)$ for the same assignments on the propositional variables p and q.

2.10 Implication and equivalence. If, for any propositional forms p and q, $p \to q$ is a tautology, it is said that p *implies* q; if, in addition, so is $q \to p$ (and, therefore, $p \leftrightarrow q$), then p and q are said *equivalent*. The important thing here is that, if p implies q, q is true whenever p is. Consequently, if p and q are equivalent, p is true if q is true and q is true if p is true. It is then needless to say that, when two propositional forms are equivalent, their truth tables are identical.

Just to have some examples, the propositional form $p \wedge (p \to q)$ implies q, and the propositional form $\neg(p \wedge q)$ is equivalent to $\neg p \vee \neg q$.

2.11. At this point, our language has gotten a little bit confusing, so let us introduce some new expressions to make it simpler. We already know that "if p then q" stands for $p \to q$. Nevertheless, based on what we know, $p \to q$ could also be read "p only if q". Take some time to think about this. Then, if we want to say "if p then q and if q then p", we could say "only if q then p and if q then p" or, in other words, "p if and only if q". There you have it! Saying that $p \leftrightarrow q$ is the same as saying "p if and only if q".

Just to finish with all this language overload, let me give you one more definition. If $p \to q$ is true, then q is said to be a necessary condition for p because, according to what we know, for $p \to q$ to hold, p cannot be true if q is false. Analogously, p is said to be a sufficient condition for q because if p is true, taking into account that $p \to q$ is true, we know, for sure, that q is true too. Thus, another fancy way of saying that $p \leftrightarrow q$ is stating that "p is a necessary and sufficient condition for q" or vice-versa.

2.12. Truth tables are not only used to describe the behaviour of logical connectives in propositional logic; they can also be used to define them. So much so that — as you probably had concluded on your own by now — there exists a perfect correspondence between truth tables and connectives. The reasons for this are obvious: every connective has its own truth table, every truth table can be used to define a connective, and any two connectives with the same truth table are equivalent.

Another way of introducing new connectives is defining them to be equivalent to some combination of known connectives. The way in which we defined the biconditional connective is a good example. The problem with this method is that, unlike with truth tables, we do not have a "correspondence" guaranteeing us that any connective can be expressed as a combination of others. Well, we did not have one...until now.

2.13 Theorem. Any propositional form p involving n propositional variables $p_1, \dots, p_n$ is equivalent to a propositional form q involving only those variables and the connectives $\wedge$, $\vee$ and $\neg$.

In particular, this shows that, given any n-ary connective $*$, the propositional form p resulting from its application on n distinct propositional variables can be written, equivalently, as a propositional form involving only the connectives $\vee$, $\wedge$ and $\neg$, which is to say that $*$ can be defined in terms of $\wedge$, $\vee$ and $\neg$.

Proof. The way we will prove this is by providing an effective algorithm for constructing the equivalent propositional form q from the truth table of the propositional form p. If p is a contradiction, it suffices to take q to be any contradiction such as $(p_1 \wedge \neg p_1) \vee \cdots \vee (p_n \wedge \neg p_n)$. If, on the other hand, there exists at least a particular assignment of truth values to $p_1, \dots, p_n$ which makes p true, we list all such assignments (mark them), and proceed as follows:

1. Set q to be an empty propositional form.
2. Find one marked assignment of truth values. Let $p_{i_1}, \dots, p_{i_r}$ with $\{i_1, \dots, i_r\} \subseteq \{1, \dots, n\}$ be all the elements that should be set to 1 in this particular assignment and, analogously, let $p_{j_1}, \dots, p_{j_s}$ with $\{j_1, \dots, j_s\} \subseteq \{1, \dots, n\}$ be all the elements that, in this assignment, need to be set to 0.
3. If q is not the empty propositional form, set it to
$$(q) \vee (p_{i_1} \wedge \cdots \wedge p_{i_r} \wedge \neg p_{j_1} \wedge \cdots \wedge \neg p_{j_s}).$$
Otherwise, if q is empty, set it to
$$p_{i_1} \wedge \cdots \wedge p_{i_r} \wedge \neg p_{j_1} \wedge \cdots \wedge \neg p_{j_s}.$$
4. Unmark the assignment we have been considering. If there are no marked assignments left, p is already equivalent to q, so we have finished. Otherwise, go back to step 2.

The reasoning we have followed is by all means valid and, with some thought on your part, should have already convinced you that what the result is true. Nevertheless, as we dive deeper into the world of mathematics, you will see that this same argument can be written in a much more elegant and clear manner. ■

2.14 Definition. An *adequate set of connectives* is a collection of connectives such that, for any propositional form involving a certain number of propositional variables, there exists an equivalent propositional form involving only those variables and the connectives in that set.

For example, $\{\wedge, \vee, \neg\}$ is an adequate set of connectives.

2.15 Corollary. The set $\{\neg, \rightarrow\}$ is an adequate set of connectives.

Proof. Since, according to 2.13, $\{\wedge, \vee, \neg\}$ is an adequate set of connectives, it suffices to show that, given any propositional variables p and q, each of $p \wedge q$ and $p \vee q$ is equivalent to a propositional form that only uses the negation and implication connectives.

It is easy to see that $p \wedge q$ is equivalent to $\neg(\mathrm{A} \rightarrow (\neg \mathrm{B}))$ and that $p \vee q$ is equivalent to $(\neg \mathrm{A}) \rightarrow \mathrm{B}$. ■

§3 Predicates

3.1. Until now, we have been concerned with propositions — statements that were either true or false — and how we could manipulate them with using logical connectives. We will now take one step further and take into consideration the particular objects which we make statements about. Welcome to the world of predicate logic.

3.2 Predicates. Just as propositional logic was all about propositions, predicate logic is all about *predicates*. An n-ary predicate $\mathrm{P}(x_1, \dots, x_n)$, where n is zero or a natural number, is a statement that depends on n variables $x_1, \dots, x_n$, known as *arguments*, that is either true or false depending on the values that those variables take. The set of values the variables are allowed to take is called the *domain of discourse* (or *domain*, for short). Of course, nullary predicates — which are mere propositions — need not be followed by parentheses since they do not take any arguments.

For example, if we were working with the set of all animals as domain, we could define a unary predicate $\mathrm{P}(x)$ as "x can fly" in such a way that P(cow) would be false but P(bird) would be true.

Notice how the proposition that arises after any assignment of values to the arguments of a predicate is written substituting the different variable symbols by the values they take, just as we did in P(cow) for $\mathrm{P}(x)$.

3.3. Predicates, just like propositions, can be combined with the connectives from propositional logic in order to form new predicates. Nonetheless, being able to work directly with the objects we are "saying things" about opens up a new world of possibilities. One of them is the ability to incorporate functions and constants into our predicates, but the most notable of all is the use of *quantifiers*.

As their name implies, quantifiers allow us to create new predicates stating the quantity of elements in the domain that verify a certain predicate. The two most fundamental are the *universal* quantifier and the *existential* quantifier. The universal quantifier is used to state that a certain predicate $\mathrm{P}(x_1, \dots, x_n)$ is true for every value taken by a variable x_i (with $i \in \{1, \dots, n\}$) in the domain. This is written as $(\forall x_i)\mathrm{P}(x_1, \dots, x_n)$ and is read as "for all x_i, $\mathrm{P}(x_1, \dots, x_n)$." On the other hand, we use the existential quantifier to state that a certain property[2] $\mathrm{P}(x_1, \dots, x_n)$ is verified by, at least, one assignment to x_i of

[2] "Property" can be used to mean "predicate."

an element in the domain. We write this as $(\exists x_i)\mathrm{P}(x_1, \dots, x_n)$ and read it as "there exists an x_i such that $\mathrm{P}(x_1, \dots, x_n)$." The predicate P over which we are quantifying is said to be the *scope* of the quantifier.

A variable x appearing in the scope of a quantifier $(\forall x)$ or $(\exists x)$ is said to be a *bound* variable. Variables that are not bound are called *free* variables. For example, if we let $\mathrm{P}(x)$ and $\mathrm{Q}(x)$ be predicates, in the predicate $\mathrm{P}(x) \land (\forall x)\mathrm{Q}(x)$, the first occurrence of x is free but the second is bound. Nevertheless, it should be taken into account that it is a pretty poor notational choice to use the same symbol to represent a free and a bound variable in the same predicate. This last predicate could have been better written as $\mathrm{Q}(y) \land (\forall x)\mathrm{Q}(x)$, but, from a purely formal point of view, there is nothing wrong with the original formulation.

3.4 Order matters. We are now ready to explore our first application of predicate logic; isn't that exciting!? And that application goes far beyond any mathematics that you will ever study: it is love.

Yesterday, I was scrolling through my social media feed and I found this statement: "No matter who you are, there exists someone who will love you". Just when I was going to click the "don't show me posts like this in the future" button, I had a brilliant idea: that was a perfect sentence waiting for us — you, my dear reader, and me — to analyse. So let us get to it!

It should be clear that, if we take the set of all human beings as our domain and let $\heartsuit(x, y)$ denote the predicate "y loves x", the statement "no matter who you are, love has already found someone for you" can be written as $(\forall x)(\exists y)\heartsuit(x, y)$. Now, I have a simple question: what would happen if you changed the order of the quantifiers and wrote $(\exists y)(\forall x)\heartsuit(x, y)$? What does this sentence mean? It means that there exists an individual who we all happen to be in love with! The first sentence read "for every person x, there exists a person y who will love them", but now it reads "there exists a person x who is loved by every person y". Do you have the slightest idea of the mess you have made by swapping two quantifiers? You have taken a cheesy sentence and transformed its meaning to postulate the existence of a love monster!

The moral of the story is simple: do not swap quantifiers. Nevertheless, if the same quantification is being used on two different variables consecutively, such as in $(\exists x)(\exists y)\heartsuit(x, y)$, there is obviously no harm in swapping the order of quantifiers; in fact, that last sentence would be often written as $(\exists x, y)\heartsuit(x, y)$.

Hey, I know you might be angry at me for having created hype with applications and all that stuff and having only given you a cheesy sentence. What can I say? I needed your attention! I hope you will forgive me.

3.5 Notation. There is a little bit of freedom in the way that quantifiers can be written and each alternative has its own advantages. The following are just some examples of the notation that one can find in the literature:

$$\forall x, \exists y, \mathrm{P}(x, y), \qquad (\forall x)(\exists y) : \mathrm{P}(x, y), \qquad \forall x : \exists y : \mathrm{P}(x, y).$$

I see myself as a liberal person when it comes to notation, but there is something I beg you to do: do not EVER write quantifiers after their scope. Every time you write $\mathrm{P}(x), \forall x$ a cute kitten cries immersed in sadness, so, please do not do that. This is not a matter of taste; it is a matter of readability. When you are about to read an expression, what variables are being quantified is the first thing you should know, not the last!

Furthermore, as we have just seen, the order in which quantifiers are set is very, very important; thus, if you write them at the end…what order are they in? Should they be read from left to right or from right to left? I mean, it is just inelegant and clumsy. Do not do that, please.

If you really want to do things right, you should extend this idea to your writing. I know; saying, for instance, “$\mathrm{P}(x)$ is true for all x” sounds natural and harmless, but as the number of quantifiers increases — and, believe me, it will increase and not just in logic, but in ordinary mathematics — you better have those quantifiers right at the start. As a rule of thumb, I would only put a quantifier at the end of a predicate in written text if it is a single universal quantifier, and I would never ever write a quantifier at the end in a symbolic expression.

Now that we are dealing with notation, let me draw your attention to an important issue. In 2.4, we said that, in propositional logic, there was no harm in getting rid of the parentheses that surround a full proposition, so there was no need to write $(p\wedge q)$ and we could just write $p \wedge q$. In predicate logic, however, we need to be somewhat careful with this idea. If, for instance, I wrote $(\forall x)\mathrm{P}(x) \wedge \mathrm{Q}(x)$, I would not mean the same as if I had written $(\forall x)(\mathrm{P}(x) \wedge \mathrm{Q}(x))$. In the former case, the scope of the quantifier is $\mathrm{P}(x)$, whereas in the latter it is $\mathrm{P}(x) \wedge \mathrm{Q}(x)$.

In order to save ourselves some parentheses and make everything a little bit cleaner, we will use the following convention. If a sequence of quantifiers is followed by a dot, it will mean that its scope is the remainder of the predicate unless an opening parenthesis is present before the quantifiers; if this happens, the scope will end at the point where the matching closing parenthesis is located. Thus, in $(\forall x)(\exists y).\ \mathrm{P}(x) \wedge \mathrm{Q}(y)$, the scope of the first quantifier is $(\exists y)(\mathrm{P}(x) \wedge \mathrm{Q}(y)$ and that of the second is $\mathrm{P}(x) \wedge \mathrm{Q}(y)$. On the other hand, in $((\forall x).\ \mathrm{P}(x)) \wedge \mathrm{Q}(x)$, the scope of the quantifier is $\mathrm{P}(x)$.

3.6 Negating predicates. If I told you that every single human loves mathematics, what would you have to do to prove me wrong? You would simply need to show the existence of a person who does not like mathematics. Therefore, the negation of the sentence “for all x, $\mathrm{P}(x)$ holds” is “there exists an x such that $\mathrm{P}(x)$ does not hold.” In other words, $\neg(\forall x)\mathrm{P}(x)$ is the same as $(\exists x)(\neg\mathrm{P}(x))$.

Now, let us just say that I postulate the existence of a person who has the superpower of flying. If you wanted to show that my statement is false, you would have to prove that no person has the superpower of flying or, equivalently, that “for every person x, x cannot fly”. What is the moral of the story? That $\neg(\exists x)\mathrm{P}(x)$ equivalent to $(\forall x)(\neg\mathrm{P}(x))$.

Putting together all that we have learnt: how would you write the predicate $\neg(\forall x)(\exists y)\mathrm{P}(x,y)$ without having a negation connective before any quantifier? Take a sheet of paper and write down the result.

[Spoiler alert] If you have understood this part, you should have reasoned as follows: $\neg(\forall x)(\exists y)\mathrm{P}(x,y)$ is the same as $(\exists x)(\neg(\exists y)\mathrm{P}(x,y))$, which is equivalent to $(\exists x)(\forall y)\neg\mathrm{P}(x,y)$. If you got this right, congrats! You are on the right track. If you did not, do not worry: make yourself a good cup of tea, go through this material again and give it some thinking.

As you can see, when negating an expression involving the existential and universal quantifiers, the only thing we need to do is “swap them and negate what is inside them". That is a pretty easy rule, but, as always happens with this kind of shortcuts, you should only apply it if you really know what is going on underneath the hood.

3.7 Defining quantifiers. There is something kind of significant in our analysis of the negation of predicates. What we have shown — probably without your noticing — is that one of our quantifiers is redundant. The predicate $(\exists x)\mathrm{P}(x)$ is equivalent to $\neg(\forall x)(\neg\mathrm{P}(x))$, so, in a way, there was no need to introduce the existential quantifier once we had the universal one. Conversely, $(\forall x)\mathrm{P}(x)$ is equivalent to $\neg(\exists x)(\neg\mathrm{P}(x))$, hence the existential quantifier could also serve us as our only quantifier.

Despite the redundancy, we introduced both of them for an obvious reason: for us humans, it is more natural to think "every x verifies $\mathrm{P}(x)$" than "there does not exits an x not verifying $\mathrm{P}(x)$".

3.8 Pseudo-quantifiers. We will now introduce some "pseudo quantifiers". I have given them this name — which is, by the way, not standard whatsoever — because they are just constructions that help us quantify over things without being proper quantifiers. Instead, they are mere logical artefacts that are limited to a certain kind of theories.

Some theories (in fact, most theories) define a predicate E that is meant to represent equality. In these theories, we can define a predicate $(\exists! x)\mathrm{P}(x)$ meaning "there exists a unique x verifying $\mathrm{P}(x)$". The equivalent real predicate behind $(\exists! x)\mathrm{P}(x)$ would be

$$(\exists x)(\forall y)(\mathrm{P}(x) \land (\mathrm{P}(y) \to \mathrm{E}(x,y))),$$

or, in English, "there exists an x verifying $\mathrm{P}(x)$ and such that, if any y verifies $\mathrm{P}(y)$, then y is equal to x".

On some occasions, one may wish to quantify only over the set of elements in the domain that verify a certain predicate P. This is very easy to do. If we wanted to restrict the quantification of $(\forall x)\mathrm{Q}(x)$ or $(\exists x)\mathrm{Q}(x)$ only to the elements x verifying $\mathrm{P}(x)$, we would just need

$$(\forall x)(\mathrm{P}(x) \to \mathrm{Q}(x)) \quad \text{and} \quad (\exists x)(\mathrm{P}(x) \land \mathrm{Q}(x))$$

respectively. You see, saying "for all x verifying $\mathrm{P}(x)$, $\mathrm{Q}(x)$ is true" is the same as saying "for all x, if x verifies $\mathrm{P}(x)$, then $\mathrm{Q}(x)$ is true", and analogously for the existential quantifier.

Many theories include binary predicates $\mathrm{P}(x,y)$ that can be written as $x\mathrm{P}y$ — for instance, the inclusion predicate $\in$ that we studied in our elementary treatment of set theory, — and, given a fixed y, one may wish to quantify over all the elements x verifying $x\mathrm{P}y$. Of course, one could write $(\forall x)(x\mathrm{P}y \to \mathrm{Q}(x))$ or $(\exists x)(x\mathrm{P}y \land \mathrm{Q}(x))$, but, instead of wasting time and ink with such lengthy expressions, it is common to simply use $(\forall x\mathrm{P}y)\mathrm{Q}(x)$ and $(\exists x\mathrm{P}y)\mathrm{Q}(x)$. Thus, if we wanted to postulate the existence of an element X inside a set Y verifying a property $\mathrm{P}(x)$, we would simply have to write $(\exists \mathrm{X} \in \mathrm{Y})\mathrm{P}(\mathrm{X})$. It is important to keep in mind that this is just shorthand notation, and we should always understand what is really going on.

Just to finish with this, I have an innocent question for you: let us assume that we are in the context of set theory and, given a set X and a predicate $\mathrm{P}(x)$, we formulate the sentence $(\exists! x \in \mathrm{X})\mathrm{P}(x)$. Is this expression well-formed and unambiguous? Take your time to think about it.

[Spoiler alert] Turns out that this expression is ambiguous. What do you mean: that there exists an element $x \in X$ verifying $P(x)$ that is unique among all the elements, or unique among all the elements in X? As far as I know, there is no widespread convention on which of these two possible interpretations is correct, so it is better not to use this construction to avoid ambiguity. If you wanted to say that it is unique among all the elements of X, you could write $(\exists! x)(x \in X \wedge P(x))$. If, on the other hand, you meant that it is unique amongst all the elements, you could write $(\exists x \in X)(P(x) \wedge (\forall y)(P(y) \to y = x))$.

3.9 Higher-order logic. Predicate logic is also known as *first-order logic*. Higher-order logic is an extension of first-order logic that not only allows quantification over variables, but also over predicates about variables, over predicates about predicates about variables, and so on.

For example, the statement "for every property P, there exists an element x such that $P(x)$ is true" would be a statement in second-order logic.

3.10 Mathematical induction. And now, let us close this chapter with a fundamental tool that we will be using extensively from now on, the principle of mathematical induction. What this principle states is the following: if a subset A of the natural numbers contains 1 and, for every $n \in A$, $n + 1 \in A$, then A is the set of natural numbers. This is kind of trivial when you think about it. We already know that $A \subseteq \mathbb{N}$ by hypothesis, so, in order to prove that $A = \mathbb{N}$, we just need to show that $\mathbb{N} \subseteq A$. Thus, let $n \in \mathbb{N}$ and let us prove that $n \in A$. We know, by hypothesis that $1 \in A$ and — since, for every $n \in A$, $n + 1 \in A$, — then $1 + 1 \in A$ and, therefore, $1 + 1 + 1 \in A$. If we apply this reasoning n times, we are led to

$$n = 1 + \overset{n}{\cdots} + 1 \in A.$$

And how does this relate to predicates? Well, let us say that we want to show that a predicate $P(n)$ is true for all natural numbers n. Let A be the set of natural numbers m such that $P(m)$ is true. If we show that $1 \in A$ and that, given any $m \in A$, $m + 1 \in A$, we will have shown that $A = \mathbb{N}$. In other words, if we show that $P(1)$ is true and that, assuming $P(n)$ to be true, $P(n + 1)$ is true, then we will know that the predicate is true for all natural numbers.

This principle can be extended to what is known as *strong induction* (we will refer to it as induction too). If, given a subset $A \subseteq \mathbb{N}$, we know that $1 \in A$ and that, assuming every natural number i such that $0 \leq i \leq n$ to belong to A, we have $n+1 \in A$, then $A = \mathbb{N}$. The reasoning that justifies this is essentially the same as that for normal induction. Of course, strong induction can also be applied to show that a certain property holds for every natural number.

Exercises

1) The exclusive-or connective acts on two propositional variables producing a propositional form $p \otimes q$ that is equivalent to $(p \wedge \neg q) \vee (\neg p \wedge q)$. Write down the truth table of this connective.

2) In a parallel universe, all mathematicians have assembled and formed an organization known as the Chalk Party. One member of the Chalk Party is being judged for allegedly having stolen chalk from the University of Oviedo.

When the prosecutor begins his exposition, he says: “This man is clearly guilty. He is a member of the chalk party, and all chalk thieves belong to this organization”. The judge is quick to dismiss his argument as a fallacy. Why? Justify your answer using propositional logic.

3) Prove that the empty set is a subset of any set A.
Hint: Show that, if $x \in \emptyset$, then $x \in \mathrm{A}$.

4) Generally, any proof relying on an algorithm can be easily transformed into a proof by induction. Rewrite the proof of 2.13 as a proof by induction.

5) Find an equivalent expression for each of the following predicates in which any negation connectives are acting directly on P, Q or R.

a) $\neg(\forall x)(\exists y)(\forall z)((\mathrm{P}(x) \wedge \mathrm{Q}(y)) \vee (\mathrm{R}(z)).$
b) $\neg(\exists x)(\forall y)(\forall z)(\mathrm{P}(x) \to (\mathrm{Q}(y) \wedge \mathrm{R}(z))).$

6) Use induction to prove that, given a natural number n and some propositional variables $p, q_1, \dots, q_n$, the following propositional forms are tautologies.

a) $p \wedge (q_1 \vee \cdots \vee q_n) \leftrightarrow (p \vee q_1) \wedge \cdots \wedge (p \vee q_n).$
b) $p \vee (q_1 \wedge \cdots \wedge q_n) \leftrightarrow (p \wedge q_1) \vee \cdots \vee (p \wedge q_n).$
c) $\neg(q_1 \vee \cdots \vee q_n) \leftrightarrow \neg q_1 \wedge \cdots \wedge \neg q_n.$
d) $\neg(q_1 \wedge \cdots \wedge q_n) \leftrightarrow \neg q_1 \vee \cdots \vee \neg q_n.$

7) Use induction and double inclusion to prove that, given a natural number n and some sets $\mathrm{A}, \mathrm{B}_1, \dots, \mathrm{B}_n$, we have

a) $\mathrm{A} \cap (\mathrm{B}_1 \cup \cdots \cup \mathrm{B}_n) = (\mathrm{A} \cup \mathrm{B}_1) \cap \cdots \cap (\mathrm{A} \cup \mathrm{B}_n),$
b) $\mathrm{A} \cup (\mathrm{B}_1 \cap \cdots \cap \mathrm{B}_n) = (\mathrm{A} \cap \mathrm{B}_1) \cup \cdots \cup (\mathrm{A} \cap \mathrm{B}_n).$

Let $\mathrm{A}_1, \dots, \mathrm{A}_n$ be subsets of a set X. Show that

c) $\mathrm{X} \setminus (\mathrm{A}_1 \cup \cdots \cup \mathrm{A}_n) = (\mathrm{X} \setminus \mathrm{A}_1) \cap \cdots \cap (\mathrm{X} \setminus \mathrm{A}_n),$
d) $\mathrm{X} \setminus (\mathrm{A}_1 \cap \cdots \cap \mathrm{A}_n) = (\mathrm{X} \setminus \mathrm{A}_1) \cup \cdots \cup (\mathrm{X} \setminus \mathrm{A}_n).$

Chapter I

The foundations of mathematics

§1 The axiomatic method. Syntax and semantics

1.1 The breakdown of logicism. In the early days of set theory, logicism was on the rise. It was by then believed — and rightly so — that set theory could be used as a foundation for all mathematics, so, if set theory could be reduced to logic, mathematics would be reducible to logic too. If you allow me to oversimplify a bit, the idea behind this endeavour was the following. How do you define a set? Well, they said, just take any predicate $\varphi(x)$ and define the set $\mathrm{R} = \{x \mid \varphi(x)\}$ of all elements x verifying φ. Thus, for instance, we could take $\varphi(x)$ to be the predicate "x is a natural number", and define the set of natural numbers as the set of elements verifying φ.

It was all good and great until, one day, Bertrand Russell found a paradox that threatened to destroy set theory and the very foundations of mathematics. Let us consider the set $\mathrm{R} = \{x \mid x \notin x\}$. By the definition of R, we would have $\mathrm{R} \in \mathrm{R}$ if and only if $\mathrm{R} \notin \mathrm{R}$. Boom! Many more paradoxes emerged that sentenced the logicist approach to set theory to death. What these paradoxes showed was that logic was not enough to define mathematics: mathematics can of course be built on top of logic, but it needs its own formal structure.

Nevertheless, the idea of identifying predicates with sets stayed alive, but not in such a wild state as it originally did. In general, there is no harm in identifying predicates with sets…provided you do it with care. What does it mean to do it carefully? Be patient, you will get a proper answer later.

If you are interested in the events that followed and preceded Russell's paradox and would like to get some more context, I invite you to read chapter II.7 of [2].

1.2. Mathematics is all about proofs. When you are doing mathematics, you are working with certain abstract objects and your job is to deduce (that is, prove) some properties about them. From an informal perspective, a proof could be defined as an argument that would convince any intelligent being that something is true; from a formal perspective, as we will soon see, things are a little bit more subtle.

If mathematics is all about proofs and proofs are all about deduction, we need an starting point. Thus, when we are working with a mathematical object, we need to understand its defining properties: the properties that are so essential and intrinsic to that object, that not only are unquestionable, but can also be used to define the object itself. Some of these defining properties properties are chosen to be *axioms*, and they are the starting point for any proof concerning the mathematical entities they define. That is what the axiomatic method is all about.

For example, let us consider the natural numbers. The statement “if $x = y$, then $x + 1 = y + 1$” could well be an axiom of the natural numbers. It is unquestionably true according to our concept of natural number, and it can be used to define what a natural number is.

Now that we have all the intuition behind how mathematical reasoning should work, we will seek to formalise everything that we have done so far. Firstly, we aim to provide an axiomatic framework for logic itself.

1.3. Before diving into the world of mathematical logic, I would like to give you some vocabulary. In mathematics, we use the words *proposition*, *lemma*, *theorem* and *corollary* when we are referring to proven statements (also known as *results*) in a particular theory. Specifically:

- A proposition is an ordinary result (and should not be confused with the “propositions” of propositional logic that we studied before),
- a lemma is an auxiliary result that is only used as a means of proving something more important,
- a theorem is an important result,
- and a corollary is an immediate consequence of another result.

By the way, when a mathematician says that something is “immediate” from something, it means that it can be somewhat mechanically deduced without using any extravagant ideas. If something is simply obvious, it is said to be trivial. Lastly, if something is somewhere in between trivial and immediate, it is said to be “direct”.

We will now be working not within any theory, but analysing theories themselves. These theories will be called *object theories* and, in rigour, we should refer to the results we obtain about those theories as *metatheorems* because they will not be results in the theory itself, but in the *metatheory* that we will be using to analyse our object theory. Nevertheless, being aware of this fact, we will still use the usual terms (proposition, theorem…) to refer to the different metatheorems.

1.4. The first notion that we will formalise is that of a language. I think we can both agree that any language needs two things: a collection of symbols and a collection of accepted constructions with those symbols. For instance, the English language has a collection of symbols (the letters of the alphabet, empty space and punctuation marks) and a collection of appropriate combinations of those symbols (the collection of all sentences).

In the context of formal languages, we can take this classification one step further. Among the allowed constructions with symbols, we can make a distinction between terms and formulas: a term would be a construction representing an object, whereas a formula would be a construction representing a well-formed statement about terms. For example, if we were to define the language or arithmetic, we should define it in such a way that $x + 1$ were a term and $x + 1 = 0$ were a formula.

In an attempt to further connect these notions with human languages, we could identify terms with nominal phrases and formulas with sentences.

1.5 Definition. Given a set of symbols Σ, the set of words Σ^* over Σ is defined as the set of finite sequences of symbols of Σ. Given any sequence of elements $a_1, a_2, \ldots, a_n$ of Σ, their corresponding element in Σ^* is represented by concatenation: $a_1 a_2 \cdots a_n \in \Sigma^*$.

A *formal language* — or *language*, for short — is a set of symbols Σ together with three subsets $\mathrm{T}, \mathrm{F} \subseteq \Sigma^*$ and $\Lambda \subseteq \mathrm{F}$. The set Σ is said to be the *alphabet* of the language, and T, F and Λ are said to be the sets of *terms*, *well-formed formulas* (or just *formulas*) and *sentences* of the language. Such a formal language is represented by the tuple $(\Sigma, \mathrm{T}, \mathrm{F}, \Lambda)$.

For must practical purposes, the set of sentences is irrelevant. Thus, we might sometimes refer to a language $(\Sigma, \mathrm{T}, \mathrm{F}, \Lambda)$ simply as $(\Sigma, \mathrm{T}, \mathrm{F})$ if there is no need to know what its sentences are.

1.6. So that was our first formal definition. It should have been clear enough, but there is a little thing I would like to clarify. In mathematics, when we define an object that consists of several pieces, we often represent it with a tuple. It is just a fancy way of representing things, do not try to find in it any sort of metaphysical nature.

As unnecessary as this may seem, it is done for a good reason. Imagine that we are defining a formal language and we have already constructed the alphabet A, the set of terms B, the set of formulas C and the set of sentences D. Instead of writing "the language having A as alphabet, B as set of terms, C as set of formulas and D as set of sentences", we can just say "the language $(\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D})$". It is more clean and convenient.

1.7 Example. The first formal language that we will define is the formal language $\mathrm{L_P}$ of propositional logic. The alphabet of this language will consist of an infinite collection of symbols $p_1, p_2, \ldots$ meant to represent propositional variables, of parentheses, and of the → and ¬ connectives. We only include these connectives because, as we showed in 0-2.15, they suffice to construct any propositional form.

Theoretically, we could add more connectives to our language, but that would simply be redundant. The purpose of formal languages is to provide a precise framework in which to analyse languages theoretically and, therefore, they should be as simple and minimal as possible. In our work as mathematicians, we can, of course, add as many connectives as we want — regarding them as aliases for their equivalent constructions involving the → and ¬ connectives.

Thus, the alphabet of the language of propositional logic will be

$$\Sigma_\mathrm{P} = \{(,), \to, \neg, p_1, p_2, \ldots, p_n, \ldots\}.$$

The set of terms $\mathrm{T_P}$ of will simply be the set of words representing propositional variables: $\mathrm{T_P} = \{p_1, p_2, \ldots\} \subseteq \Sigma_\mathrm{P}^*$. The terms of $\mathrm{L_P}$ will be known as *propositional symbols.* Finally, the set of formulas $\mathrm{F_P}$ is defined following what is known as an *inductive procedure*:

1. All terms are well-formed formulas: $\mathrm{T_P} \subseteq \mathrm{F_P}$.
2. If A and B are well-formed formulas, so are $(\neg \mathrm{A})$ and $(\mathrm{A} \to \mathrm{B})$.

Just to better understand what is going on here, how could we prove that $((\neg p_1) \to p_2)$ is a formula of propositional logic? By the first rule, we know that both p_1 and p_2 are formulas. By the second one, we know that $(\neg p_1)$ is also a formula; therefore, as both $(\neg p_1)$ and p_2 are formulas, if we apply the second rule again, we know $((\neg p_1) \to p_2)$ to be a formula.

Just to conclude the definition of $\mathrm{L_P}$, we will take any formula of this language to be a sentence.

The language of propositional logic is then $\mathrm{L_P} = (\Sigma_\mathrm{P}, \mathrm{T_P}, \mathrm{F_P}, \mathrm{F_P})$. It is easy to see that the formulas of this language are, precisely, the propositional forms on the propositional variables $p_1, p_2, \ldots$. That is why we will refer to the formulas in $\mathrm{F_P}$ as *propositional forms* too.

Informally, for any formulas $\mathrm{A}, \mathrm{B} \in \mathrm{F_P}$, we will define the formula $(\mathrm{A} \wedge \mathrm{B})$ as an abbreviation of $(\neg(\mathrm{A} \to (\neg\mathrm{B})))$, the formula $\mathrm{A} \vee \mathrm{B}$ as an abbreviation of $((\neg\mathrm{A}) \to \mathrm{B})$, and the formula $(\mathrm{A} \leftrightarrow \mathrm{B})$ as an abbreviation of $((\mathrm{A} \to \mathrm{B}) \wedge (\mathrm{B} \to \mathrm{A}))$.

1.8 Definition. Let S be a collection of symbols together with a *signature* function $\sigma : \mathrm{S} \longrightarrow \mathbb{Z}$. For any $s \in \mathrm{S}$, we say that s is an *n-ary function symbol* if $\sigma(s) = n > 0$; we say that s is an *n-ary predicate symbol* if $\sigma(s) = -n < 0$, and we say that s is a *constant symbol* if $\sigma(s) = 0$. Let $\{x_1, \ldots, x_n, \ldots\}$ be an arbitrary collection of symbols such that, for any $i \in \mathbb{N}$, $x_i \notin \mathrm{S}$. The alphabet $\Sigma_{(\mathrm{S},\sigma)}$ associated to (S, σ) is

$$\Sigma = \mathrm{S} \cup \{\neg, \to, \forall, (,)\} \cup \{,\} \cup \{x_1, x_2, \ldots, x_n, \ldots\}.$$

The set of terms $\mathrm{T}_{(\mathrm{S},\sigma)}$ is defined inductively according to the following rules.

1. All the elements of the set $\{x_1, x_2, \ldots, x_n, \ldots\}$ of *variables* are terms.
2. All constant symbols are terms.
3. Given any n-ary function symbol f and given any n terms $t_1, \ldots, t_n$, the construction $f(t_1, \ldots, t_n)$ is a term.

The set of formulas $\mathrm{F}_{(\mathrm{S},\sigma)}$ is defined inductively according to:

1. Given any n-ary predicate symbol P and any n terms $t_1, \ldots, t_n$, the *atomic formula* $\mathrm{P}(t_1, \ldots, t_n)$ is a formula.
2. If A and B are formulas and $i \in \mathbb{N}$, then $(\neg\mathrm{A})$, $(\mathrm{A} \to \mathrm{B})$ and $(\forall x_i)\mathrm{A}$ are formulas.

As in the informal treatment of predicate logic, given any formula of the form $(\forall x_i)\mathrm{A}$, the formula A is said to be the *scope* of the quantifier $(\forall x_i)$. Moreover, any occurrence of a variable x_i within the scope of a quantifier $(\forall x_i)$ is said to be *bound.* If an appearance of a variable is not bound, we say that it is *free.*

The set of sentences $\Lambda_{(\mathrm{S},\sigma)}$ is the set of all the *closed formulas*: the collection of all formulas having no free occurrences of variables.

The language $\mathrm{L}_{(\mathrm{S},\Sigma)} = (\Sigma_{(\mathrm{S},\sigma)}, \mathrm{T}_{(\mathrm{S},\sigma)}, \mathrm{F}_{(\mathrm{S},\sigma)}, \Lambda_{(\mathrm{S},\sigma)})$ is said to be the *first-order language* associated to (S, σ).

Just as we did in the language of propositional logic, given any formulas A and B in any first order language, we can define the formula $(\mathrm{A} \wedge \mathrm{B})$ as an abbreviation of $(\neg(\mathrm{A} \to (\neg\mathrm{B})))$, the formula $\mathrm{A} \vee \mathrm{B}$ as an abbreviation of $((\neg\mathrm{A}) \to \mathrm{B})$, and the formula $(\mathrm{A} \leftrightarrow \mathrm{B})$ as an abbreviation of $((\mathrm{A} \to \mathrm{B}) \wedge (\mathrm{B} \to \mathrm{A})$. In addition, for any variable x_i, we will define $(\exists x_i)\mathrm{A}$ to be an alias for $(\neg(\forall x_i)(\neg\mathrm{A}))$.

1.9. First-order languages will make it easy to define the language of any theory built in the framework of first-order logic. If you have perfectly understood the definition I have given you, that's wonderful! Just move on. Otherwise, if you feel a little bit overwhelmed and are wondering why on earth I have just done this to you…let us have a small chat here. One of the most important skills you will learn in your journey through the wonders of mathematics is being able to see beyond formal concepts. I have given you a purely formal definition, and now you need to work on it to get the intuition behind it. This might be the first time you are facing this kind of challenge, so let me guide you through.

The idea behind all of this is simple. The alphabet of any first-order language always has an infinite amount of symbols representing variables ($x_1, \dots, x_n, \dots$), two connectives ($\neg$ and $\rightarrow$), one quantifier ($\forall$) and some punctuation symbols (parentheses and a comma). In addition, it may have some constants, some symbols representing predicates and some symbols representing functions. These last objects are what can make first-order languages distinct, and they are encoded in S and σ. Instead of using a set for the constants, a set for the predicate symbols and another set for the function symbols, I thought it would be better to put them all in one set and use the σ function to distinguish them. Why? Because we were going to need such a function anyway to specify the arity of each of the symbols for, as you have seen, knowing the arity of a function or predicate symbol is necessary when defining how terms and formulas can be constructed. So, instead of having three sets and two arity functions (one for the set of predicates and another one for the set of function symbols), why not use just one set and a "signature" function? A signature of zero just means that symbol is a constant, a positive signature that a symbol is a function, and a negative signature that it is a predicate. Moreover, the absolute value indicates the arity. It is all more compact and simple to encode.

1.10 Example. (i) Any first-order language can be used as a language for predicate logic. Nonetheless, we will do it for the most general first-order language available: one that contains an infinite amount of constants, predicates and functions. Our first-order language L_Q will be the one associated to the set S of all the symbols of the form

$$\mathrm{P}_i^j, \quad f_i^j, \quad c_i$$

with $i, j \in \mathbb{N}$, and to the signature function σ defined, for every $i, j \in \mathbb{N}$ as

$$\sigma\left(\mathrm{P}_i^j\right) = -j, \qquad \sigma\left(f_i^j\right) = j, \qquad \sigma(c_i) = 0.$$

(ii) The language of first-order arithmetic is the first-order language associated to the set $\mathrm{S} = \{=, 1, s, +, \cdot\}$ and the signature $\sigma(=) = -2$, $\sigma(1) = 0$, $\sigma(s) = 1$, $\sigma(+) = 2$ and $\sigma(\cdot) = 2$. Instead of witting $=(x, y)$, $+(x, y)$ and $\cdot(x, y)$, we will use *infix notation* and write $x = y$, $x + y$ and $x \cdot y$ respectively. Furthermore, $x \cdot y$ will often be abbreviated as xy.

The symbol $=$ is meant to represent equality, $+$ is meant for addition, $\cdot$ for multiplication, 1 is meant for the number one, and s for the successor function. I suppose that wasn't very surprising.

In order to simplify the notation and reduce the number of parentheses, by convention, $\cdot$ will have higher precedence than $+$ in the creation of terms. Thus $x + y \cdot z$ will be parsed as $x + (y \cdot z)$ and not as $(x + y) \cdot z$.

1.11 Syntax and semantics. With formal languages, we now have a formal device enabling us to precisely express statements about a particular theory. What remains to be done is providing mechanical rules for transforming formulas in order to enable deductions, and providing ways of deciding the truth or falsity of such formulas and the validity of the methods of deduction. These two different tasks are tackled by syntax and semantics respectively.

As we will soon see, the central object in syntax is that of a *formal system*: a framework in which we can use a deductive machinery to operate with the formulas of a language.

The fundamental tools in semantics are *interpretations*. An interpretation is nothing more than an assignment of meaning to the elements of a language. Thus, once a formula is considered under a particular interpretation, we are able to decide its truth or falsity.

The informal analysis of propositional and predicate logic that we have made has been purely semantic. When we studied propositional logic, we developed a complete and systematic understanding of the truth of propositional forms for each possible assignment of truth values to its variables; as we will soon formalise, those assignments of truth values are the possible interpretations of propositional logic. In our study of predicate logic, we learned to identify the truth or falsity of formulas in a certain domain of discourse; following the same ideas, an interpretation of any first-order language will be nothing more than a choice of a domain of discourse and an assignment of meaning to the symbols of the language under that domain.

1.12 Definition. A *formal system* S is a language $L = (\Sigma_L, T_L, F_L)$ together with a subset $\Xi \subseteq F_L$ of formulas known as its set of *axioms* and a set R of *rules of inference*. We write $\mathsf{S} = (L, \Xi, R)$.

We say that a formula $P \in F_L$ can be *deduced* from a set of formulas $\Gamma \subseteq F_L$ in the formal system S if there exists a sequence of formulas $P_1, \dots, P_n$ with $P_n = P$ such that, for every $i \in \{1, \dots, n\}$, P_i belongs to Ξ or Γ, or has been obtained by applying a rule of inference on previous formulas of the sequence. If $\Gamma = \emptyset$, P is said to be a *theorem* of S and the sequence $P_1, \dots, P_n$ is said to be a *proof* of P in S.

If P can be deduced from Γ, we write $\Gamma \vdash_{\mathsf{S}} P$. Furthermore, if $\Gamma = \{H_1, \dots, H_m\}$, we can also write $H_1, \dots, H_m \vdash_{\mathsf{S}} P$. We can denote the fact that P is a theorem of S by $\vdash_{\mathsf{S}} P$ or by $\mathsf{S} \vdash P$. In addition, if the context makes it clear that the deduction or proof is taking place in S, the symbol $\vdash_{\mathsf{S}}$ can be safely replaced by $\vdash$.

1.13. I would like to make a small remark about the definitions I have just given you. What we have done is a mere formalisation. Thus, if we are ever proving or deducing something in any particular formal system, we will probably not find ourselves just applying inference rules in such a mechanical way as that described in the formal definition of proofs.

Instead, in practice, we will use some methods that will enable us to reason in a more intuitive and human-like manner, but always with the assurance that a “formal proof” like the one we have defined can be constructed from that reasoning — even if that would involve an unfeasible amount of mechanical work.

1.14 Notation. Instead of writing “$P_1, \dots, P_n$ with $P_n = P$” as we did in the previous definition, it is customary and convenient to write $P_1, \dots, P_n = P$.

This is quite a bit more than something specific of this context. In fact, one should always be flexible in the way they read notation. For instance, if I wanted to say “this applies to the set A that is a subset of B”, I could just say “this applies to $A \subseteq B$”.

1.15 Proposition. Let FS be a formal system on a language $L = (\Sigma, T, F)$. If $\Gamma, \Gamma' \subseteq F$ are collections of formulas and A and B are two arbitrary formulas, then:

(i) If $\Gamma \vdash A$, then $\Gamma \cup \{B\} \vdash A$.
(ii) If $\Gamma \vdash A$ and $\Gamma' \cup \{A\} \vdash B$, then $\Gamma \cup \Gamma' \vdash B$.
(iii) If $\Gamma \vdash A$, then $\Gamma \vdash B$ if and only if $\Gamma \cup \{A\} \vdash B$.

1.16 Definition. Given a language $L = (\Sigma, T, F)$, an *interpretation* of L is an assignment of meaning to the elements of L. The elements of L are nothing more than a bunch of meaningless symbols. Under a certain interpretation, however, those symbols are given an abstract reference. For example, in the formal language of arithmetic, the symbol 1 is nothing more than a meaningless character, but we can construct an interpretation in which 1 represents "the number one". The way in which these interpretations can be constructed is different for each of the two "flavours" of logic that we have studied and will be described, with full precision, later.

An interpretation of L will define two subsets of F: a subset of *true* formulas, and a subset of *false* formulas; we denote that a formula $A \in F$ is true in an interpretation I by $I \vDash A$. Bear in mind, though, that there might be formulas that are neither true or false.

A formula $A \in F$ is said to be *valid* in L if it is true in any interpretation. That is denoted by $\vDash A$.

Given a formula $B \in F$ and a collection of formulas $\Gamma \subseteq F$, if whenever an interpretation I satisfies $I \vDash A$ for every $A \in \Gamma$, we have $I \vDash B$, we say that B is a *semantic consequence* of Γ or that Γ *semantically entails* B. In this case, we write $\Gamma \vDash B$. If $\Gamma = \{A_1, \dots, A_n\}$, we may also write $A_1, \dots, A_n \vDash B$. Notice how $\varnothing \vDash B$ is the same as $\vDash B$.

1.17 Why bother?. At this point, one question might be popping up in your mind. You should by now have a fully developed understanding of propositional and predicate logic — an understanding that we have labelled as "semantic". why isn't that everything you need? Why bother with the axiomatic method and with syntax altogether?

While what you already know is, probably, everything you will ever need to know about logic in your future career as a mathematician, syntactical approaches in general and the axiomatic method in particular have a purpose.

You see, logic is very intuitive, mechanical and simple; so much so that it sometimes seems just like a linguistic artefact. This will not be the case with mathematics. When arguing about the truth or falsity of, let us say, a proposition (in propositional logic), we have systematic semantic methods that enable us to do so. In mathematical theories, such a thing does not exist.

How would you show me that there exists an infinite amount of prime numbers in a semantic way? You cannot — at least unless you are able to write down an infinite list of prime numbers to prove it. While it is semantically clear that there either is or there is not an infinite amount of prime numbers, there is no way to decide semantically. We need to have some structure enabling deduction: that is why we need syntax. The purpose of a formal system is capturing the "essence" of the entities it aims to add syntax to. Syntax aims to model an abstract reality in such a way that we can work in it whereas semantics is concerned with the abstract realities themselves.

Of course, one should expect syntax and semantics — that is, formal systems and interpretations — to work well together. In an ideal situation, we should all expect that a formula be a theorem in a formal system if and only if it be true in any suitable interpretation of its language. That is what should happen and, indeed, that is what we know happens in the formalisations of propositional and predicate logic. Unfortunately, that is everything we have. As we will later discuss, Gödel showed that this perfect harmony can never be reached for any formal system powerful enough to capture basic arithmetic. I know, that is a real pain in the neck: syntax can only do its job perfectly wherever it is not necessary.

As catastrophic as this may seem, this is not that bad; in fact, depending on your views on the philosophy of mathematics, it can be great news. Please, bear with me.

§2 Propositional logic. General definitions

2.1 A note on circularity. From now on, we will be working with propositional logic and predicate logic as object theories. This unavoidably leads one to wonder what kind of reasoning they should be able to use in their metatheory. I mean, it does not seem legit to use propositional logic in order to study propositional logic, right?

Turns out we can safely do it. What we will be studying is not propositional logic itself, but a formalisation of propositional logic. It should be out of question to any rational being that propositional logic is perfectly valid (and, thus, that there is no harm in using it). What will not be obvious, however, is that the formalisation of propositional logic that we are about to present is correct and faithfully represents it.

2.2 Definition. The formal system P of propositional logic is defined by the tuple $(\mathrm{L_P}, \Xi_\mathrm{P}, \mathrm{R_P})$ where $\mathrm{L_P}$ is the language defined in 1.7, the set of axioms Ξ_P consists of all the formulas of the form

(P1) $\mathrm{A} \to (\mathrm{B} \to \mathrm{A})$,
(P2) $(\mathrm{A} \to (\mathrm{B} \to \mathrm{C})) \to ((\mathrm{A} \to \mathrm{B}) \to (\mathrm{A} \to \mathrm{C}))$,
(P3) $(\neg\mathrm{A} \to \neg\mathrm{B}) \to (\mathrm{B} \to \mathrm{A})$,

for $\mathrm{A}, \mathrm{B}, \mathrm{C} \in \mathrm{F_P}$, and $\mathrm{R_P}$ has a single rule of inference:

$$\{\mathrm{A}, (\mathrm{A} \to \mathrm{B})\} \vdash_\mathrm{P} \mathrm{B}$$

for all $\mathrm{A}, \mathrm{B} \in \mathrm{F_P}$. This rule is known as *modus ponens*, MP for short.

2.3. As you may have already noticed, axioms such as (P1), (P2) and (P3) are not axioms per se. Instead, they are "rules" for constructing the axioms; that is why they are called *axiom schemata* instead of axioms. The axioms that are constructed using a particular axiom schema are said to be *instances* of the schema. For example, the axiom $(p_1 \to p_2) \to p_1$ is an instance of the axiom schema (P1).

2.4 Definition. An interpretation of $\mathrm{L_P}$ based on the semantics of propositional logic is a function $i : \mathrm{T_P} \longrightarrow \{0, 1\}$. Each interpretation i induces a *valuation* function $v_i : \mathrm{F_P} \longrightarrow \{0, 1\}$ verifying the following semantic rules:

1. The function v_i takes the same values as i on the set of propositional variables $\mathrm{T_P} \subseteq \mathrm{F_P}$.
2. If $\mathrm{A} \in \mathrm{F_P}$, then $v_i(\mathrm{A}) \neq v_i(\neg\mathrm{A})$.
3. If $\mathrm{A}, \mathrm{B} \in \mathrm{F_P}$, then $v_i(\mathrm{A} \to \mathrm{B}) = 0$ if and only if $v_i(\mathrm{A}) = 1$ and $v_i(\mathrm{B}) = 0$.

Given any propositional form A, we say that A is true in i if $v_i(\mathrm{A}) = 1$. If, instead, $v_i(\mathrm{A}) = 0$, we say that A is false.

Valid propositional forms — that is, valid formulas in the context of propositional logic — are referred to as *tautologies*. Propositional forms that are false under any interpretation of propositional logic are said to be *contradictions.*

2.5. It should be obvious that an interpretation is nothing more than an assignment of truth values. The interpretation function i defined for the propositional symbols represents the assignment and then the "truth value" of the remaining formulas, given by their image under v_i, is obtained inductively by applying the semantic rules that define the connectives $\neg$ and $\rightarrow$.

It should be obvious, according to the semantic rules that we have used, that getting the value of $v_i(\mathrm{A})$ for a propositional form A and an interpretation i is the same as getting the truth value of A under the assignment of truth values induced by i. Thus, all the informal methods that can used to decide the truth or falsity of a propositional forms can be safely used to compute valuations.

Notice, by the way, how we have redefined tautologies. The underlying meaning, however, is the same.

2.6 Definition. Let $\mathrm{L} = (\Sigma, \mathrm{T}, \mathrm{F}, \Lambda)$ be a language and $\mathsf{FS} = (\mathrm{L}, \Xi, \mathrm{R})$ be a formal system containing all instances of axiom schemata (P1), (P2) and (P3) among its axioms.

- If there exists no formula $\mathrm{A} \in \mathrm{F}$ such that $\vdash_{\mathsf{FS}} \mathrm{A}$ and $\vdash_{\mathsf{FS}} \neg\mathrm{A}$, the system FS is said to be *consistent.*
- If, for every sentence $\mathrm{A} \in \Lambda$, we have $\vdash_{\mathsf{FS}} \mathrm{A}$ or $\vdash \neg\mathrm{A}$, we say that FS is *syntactically complete* or, for short, *complete.*
- The system FS is said to be *semantically complete* if every valid formula is a theorem in FS. Conversely, if every theorem is a valid formula, FS is said to be *sound.*

Let A be an axiom and let FS^* be the formal system obtained by removing the axiom A from FS. If neither A nor $\neg$A are theorems of FS^*, A is said to be an *independent axiom.* Ideally, we want all axioms to be independent in order to avoid redundancy.

2.7 Example. We will prove that $\vdash \mathrm{A} \rightarrow \mathrm{A}$ for any formula A. We will write down all the steps of the deduction together with their justification.

(1) [(P1)] $\mathrm{A} \rightarrow ([\mathrm{A} \rightarrow \mathrm{A}] \rightarrow \mathrm{A})$,

(2) [(P2)] $(\mathrm{A} \rightarrow ([\mathrm{A} \rightarrow \mathrm{A}] \rightarrow \mathrm{A})) \rightarrow ((\mathrm{A} \rightarrow [\mathrm{A} \rightarrow \mathrm{A}]) \rightarrow (\mathrm{A} \rightarrow \mathrm{A}))$,

(3) [MP on (1), (2)] $((\mathrm{A} \rightarrow [\mathrm{A} \rightarrow \mathrm{A}]) \rightarrow (\mathrm{A} \rightarrow \mathrm{A}))$,

(4) [(P1)] $\mathrm{A} \rightarrow (\mathrm{A} \rightarrow \mathrm{A})$,

(5) [MP on (4), (3)] $\mathrm{A} \rightarrow \mathrm{A}$.

2.8 Theorem (Deduction theorem in P)**.** For any collection of propositional forms $\Gamma \subseteq \mathrm{F}_{\mathrm{P}}$ and any formulas $\mathrm{A}, \mathrm{B} \in \mathrm{F}_{\mathrm{P}}$, one can deduce $\Gamma \cup \{\mathrm{A}\} \vdash_{\mathrm{P}} \mathrm{B}$ if and only if $\Gamma \vdash_{\mathrm{P}} (\mathrm{A} \rightarrow \mathrm{B})$.

Proof. We will first show that $\Gamma \cup \{\mathrm{A}\} \vdash \mathrm{B}$ implies $\Gamma \vdash (\mathrm{A} \rightarrow \mathrm{B})$ following a proof by induction on the length n of the deduction of $\Gamma \cup \{\mathrm{A}\} \vdash \mathrm{B}$. If $n = 1$, there are only three possibilities: either $\mathrm{B} = \mathrm{A}$, $\mathrm{B} \in \Gamma$, or $\mathrm{B} \in \Xi_{\mathrm{P}}$.

According to 2.7, any formula A verifies $\vdash (\mathrm{A} \rightarrow \mathrm{A})$, so, in particular, $\Gamma \vdash (\mathrm{A} \rightarrow \mathrm{A})$. Thus, if $\mathrm{B} = \mathrm{A}$, it is obvious that $\Gamma \cup \{\mathrm{A}\} \vdash \mathrm{A}$ implies $\Gamma \vdash (\mathrm{A} \rightarrow \mathrm{A})$.

If $\mathrm{B} \in \Gamma$ or $\mathrm{B} \in \Xi_{\mathrm{P}}$, $\Gamma \vdash (\mathrm{A} \rightarrow \mathrm{B})$ follows from a trivial application of MP to B and

(P1).[1]

Let us now assume the result to hold for deductions of an arbitrary length n and prove it for those of length $n + 1$. If the deduction is of length $n + 1$, the formula B may be, as in the base case, an axiom, equal to A, or an element of Γ. But it may also have been obtained from an application of MP on two previous elements of the deduction. In this case, those elements have deductions of length smaller than $n + 1$ and, therefore, they satisfy the result by the inductive hypothesis. Let us then assume that B has been obtained from an application of MP to two formulas of the form X and $\mathrm{X} \to \mathrm{B}$ verifying $\Gamma \vdash (\mathrm{A} \to \mathrm{X})$ and $\Gamma \vdash (\mathrm{A} \to (\mathrm{X} \to \mathrm{B}))$. Under these conditions, we can make the following deduction from Γ.

(1) [By hypothesis, can be deduced from Γ] $\mathrm{A} \to \mathrm{X}$,

(2) [By hypothesis, can be deduced from Γ] $\mathrm{A} \to (\mathrm{X} \to \mathrm{B})$,

(3) [(P1)] $(\mathrm{A} \to (\mathrm{X} \to \mathrm{B})) \to ((\mathrm{A} \to \mathrm{X}) \to (\mathrm{A} \to \mathrm{B}))$,

(4) [MP on (2), (3)] $(\mathrm{A} \to \mathrm{X}) \to (\mathrm{A} \to \mathrm{B})$,

(5) [MP on (1), (4)] $\mathrm{A} \to \mathrm{B}$.

We shall now prove the converse: assuming that $\Gamma \vdash (\mathrm{A} \to \mathrm{B})$, we will show that $\Gamma \cup \{\mathrm{A}\} \vdash \mathrm{B}$. This is easy. Keeping in mind that $\Gamma \vdash (\mathrm{A} \to \mathrm{B})$ and, therefore, that $\Gamma \cup \{\mathrm{A}\} \vdash (\mathrm{A} \to \mathrm{B})$, we can write an explicit deduction of $\Gamma \cup \{\mathrm{A}\} \vdash \mathrm{B}$:

(1) [By hypothesis, can be deduced from $\Gamma \cup \{\mathrm{A}\}$] $\mathrm{A} \to \mathrm{B}$,

(2) [Belongs to $\Gamma \cup \{\mathrm{A}\}$] A,

(3) [MP on (2), (1)] B.

This concludes the proof. ■

2.9. The deduction theorem is, perhaps, one of the most significant results in this section, for it explains the confusion that the implication connective → generates.

The formula $\mathrm{A} \to \mathrm{B}$ is a formula in the object language, full stop. What we have shown is that the meta-theoretic statement $\vdash (\mathrm{A} \to \mathrm{B})$ meaning "$(\mathrm{A} \to \mathrm{B})$ is a theorem in P" is equivalent to the meta-theoretic statement $\mathrm{A} \vdash \mathrm{B}$ meaning "B can be deduced from A".

We will later introduce the first-order version of this metatheorem, which will shed even more light on this matter.

2.10 Theorem. The following metatheorems about propositional logic are true:

(i) If $\Gamma \subseteq \mathrm{F}_\mathrm{P}$ and $\mathrm{X} \in \mathrm{F}_\mathrm{P}$ are such that $\Gamma \vdash \mathrm{X}$, then $\Gamma \vDash \mathrm{X}$. In particular, if a formula $\mathrm{X} \in \mathrm{F}$ is a theorem in P, it is a tautology; which is to say that P is sound.
(ii) The formal system P is consistent.
(iii) Every tautology is a theorem in P. In other words, P is semantically complete.

These results show beyond any doubt that P is a correct formalisation of propositional logic.

[1] Be aware that, when applying (P1), "A" in (P1) should be substituted by "B" and "B" by "A."

Proof. (i) We proceed by induction on the length n of the deduction. If $n = 1$, then either $X \in \Gamma$ (in which case the result is obvious) or X is an axiom of P. In order for the result to hold, we need to see that every axiom of P is true under any interpretation, i.e., that it is a tautology.

We shall first analyse (P1). I think we can both agree that, given any $A, B \in F$, the formula $A \to (B \to A)$ either is or is not a tautology. Thus, we just need to show that it is impossible for $A \to (B \to A)$ not to be a tautology. Were that formula not a tautology, there would necessarily exist an interpretation i under which it would be false. Nonetheless, according to the semantic rules of propositional logic, that would mean that $v_i(A) = 1$ yet $v_i(B \to A) = 0$. But, by those same rules, $v_i(B \to A) = 0$ can only mean that $v_i(B) = 1$ and $v_i(A) = 0$. Consequently, $A \to (B \to A)$ can only be false under an interpretation i verifying both $v_i(A) = 0$ and $v_i(A) = 1$. As that is impossible, we can safely conclude that all the axioms defined by the schema (P1) are tautologies.

We can proceed in a similar fashion regarding (P2). Let $A, B, C \in F_P$ be formulas. If an interpretation i is such that

$$v_i(\underbrace{(A \to (B \to C))}_{D_1} \to \overbrace{((A \to B) \to (A \to C))}^{D_2}) = 0,$$

then we necessarily have $v_i(D_1) = 1$ and $v_i(D_2) = 0$. Having $v_i(D_2) = 0$ implies that $v_i(A \to C) = 0$ and $v_i(A \to B) = 1$, which can only mean that $v_i(A) = 1$, that $v_i(C) = 0$ and that $v_i(B) = 1$. Simultaneously, $v_i(D_1) = 1$ with $v_i(A) = 1$ leads to $v_i(B \to C) = 1$, which — with $v_i(C) = 0$ — could only be the case if $v_i(B) = 0$. Having reached a contradiction, we can conclude that all the axioms defined by (P2) need be true under any interpretation and, therefore, that they are all tautologies.

Finally, let us tackle (P3). Given any two formulas $A, B \in F_P$, if we assume an interpretation i to exist such that

$$v_i((\neg A \to \neg B) \to (B \to A)) = 0,$$

it will need to verify $v_i(B \to A) = 0$ and $v_i(\neg A \to \neg B) = 1$. The former of these conditions implies that $v_i(A) = 0$ and $v_i(B) = 1$, which — according to the semantic rules — is equivalent to having $v_i(\neg A) = 1$ and $v_i(\neg B) = 0$. If $v_i(\neg A \to \neg B) = 1$ and, as we have just shown, $v_i(\neg A) = 1$, then we necessarily have $v_i(\neg B) = 1$, which is, as expected, a contradiction. This proves that all the axioms defined by (P3) are tautologies.

Now that we have completed the base case, let us assume the result to hold for deductions of length equal to or smaller than n, and we will prove it for those of length $n + 1$. If the deduction has length $n + 1$, X may be an element of Γ or an axiom as in the base case, or it may have been obtained from an application of MP on two previous elements of the deduction that, therefore, verify the result according to the inductive hypothesis. Thus, we need to show that, given any two formulas of the form A and $A \to B$, if they are true in any particular interpretation i, so is B.

If $v_i(A) = 1$ and $v_i(A \to B) = 1$, it is obvious that we need to have $v_i(B) = 0$. Indeed, if $v_i(B) = 0$ and $v_i(A) = 1$, that would yield $v_i(A \to B) = 0$. Thus, we have shown that, for any interpretation i in which any two formulas A and $A \to B$ are true, B is true too.

(ii) Let A be a theorem in P. By (i), A need be a tautology and, therefore, for an interpretation i, we will have $v_i(A) = 1$. According to the semantic rules, this means that, under any interpretation i, $v_i(\neg A) = 0$. Consequently, ¬A is not a tautology and since, by (i), being a tautology is a necessary condition for any formula to be a theorem in P, ¬A cannot be a theorem.

This shows that no formula A can verify both $\vdash_P A$ and $\vdash_P \neg A$, and, therefore, that P is consistent.

(iii) The details of this proof are pretty tedious to go through. If you feel motivated enough to do it, feel free to visit A.3 in the appendices.

■

2.11 Lemma (Conjunction introduction rule)**.** Let Γ be a collection of propositional forms and let A and B be two arbitrary propositional forms. One can deduce $\Gamma \vdash_P (A \wedge B)$ if and only if one can deduce both $\Gamma \vdash_P A$ and $\Gamma \vdash_P B$.

Proof. If $\Gamma \vdash (A \wedge B)$, then, since $(A \wedge B) \rightarrow A$ and $(A \wedge B) \rightarrow B$ are both tautologies — and, therefore, theorems in P — it follows by a direct application of the MP rule that $\Gamma \vdash A$ and $\Gamma \vdash B$.

Conversely, if $\Gamma \vdash A$ and $\Gamma \vdash B$, we know $A \rightarrow (B \rightarrow (A \wedge B))$ to be another tautology. Two applications of MP yield $\Gamma \vdash (A \wedge B)$. ■

2.12 Proposition. Given any propositional forms A and B, $\Gamma \vdash_P (A \leftrightarrow B)$ if and only if $\Gamma \vdash_P (A \rightarrow B)$ and $\Gamma \vdash_P (B \rightarrow A)$.

In particular, if $\Gamma = \emptyset$, $A \leftrightarrow B$ is a theorem of P if and only if so are $A \rightarrow B$ and $B \rightarrow A$.

Proof. If we have both $\Gamma \vdash (A \rightarrow B)$ and $\Gamma \vdash (B \rightarrow A)$, by 2.11, we know that $\Gamma \vdash ((A \rightarrow B) \wedge (B \rightarrow A))$, which is, according to our definition of ↔, the same as $\Gamma \vdash (A \leftrightarrow B)$. The converse is also a direct consequence of 2.11. ■

2.13 Lemma. Let A, X and Y be propositional forms. If $X \leftrightarrow Y$ is a tautology and A' denotes the propositional form resulting from replacing each appearance of X in A by Y, then $A \leftrightarrow A'$ is a tautology and, therefore, a theorem in P.

Proof. Let i be any interpretation. It suffices to notice that, as $X \leftrightarrow Y$ is a tautology, we always have $v_i(X) = v_i(Y)$. Therefore, as A' is obtained by replacing every occurrence of X by an occurrence of Y, we necessarily have $v_i(A) = v_i(A')$. Consequently, $v_i(A \leftrightarrow A') = 1$. ■

2.14. It is very easy to see that the following formulas are tautologies for any propositional forms A, B and C:

(i) $(A \leftrightarrow B) \leftrightarrow (B \leftrightarrow A)$,
(ii) $(A \wedge B) \leftrightarrow (B \wedge A)$,
(iii) $((A \wedge B) \wedge C) \leftrightarrow (A \wedge (B \wedge C))$,
(iv) $((A \vee B) \vee C) \leftrightarrow (A \vee (B \vee C))$,

This, together with 2.13, should be enough to convince you that the conventions we introduced in 0-2.4 can be safely used when working with L_P in P. In particular, this shows that there is no harm in swapping formulas around the ∨, ∧ and ↔ connectives.

2.15 Proposition. Let A, B, A_1, A_2, B_1 and B_2 be propositional forms and Γ a set of propositional forms.

(i) One can deduce $\Gamma \vdash_{P} (A \to (B_1 \wedge B_2))$ if and only if one can deduce both $\Gamma \vdash_{P} (A \to B_1)$ and $\Gamma \vdash_{P} (A \to B_2)$. In particular, if $\Gamma = \emptyset$, $A \to (B_1 \wedge B_2)$ is a theorem of P if and only if so are $A \to B_1$ and $A \to B_2$.

(ii) One can deduce $\Gamma \vdash_{P} ((A_1 \wedge A_2) \to B_1)$ if and only if one can deduce $\Gamma \vdash_{P} (A_1 \to (A_2 \to B))$ or $\Gamma \vdash_{P} (A_2 \to (A_1 \to B)$. In particular, if $\Gamma = \emptyset$, $(A_1 \wedge A_2) \to B$ is a theorem if and only if so are $A_1 \to (A_2 \to B)$ or $A_2 \to (A_1 \to B)$.

Proof. (i) Follows from a direct application of MP taking into account 2.12 and the fact that

$$(A \to (B_1 \wedge B_2)) \leftrightarrow ((A \to B_1) \wedge (A \to B_2))$$

is a tautology.

(ii) Follows from a direct application of MP taking 2.12 into account together with the fact that both

$$((A_1 \wedge A_2) \to B) \leftrightarrow (A_1 \to (A_2 \to B)),$$
$$((A_1 \wedge A_2) \to B) \leftrightarrow (A_2 \to (A_1 \to B))$$

are tautologies. ■

2.16 Proposition (Principle of explosion)**.** Anything can be deduced from a false premise: given any two propositional forms $A, B \in F_P$, we have $A, \neg A \vdash B$.

Proof. It is easy to see that the formulas $A \to (A \vee B)$ and $\neg A \to ((A \vee B) \to B)$ are tautologies. We know P to be semantically complete and, therefore, we know those tautologies to be theorems of P. The deduction of B from $A, \neg A$ is then very simple.

(1) [Premise] A,

(2) [Premise] $\neg A$,

(3) [Theorem] $A \to (A \vee B)$,

(4) [Theorem] $\neg A \to ((A \vee B) \to B)$,

(5) [MP on (1), (3)] $A \vee B$,

(6) [MP on (2), (4)] $(A \vee B) \to B$,

(7) [MP on (5), (6)] B.

This completes the proof. ■

§3 Predicate logic

3.1 Definition. Let $L = (\Sigma, T, F)$ be a first-order language and $\xi \subseteq F$ a collection of formulas. The first order system H on the language L defined by the *non-logical axioms* ξ is the formal system $H = (L, \Xi, R)$ where Ξ and R are defined as follows.

The set of axioms Ξ consists of ξ and all the axioms defined by the schemata (P1), (P2) and (P3) in addition to the following.

(Q1) If A is a formula, x_i a variable and t a term not containing any variables quantified in A, then $((\forall x_i)\mathrm{A} \to \mathrm{A}(x_i \| t))$. The formula A may or may not contain occurrences of x_i. When we use $\mathrm{A}(x_i \| t)$, we refer to the formula obtained by replacing every occurrence of x_i in A, should there be any, by t.

(Q2) If A and B are formulas, x_i is a variable and A does not contain x_i as a free variable, then $((\forall x_i)(\mathrm{A} \to \mathrm{B}) \to (\mathrm{A} \to (\forall x_i)\mathrm{B}))$.

The set R of rules of inference contains modus ponens (which we have already studied) and the rule of *generalisation*. The rule of generalisation states that, for any formula $\mathrm{A} \in \mathrm{F}$ and any variable $x_i \in \mathrm{T}$, $\mathrm{A} \vdash (\forall x_i)\mathrm{A}$.

The formal system Q of first-order logic is the first-order system on the language $\mathrm{L_Q}$ defined in 1.10(i) with an empty set of non-logical axioms.

3.2. There are some important remarks that need to be made about the definition I have just given you. The first of them concerns the purpose of the generalisation rule. When we work in first-order logic, we are only concerned with closed formulas (sentences), so the generalisation rule fixes the oddity of free variables by saying "hey, is there a free variable in this formula? Well, that is the same as saying that the formula holds for every possible value this variable can take".

You see, when setting out the axioms for a first order-system, we could have done so in such a way that no axioms would contain free variables. Nevertheless, that would not have stopped anyone from trying to make deductions from non-closed formulas. Hence, the only way to prevent the appearance of free variables in our variables would have been to define first order languages in such a way that all its formulas were closed. Doing so in a proper, inductive way is — as far as I know — unnecessarily complicated. Furthermore, we would be losing a lot of expressive power. You know, sometimes free variables have their place! Maybe not as theorems of a formal system, but would you dare say that, if we were considering a first-order language of set theory, the formula $x = \varnothing$ is not a formula? That does not seem right. So, instead of doing weird things when defining our language, the most simple thing to do is to introduce the generalisation rule.

Let us now see the generalisation rule in action with a simple example. Notice how, when combined with (Q1) and MP, it allows us to deduce, for any given formula A dependent on a variable x_i, that $\mathrm{A} \vdash \mathrm{A}(x_i \| x_j)$. This enables us to relabel our variables in our formulas freely, which is a pretty good deal.

With that sorted, there is a part of the definition that might have created some confusion. Unless you paid close attention to 0-3.5, you might be having issues understanding how (Q1) and (Q2) are parenthesised. In (Q1), the scope of the first quantifier is only A, whereas, in (Q2), the scope of the first quantifier is $(\mathrm{A} \to \mathrm{B})$ and the scope of the second is B.

If this has not been enough for you to fully understand what is going on, I invite you to revisit 0-3.5 and take things a little bit more slowly!

3.3 Example. (i) A very interesting kind of first-order system is that of first-order systems with equality. These systems use the first-order language associated to any set of symbols containing $\{=\}$ with signature $\sigma(=) = -2$; in other words, it uses any language incorporating one additional symbol representing a 2-ary predicate. Instead of writing $=(x, y)$, it is customary to use infix notation and write $x = y$.

The set of non-logical axioms of these formal systems contains the following formulas and schemata.

(E1) $(\forall x).\ x = x$.
(E2) $(\forall x)(\forall y).\ x = y \to y = x$.
(E3) If f is an n-ary function symbol and $t_1, \dots t_n$ are terms, $(\forall x)(\forall y).\ x = y \to f(t_1, \dots, x, \dots, t_n) = f(t_1, \dots, y, \dots, t_n)$.
(E4) If P is an n-ary predicate symbol and $t_1, \dots, t_n$ are terms, $(\forall x)(\forall y).\ x = y \to (\mathrm{P}(t_1, \dots, x, \dots, t_n) \to \mathrm{P}(t_1, \dots, y, \dots, t_n))$.

Only (E3) and (E4) are axiom schemata. The remaining items are axioms.

(ii) The standard first-order formalisation of arithmetic is known as Peano Arithmetic. The formal system PA of Peano Arithmetic uses the first-order language of arithmetic defined in 1.10(ii). It is a formal system using first-order logic with equality and — in addition to (E1), (E2), (E3) and (E4) — its non-logical axioms are:

(PA1) $(\forall x).\ \neg(s(x) = 1)$.
(PA2) $(\forall x, y).\ s(x) = s(y) \to x = y$.
(PA3) $(\forall x).\ x + 1 = s(x)$.
(PA4) $(\forall x, y).\ x + s(y) = s(x + y)$.
(PA5) $(\forall x).\ x \cdot 1 = x$.
(PA6) $(\forall x, y).\ x \cdot s(y) = x \cdot y + x$.
(PA7) If $\mathrm{A} \in \mathrm{F}$ has a free variable x, $(\mathrm{A}(x \| 0) \wedge (\forall x)(\mathrm{A} \to \mathrm{A}(x \| s(x)))) \to (\forall x)\mathrm{A}$.

Axiom (PA1) simply states that, within the theory of arithmetic, 1 cannot be the successor of any number. Axiom (PA2) establishes that two numbers are equal whenever the numbers succeeding them are equal. Axioms (PA3) and (PA4) define addition, and (PA5) and (PA6) define multiplication. Finally, the axiom schema (PA7) formalises the principle of induction.

When I first studied this, the axiom schema of induction made me feel suspicious. It kind of looks like a second-order axiom, doesn't it? It seems like we are quantifying over all the predicates! Isn't that cheating? Well, no. There is a second-order formalisation of Peano Arithmetic that, instead of using this axiom schema, uses a proper quantification over all the possible predicates, and the consequences of that seemingly innocent difference are very significant. We will come back to this later.

3.4 Definition. Let S be a collection of symbols together with a signature function σ. Let $\mathrm{L} = (\Sigma, \mathrm{T}, \mathrm{F})$ be the first-order language associated to (S, σ). An interpretation I of L is the assignment of a set D_I as the *domain* of the interpretation together with an *interpretation function* ι_I mapping every $c \in \mathrm{S}$ with $\sigma(c) = 0$ to an element $\iota_\mathrm{I}(c) \in \mathrm{D}$, every $f \in \mathrm{S}$ with $\sigma(c) = n > 0$ to an n-ary function $\iota_\mathrm{I}(f) : \mathrm{D} \times \cdots \times \mathrm{D} \longrightarrow \mathrm{D}$, and every $\mathrm{P} \in \mathrm{S}$ with $\sigma(\mathrm{P}) = -n < 0$ to an n-ary predicate $\iota_\mathrm{I}(\mathrm{P})$ taking values in D. Instead of writing $\iota_\mathrm{I}(c)$, $\iota_\mathrm{I}(f)$ and $\iota_\mathrm{I}(\mathrm{P})$, we will often use $\underline{c}$, $\underline{f}$ and $\underline{\mathrm{P}}$, provided there is no room for ambiguity, in order to make our notation more clear.

To put it in less formal terms, an interpretation is nothing more than the definition of a domain (in which the variables of the languages are meant to take values) and an assignment of constants in that domain to the constant symbols of the language, of n-ary predicates taking arguments in the domain to each n-ary predicate symbol of the language, and of n-ary functions taking arguments in values in S to every n-ary function symbol of the language.

Let us then consider an arbitrary interpretation I on L. An *assignment of values* (*assignment*, for short) is any function $\alpha : \{x_1, \dots, x_n, \dots\} \longrightarrow \mathrm{D_I}$ that can be extended to a function $\tilde{\alpha} : \mathrm{T} \longrightarrow \mathrm{D_I}$ by defining, for every n-ary function letter $f \in \mathrm{S}$,

$$\tilde{\alpha}(f(x_1, \dots, x_n)) = \underline{f}(\tilde{\alpha}(x_1), \dots, \tilde{\alpha}(x_n)).$$

Each of these assignments induces a valuation function $v_\alpha : \mathrm{F} \longrightarrow \{0, 1\}$ defined by the following inductive rules:

1. If $\mathrm{P} \in \mathrm{S}$ is an n-ary predicate letter and $t_1, \dots, t_n \in \mathrm{T}$, then $v_\alpha(\mathrm{P}(t_1, \dots, t_n)) = 1$ if and only if the predicate $\underline{\mathrm{P}}$ holds for $(\tilde{\alpha}(t_1), \dots, \tilde{\alpha}(t_n))$.
2. If $\mathrm{A}, \mathrm{B} \in \mathrm{F}$, then $v_\alpha(\mathrm{A} \to \mathrm{B}) = 0$ if and only if $v_\alpha(\mathrm{A}) = 1$ and $v_\alpha(\mathrm{B}) = 0$.
3. If $\mathrm{A} \in \mathrm{F}$, then $v_\alpha(\neg\mathrm{A}) = 1$ if and only if $v_\alpha(\mathrm{A}) = 0$.
4. If $\mathrm{A} \in \mathrm{F}$, then $v_\alpha((\forall x_i)\mathrm{A}) = 1$ if and only if, for every assignment α' verifying $\alpha'(x_j) = \alpha(x_j)$ for every index $j \neq i$, $v_{\alpha'}(\mathrm{A}) = 1$.

If, given a formula $\mathrm{A} \in \mathrm{F}$, $v_\alpha(\mathrm{A}) = 1$, it is said that the assignment α *satisfies* the formula A. Informally speaking, the only purpose of assignments is, as their name suggests, assigning a value to the variables in the formulas. A formula is true in an interpretation I if and only if it is satisfied by every assignment in I. Analogously, a formula is false in I if and only if it is not satisfied by any assignment in I.

Two interpretations I and J of a language L are said to be *isomorphic* if they are equal up to a relabelling of the elements of their domains. To put it in formal terms, they are said to be isomorphic if there exists a bijective function $\theta : \mathrm{D_I} \longrightarrow \mathrm{D_J}$ verifying, for every constant symbol $c \in \mathrm{S}$, $\iota_\mathrm{J}(c) = \theta(\iota_\mathrm{I}(c))$; for every n-ary function symbol f,

$$\iota_\mathrm{I}(f) : (x_1, \dots, x_n) \longmapsto \theta^{-1}(\iota_\mathrm{J}(f)(\theta(x), \dots, \theta(x_n)),$$

and, for every n-ary predicate letter P, $\iota_\mathrm{I}(\mathrm{P})(x_1, \dots, x_n)$ if and only if $\iota_\mathrm{J}(\mathrm{P})(\theta(x_1), \dots, \theta(x_n))$.

Given a first-order system H over a first-order language L, an interpretation I of L is said to be a *model* of H if all the axioms of H are true in I. Some first-order languages and systems are built with a particular model in mind; these models are known as the *intended interpretations* or *standard models.* Any model of a formal system that is non-isomorphic to the intended interpretation is said to be a *non-standard model.*

3.5 Example. The intended interpretation $\mathcal{N}$ of the first-order language of arithmetic defined in 1.10(ii) is defined by taking the set of natural numbers $\mathbb{N}$ as the domain of discourse and by making the following assignments:

- The constant $\underline{1}$ is the number $1 \in \mathbb{N}$.
- The predicate $\underline{=}$ is the binary predicate that is true if and only if its two arguments are the same number.
- The function $\underline{s}$ is the function taking every $x \in \mathbb{N}$ to $x + 1 \in \mathbb{N}$.
- The function $\underline{+}$ is the function taking every $x, y \in \mathbb{N}$ to $x + y \in \mathbb{N}$.
- The function $\underline{\cdot}$ is the function taking every $x, y \in \mathbb{N}$ to $x \cdot y \in \mathbb{N}$.

It should be obvious that $\mathscr{N}$ is a model of PA.

The formula $(x + 1) + 1 = s(x) + 1$ is true in $\mathscr{N}$ because, regardless of the value that x is given in any assignment α, the formula is satisfied; given any $n \in \mathbb{N}$, it is true that $(n + 1) + 1 = (n + 1) + 1$. Notice how this last expression is not meant to be a formula of the formal language, but a "real" (semantic) statement about $\mathbb{N}$. A formula in a formal language is a meaningless sequence of symbols. But what we have written is just a representation, with the usual notation of arithmetic, of $(n\underline{+1})\underline{+1} = \underline{s}(n)+\underline{1}$, which in turn represents the statement "adding one to n plus one is the same as adding one to the number that goes after n". Analogously, it is easy to see how the formula $(x+1) = x$ is false in $\mathscr{N}$. Now, let us consider the formula $x = 1$. Is it true? Certainly not! It suffices to consider any assignment α with $\alpha(x) \neq 1$. Nevertheless, it is not false either, for it is satisfied by any assignment α with $\alpha(x) = 1$.

Let us now define a model isomorphic to $\mathscr{N}$. We just need to consider the interpretation $\mathscr{N}'$ obtained by

- using $\mathbb{N}' = \{1', 2', 3', \dots\}$ instead of $\mathbb{N}$;
- using the constant $1'$ instead of 1;
- using the functions $x' +' y' = (x+y)'$ instead of $+$, and $x' \cdot' y' = (x \cdot y)'$ instead of $\cdot$,
- and using the relation $x' =' y'$ defined to be true if and only if $x = y$ instead of using the relation $=$.

To conclude this example, there is one question that we need to answer: are there any non-standard models of Peano Arithmetic? Yes, there are. Nonetheless, getting to those models truly goes beyond the scope of this book.

The only way to have a formal system of arithmetic without non-standard models is to use second-order logic: and that is because of the axiom schema we discussed before! You see, the axiom schema we used when defining PA is much weaker than a quantification over all the possible predicates, for it only takes into consideration the predicates that can be defined within the language of arithmetic, and those predicates need to be formulated with a finite combination of symbols! Thus, there is an infinite amount of predicates that the formulas of our language cannot capture.

You may then wonder, why don't we study second-order logic? Well, second-order logic has its oddities too, and, as we will later see in our study of set theory, first-order logic will suffice to define a formal system able to "contain" all mathematics (and, yes, that *kind of* includes second-order arithmetic).

3.6 Theorem (First-order deduction theorem)**.** Let H be a first-order formal system over a first-order language $\mathrm{L} = (\Sigma, \mathrm{T}, \mathrm{F})$ with a set of axioms Ξ. For any collection of formulas $\Gamma \subseteq \mathrm{F}$, any *closed* formula $\mathrm{A} \in \mathrm{F}$ and any formula $\mathrm{B} \in \mathrm{F}$, if one can deduce $\Gamma \cup \{\mathrm{A}\} \vdash_{\mathsf{H}} \mathrm{B}$, then $\Gamma \vdash_{\mathsf{H}} (\mathrm{A} \to \mathrm{B})$.

Conversely, even if A is not closed, if $\Gamma \vdash_{\mathsf{H}} (\mathrm{A} \to \mathrm{B})$, then $\Gamma \cup \{\mathrm{A}\} \vdash_{\mathsf{H}} \mathrm{B}$.

Proof. The reasoning followed in 2.8 to show that $\Gamma \cup \{\mathrm{A}\} \vdash \mathrm{B}$ implies $\Gamma \vdash (\mathrm{A} \to \mathrm{B})$ is perfectly valid for first-order systems. We just need to extend it in order to take into consideration the generalisation rule.

Proceeding by induction as in 2.8, let us then assume that B has been obtained by an application of the generalisation rule on a formula X such that, according to our inductive hypothesis, $\Gamma \vdash (A \to X)$. We will assume that, for a certain variable x_i, B is of the form $(\forall x_i)X$. In order to show that $\Gamma \vdash (A \to B)$, we just need to consider the following deduction from Γ.

(1) [By hypothesis, can be deduced from Γ] $A \to X$,

(2) [Generalisation on (1)] $(\forall x_i)(A \to X)$,

(3) [(Q2)] $(\forall x_i)(A \to X) \to (A \to (\forall x_i)X)$,

(4) [MP on (1), (3)] $A \to (\forall x_i)X$.

The proof of the converse is completely analogous to that of 2.8. Furthermore, as we pointed out in the statement of the theorem, for the converse to be true, A need not be closed. ■

3.7. Let $A \in F_P$ be any tautology of propositional logic. Given any first-order language L, any first-order formula obtained by replacing the propositional symbols in A by formulas of L is also said to be a tautology in L.

For instance, we know $A \to A \in F_P$ to be a tautology. Thus, the formula $(\forall x_1)P^1_1(x_1) \to (\forall x_1)P^1_1(x_1)$ is a tautology in L_Q.

3.8 Proposition. Let H be an arbitrary first-order system defined on a first-order language $L = (\Sigma, T, F)$.

(i) Any tautology in L is a theorem of H.
(ii) Any tautology in L is a valid formula.

Proof. (i) Let A be a tautology in propositional logic and let $A' \in F$ be the first-order tautology obtained by substituting the propositional symbols of A by formulas in L. As P is semantically complete, $\vdash_P A$ and, therefore, we there know to exist a proof $(A_1, \dots, A_n)$ in P with $A_n = A$. We can then consistently substitute all the propositional symbols in the proof by their corresponding formulas in L in such a way that $A'_n = A'$. Thus, we are led to a proof of A' in H because the MP inference rule and the axiom schemata (P1), (P2) and (P3) are included in H and, consequently, $\vdash_H A'$.

(ii) As before, let $A \in F_P$ be a tautology of propositional logic and let $A' \in F$ be a first-order formula obtained by substituting the propositional symbols in A by formulas in L. Let $B_1, \dots, B_n$ be those distinct formulas and, without loss of generality, let $p_1, \dots, p_n$ be the respective propositional symbols being substituted by them.

Before diving into the proof, let us extend our notation. For any formula $P \in F_P$ using, exclusively, the propositional symbols $p_1, \dots, p_n$, we will denote by P' the first-order formula in L obtained by replacing every propositional symbol p_k in P by B_k.

Let us consider an arbitrary interpretation $\mathscr{I}$ of L. We need to show that, for any assignment of values α in $\mathscr{I}$, $v_\alpha(A) = 1$. In order to achieve this, we will first prove that — for any interpretation i of propositional logic verifying, for any $k \in \{1, \dots, n\}$, $v_i(p_k) = v_\alpha(B_k)$ — we have $v_\alpha(A') = v_i(A)$, which will be equal to 1, since A is a tautology.

We will prove our claim by induction on the number of connectives m in A. For the base case $m = 0$, A' will be p_1. Thus, by the definition of i, it is clear that $v_i(p_1) = v_\alpha(B_1)$. Let us now assume our claim to hold for an arbitrary $m \in \mathbb{N} \cup \{0\}$ and prove it for $m + 1$. There are three cases we need to consider: A may be of the form

$X \to p_k$, $p_k \to X$ or $\neg X$ for some $k \in \{1, \dots, n\}$ and for some $X \in F_P$ having m connectives. In the first case we mentioned, we have a formula of the form $X \to p_k$ with $v_i(X) = v_\alpha(X')$ and $v_i(p_k) = v_\alpha(B_k)$. By the semantics of propositional and predicate logic, we know that both $v_\alpha(X' \to B_k)$ and $v_i(X \to p_k)$ are 0 if and only if $v_\alpha(X') = v_i(X) = 0$ and $v_\alpha(B_k) = v_i(p_k) = 1$, and they are both 1 otherwise. Hence, they always have the same value. The proof for the second case is analogous. Lastly, the third case is trivial: if $v_i(X) = v_\alpha(X')$, then $v_i(\neg X)$ and $v_\alpha(X')$ will both be 0 if and only if $v_i(X) = v_\alpha(X') = 1$ and 1 otherwise.

■

3.9. Let $A \in F_P$ be any contradiction of propositional logic. Given any first-order language L, any first-order formula obtained by replacing the propositional symbols in A by formulas of L is also said to be a contradiction in L. It can be easily shown, in full analogy with the proof of 3.8(ii), that any first-order contradiction is false under any interpretation. I will leave the details for you.

3.10. The fact that any tautology in a first order system is both a theorem and a valid formula enables us to import effortlessly many results from propositional logic into predicate logic. In the remainder of this section, we will present a bunch of those results referencing their analogues in our treatment of propositional logic. If no proofs are given, it is because — once 3.8 is taken into consideration — they are practically identical to those provided in our study of propositional logic.

3.11 Lemma (Analogue of 2.11)**.** Let H be a first-order system on a first-order language $L = (\Sigma, T, F)$. Let $\Gamma \subseteq F$ and let A and B be two arbitrary formulas. One can deduce $\Gamma \vdash_H (A \wedge B)$ if and only if one can deduce both $\Gamma \vdash_H A$ and $\Gamma \vdash_H B$.

3.12 Proposition (Analogue of 2.12)**.** Let H be a first-order system on a first-order language L. Given any formulas A and B of L, $\Gamma \vdash_H (A \leftrightarrow B)$ if and only if $\Gamma \vdash_H (A \to B)$ and $\Gamma \vdash_H (B \to A)$.

In particular, if $\Gamma = \varnothing$, $A \leftrightarrow B$ is a theorem of H if and only if so are $A \to B$ and $B \to A$.

3.13 Proposition (Principle of explosion. Analogue of 2.16)**.** In an arbitrary first-order system H, anything can be deduced from a false premise: given any two formulas A, B in its language, we have $A, \neg A \vdash_H B$.

3.14 Lemma. Let H be a consistent first-order system with a set of non-logical axioms ξ. If A is a closed formula that is not a theorem in H, the formal system H^* obtained from H by adding $\neg A$ as a non-logical axiom is consistent.

Proof. Let us assume that H^* is inconsistent and, therefore, that, for a formula $B \in F$, both B and $\neg B$ are theorems of H^*. From the principle of explosion 3.13, it follows that $\vdash_{H^*} A$. Nevertheless, since H^* is nothing more than H with $\neg A$ as an additional axiom, any proof in H^* is a deduction from $\neg A$ in H. Therefore, $\neg A \vdash_H A$.

By hypothesis, A is closed and so must be $\neg A$. Under these conditions, we can apply the deduction theorem to conclude that $\vdash_H \neg A \to A$. In addition, since the tautology $(\neg A \to A) \to A$ is a theorem of H, so an application of MP yields $\vdash_H A$, which cannot be the case according to our hypotheses and proves that H^* need be consistent. ■

3.15 Theorem. Let H be any first-order system on a first-order language $\mathrm{L} = (\Sigma, \mathrm{T}, \mathrm{F})$. The following metatheorems are true.

(i) Every instance of the axiom schemata (P1), (P2), (P3), (Q1) and (Q2) in L is valid. Consequently, verifying the non-logial axioms of a first-order system is sufficient to show an interpretation to be a model.

(ii) Let $\mathscr{M}$ be a model of H. If, given any collection of formulas $\Gamma \subseteq \mathrm{F}$ and any formula $\mathrm{A} \in \mathrm{F}$, it can be deduced that $\Gamma \vdash \mathrm{A}$, then $\mathscr{M} \vDash \Gamma$ implies $\mathscr{M} \vDash \mathrm{A}$. In other words, if $\Gamma \vdash \mathrm{A}$, any model in which all the formulas in Γ are true makes A true too. If, in particular, we take $\Gamma = \varnothing$, this means that any theorem of H is true in any model of H.

(iii) If H has a model, it is consistent.

(iv) If H is consistent, it has a model.

(v) Any valid formula A in L is a theorem in H. In other words, any first-order system is semantically complete.

Proof. (i) All instances of the axiom schemata (P1), (P2) and (P3) are, undoubtedly, tautologies; therefore, applying 3.8(ii), we already known them to be valid formulas. Let us then focus on the axiom schemata (Q1) and (Q2) and show that all of their instances are valid. For this purpose, we will consider an arbitrary interpretation $\mathscr{I}$ of L and an arbitrary assignment of values α in $\mathscr{I}$.

Let us begin with (Q1). Let $\mathrm{A} \in \mathrm{F}$ be an arbitrary formula, $x_i \in \mathrm{T}$ be any variable and $t \in \mathrm{T}$ be any term containing no variables that are quantified in A. We need to show that $v_\alpha((\forall x_i)\mathrm{A} \to \mathrm{A}(x_i \| t)) = 1$. If we had $v_\alpha((\forall x_i)\mathrm{A} \to \mathrm{A}(x_i \| t)) = 0$, then, necessarily, $v_\alpha((\forall x_i)\mathrm{A}) = 1$ and $v_\alpha(\mathrm{A}(x_i \| t)) = 0$. Nevertheless, the fact that $v_\alpha((\forall x_i)\mathrm{A}) = 1$ means that, for any valuation α' with $\alpha(x_j) = \alpha'(x_j)$ for any $i \neq j$, we have $v_{\alpha'}(\mathrm{A}) = 1$. We will consider a particular α' satisfying $\alpha'(x_i) = t$. Since none of the variables present in t are quantified in A, it follows that $v_\alpha(\mathrm{A}(x_i \| t)) = v_{\alpha'}(\mathrm{A}) = 1$. As it is impossible for $v_\alpha(\mathrm{A}(x_i \| t))$ to be both 0 and 1, we can conclude that it is impossible for $v_\alpha((\forall x_i)\mathrm{A} \to \mathrm{A}(x_i \| t))$ to be 0.

Lastly, let us analyse (Q2). Let $\mathrm{A}, \mathrm{B} \in \mathrm{F}$ be any formulas and let x_i be any variable not appearing free in A. Let us assume that

$$v_\alpha((\forall x_i)(\mathrm{A} \to \mathrm{B}) \to (\mathrm{A} \to (\forall x_i)\mathrm{B})) = 0.$$

In this scenario, we have $v_\alpha((\forall x_i)(\mathrm{A} \to \mathrm{B})) = 1$ and $v_\alpha(\mathrm{A} \to (\forall x_i)\mathrm{B}) = 0$ simultaneously. If $v_\alpha((\forall x_i)(\mathrm{A} \to \mathrm{B})) = 1$, then, for any assignment α' with $\alpha'(x_j) = \alpha(x_j)$ for any $i \neq j$, we have $v_{\alpha'}(\mathrm{A} \to \mathrm{B}) = 1$. Since x_i does not appear as a free variable in A and all such assignments α' only differ in $\alpha'(x_i)$, their valuations of A are either all equal to 1 or all equal to 0. If they were all equal to 0, then, in particular, $v_\alpha(\mathrm{A}) = 0$, which would mean that $v_\alpha(\mathrm{A} \to (\forall x_i)\mathrm{B}) = 1$, and that would contradict our hypothesis. Thus, all those assignments α' need to verify $\alpha'(x_i) = 1$. Nevertheless, since, $v_{\alpha'}(\mathrm{A} \to \mathrm{B}) = 1$, this means that $v'_\alpha(\mathrm{B}) = 1$ for any α'. It is then immediate that $v_\alpha((\forall x_i)\mathrm{B}) = 1$ and, therefore, that $v_\alpha(\mathrm{A} \to (\forall x_i)\mathrm{B}) = 1$, which, again, would contradict our initial assumptions. It follows that, necessarily, $v_\alpha((\forall x_i)(\mathrm{A} \to \mathrm{B}) \to (\mathrm{A} \to (\forall x_i)\mathrm{B})) = 1$.

(ii) We proceed by induction on the length n of the deduction $(\mathrm{A}_1, \dots, \mathrm{A}_n)$ with $\mathrm{A}_n = \mathrm{A}$ of $\Gamma \vdash \mathrm{A}$. The base case $n = 1$ is trivial: if the deduction is (A) then A may be an axiom (which is, by definition, true in any model) or an element of Γ (which is, by hypothesis, true in $\mathscr{M}$).

Let us assume the result to be true for any natural number smaller than or equal to an arbitrary $n \in \mathbb{N}$ and let us prove it for $n + 1$. If we have a deduction with $n + 1$ elements, A may still be an axiom or an element of Γ (if that were the case, we have nothing to worry about), but it may also have been obtained by applying the MP rule or the generalisation rule on previous elements of the deduction. These elements have deductions of length smaller than $n + 1$ and, therefore, the inductive hypothesis applies to them.

If A has been obtained by an application of MP to two formulas of the form X and $X \to A$ verifying $\mathscr{M} \models X$ and $\mathscr{M} \models X \to A$, is A true in $\mathscr{M}$? We know that any assignment of values α verifies $v_\alpha(B) = 1$ and $v_\alpha(B \to A) = 1$. That can only mean that $v_\alpha(A) = 1$ for any assignment α and, therefore, that $\mathscr{M} \models A$.

If, on the other hand, A is a formula of the form $(\forall x_i)X$ and has been obtained by an application of the rule of generalisation to X with $\mathscr{M} \models X$, is A true in $\mathscr{M}$? Any assignment α verifies $v_\alpha(X) = 1$ and, for $(\forall x_i)X$ to be true, we need to have $v_\alpha((\forall x_i)X) = 1$ for any assignment α. Let us consider an arbitrary assignment α and show it. By definition, $v_\alpha((\forall x_i)X) = 1$ if and only if, for every assignment α' with $\alpha'(x_j) = \alpha(x_j)$ for $j \neq i$, we have $v_{\alpha'}(X) = 1$. Is that the case? Of course it is...didn't we have $\mathscr{M} \models X$?

By the principle of mathematical induction, the proof is complete.

(iii) According to (ii), any theorem in H needs to be true in $\mathscr{M}$. Thus, if A is a theorem, every assignment α verifies $v_\alpha(A) = 1$ and, consequently, $v_\alpha(\neg A) = 0$, so $\neg A$ is false in $\mathscr{M}$ and cannot be a theorem of H.

(iv) This is an important result, but the proof is pretty lengthy and technical. If you are curious about the proof and want to go through it, check out A.5 in the appendices.

(v) Let $A \in F$ be a valid formula and let us assume that $\nvdash_H A$ and that H is consistent. According to 3.14, the formal system H^* obtained by adding the axiom $\neg A$ to H is consistent. Applying (iv), we know H^* to have a model $\mathscr{M}$, and, by the very definition of model, $\mathscr{M} \models \neg A$, which means that A will be false in $\mathscr{M}$. Nonetheless, that is impossible for, by hypothesis, A is valid and, therefore, true in every interpretation. Consequently, every valid formula is, necessarily, a theorem in H.

If H is not consistent, any formula is a theorem by the explosion principle and the result is trivial.

■

3.16 Theorem. The following metatheorems about the formal system of predicate logic are true:

(i) All the theorems of Q are valid formulas: Q is sound.
(ii) The formal system Q is consistent.

Proof. (i) According to 3.15(i), any interpretation is a model of Q. Moreover, applying 3.15(ii), any theorem of Q must be true in every model of Q, i.e. in every interpretation. This means that every theorem of Q is a valid formula.

(ii) We know, thanks to 3.15(i), that Q has a model (it can be any interpretation). Thus, the result is a direct consequence of 3.15(iii).

■

3.17 Lemma (Analogue of 2.13)**.** Let H be a first-order system on a language L. Let A, X and Y be formulas of L. If $X \leftrightarrow Y$ is a valid formula in L and A′ denotes the formula resulting from replacing each appearance of X in A by Y, then $A \leftrightarrow A'$ is a valid formula and, by 3.15(v), a theorem in H.

3.18. In 2.14, we saw how the informal manipulations of propositional forms — e.g., removing parentheses or swapping formulas around ∧ or ∨ — can be safely used in P. That same reasoning should also be enough to convince you by now that those manipulations, together with the conventions set out in 0-2.4 and 0-3.5, can be used in any first-order system.

In addition, it should be clear that the informal use of the pseudo-quantifiers that were introduced in 0-3.8 is perfectly safe in any first-order system accepting them.

3.19 Proposition (Analogue of 2.15)**.** Let H be a formal system on a language L. Let A, B, A_1, A_2, B_1 and B_2 be formulas of L and Γ a set of formulas of L.

(i) One can deduce $\Gamma \vdash_H (A \to (B_1 \wedge B_2))$ if and only if one can deduce both $\Gamma \vdash_H (A \to B_1)$ and $\Gamma \vdash_H (A \to B_2)$. In particular, if $\Gamma = \emptyset$, $A \to (B_1 \wedge B_2)$ is a theorem of H if and only if so are $A \to B_1$ and $A \to B_2$.

(ii) One can deduce $\Gamma \vdash_H ((A_1 \wedge A_2) \to B_1)$ if and only if one can deduce $\Gamma \vdash_H (A_1 \to (A_2 \to B))$ or $\Gamma \vdash_H (A_2 \to (A_1 \to B)$. In particular, if $\Gamma = \emptyset$, $(A_1 \wedge A_2) \to B$ is a theorem if and only if so are $A_1 \to (A_2 \to B)$ or $A_2 \to (A_1 \to B)$.

3.20. The remarks that we made about the deduction theorem in 2.9 are as valid for first-order closed formulas as they were for propositional forms. Nevertheless, they are not true for formulas with free variables.

In a first-order system, given two formulas A and B, $A \vdash B$ and $\vdash A \to B$ are not necessarily equivalent if A is not closed. If A has free variables, the statement $\vdash A \to B$ is stronger than $A \vdash B$; and, from a semantic perspective, it is obvious why this is the case.

Going back to 3.15(ii), we know that if $A \vdash B$, then any model *making* A *true* makes B true. If, instead, $\vdash (A \to B)$, we know that $A \to B$ is true in *any* model of the system and, therefore, that — in any model — any assignment satisfying A also satisfies B.

For example, in PA, the generalisation rule yields $x = 1 \vdash (\forall x).\ x = 1$ while, clearly, $\nvdash (x = 1 \to (\forall x).\ x = 1)$. On the other hand, $\vdash (x = 1 \to s(x) = s(1))$ and, consequently, $x = 1 \vdash s(x) = s(1)$. You see, saying that $A \vdash B$ is the same as saying that B is a theorem if we add A as a general assumption (i.e., as an axiom of the formal system), whereas $\vdash A \to B$ means that, in our formal system, whenever A is true, B is true. Do you remember when I told you that removing formulas with free variables would make us lose expressive power? This is what I was talking about.

In mathematics, when working inside any first order system, it is extremely common to represent — for any two formulas A and B — the statement $\vdash A \to B$ as $A \implies B$. Be aware that $A \implies B$, unlike $A \to B$, is a statement in the metalanguage. Using $A \implies B$ instead of $\vdash A \to B$ is so common that some mathematicians do not know what $\vdash A \to B$ means, so, unless you want to be frowned upon, always stick to $A \implies B$ unless, of course, you are working on mathematical logic as we have been doing.

In full analogy, given any two formulas A and B in a first-order system, $A \iff B$

is used to mean $\vdash A \leftrightarrow B$. This, as with $\implies$, is stronger a statement than saying both $A \vdash B$ and $B \vdash A$.

3.21 Theorem. Let us consider an arbitrary first-order system.

(i) Let A and B be any formulas. Proving that $A \implies B$ is equivalent to proving the *contrapositive*: $\neg B \implies \neg A$.

(ii) Let X be any formula and let C be any contradiction. Proving that X is a theorem is equivalent to showing that $\neg X \implies C$. In particular, if X is of the form $A \to B$, proving $A \implies B$ is the same as showing

$$A \wedge \neg B \implies (A \wedge \neg A),$$

where we have used the fact that $A \wedge \neg B$ is equivalent to $\neg(A \to B)$. The use of this technique is known as doing a *proof by contradiction.*

Notice how any proof that makes use of the contrapositive can be trivially transformed into a proof by contradiction, but not conversely.

Proof. (i) Let us assume that $A \implies B$, this is, that $\vdash A \to B$. As we know $(A \to B) \leftrightarrow (\neg B \to \neg A)$ to be a tautology and, therefore, a theorem in our formal system, a direct application of MP and 3.12 yields $\vdash \neg B \to \neg A$. The converse is analogous.

(ii) The proof of this statement is analogous to that of (i) and relies on the fact that the formula $X \leftrightarrow (\neg X \to C)$ is a tautology.

■

3.22. And now, as we reach then end of the chapter, it is time for us to address the issue that we considered in 1.17 and discuss a deep topic: Gödel's incompleteness theorem.

Now that we leave our analysis of mathematical logic behind, we will begin working in the formal system of set theory that unifies all mathematics. What properties would we like that system to have? We would certainly like it to be consistent and syntactically complete. In other words, given any formula A, we want either A or ¬A to be a theorem because, whenever you have a property of set theory, you know that either it or its negation are true and, of course, you would like your system to be powerful enough to deduce the true one (and only the true one).

If you were a committed formalist and you believed that mathematics is just a game of symbols with some rules, you would also want the formal system to be able to prove its own consistency in order for everything to fit nicely. In fact, were you not able to do that for basic arithmetic, all the work we have done would render meaningless for you because, in order to prove results about the formal system of propositional logic, we have been using some basic properties of the natural numbers and arithmetic. Thus, from the point of view of a pure formalist, the only way to see what we have done as valid would be formalising our reasoning in a formal metatheory that would need to capture basic arithmetic and prove its own consistency.

There are some nice properties that we could also ask our formal system to have (like all their axioms' being independent), but that is insignificant when compared with the importance of what we have just discussed.

Now, the question is: has anyone managed to do such a thing? Has anyone been able to prove the consistency of mathematics within mathematics? The answer is no, and that is for a very good reason...

3.23 Theorem (Kurt Gödel)**.** Let H be any consistent formal system capturing elementary arithmetic.

(i) The formal system H is not syntactically complete.
(ii) The formal system H cannot prove its own consistency.

Proof. This one does go far beyond the scope of this book. Nonetheless, if you have the time for it — depending on you level of understanding, it may take you a few days, — I invite you to read Gödel's original proof at some point. You can find it in [3]. It is a pretty illuminating experience. ■

3.24. I should warn you that what follows is a personal, partially subjective remark. That theorem was...intense, wasn't it? If you are a formalist, please accept my condolences. Just as Russel's paradox sentenced logicism by showing that we cannot reduce mathematics to logic, Gödel's theorem killed pure formalism by proving that no formal system is able to capture all mathematics or even formally prove its consistency. In other words: syntax is not enough; symbols are not enough; there is something else.

Now does this mean that arithmetic may be inconsistent? From a purely formal point of view, yes. From a human point of view, of course not. We, as humans, can see further beyond mere formalisations of theories, and, even if we cannot prove it formally, we know that Peano Arithmetic is consistent. Why? Because we see arithmetic: the reality of arithmetic is one that we have already explored through our minds and know to exist and be consistent. This fact cannot be captured formally, but that does not make it any less real.

This very same reasoning will also be applicable to the formalisation of set theory that we will soon introduce and that — as we will see — includes Peano Arithmetic and is thus affected by Gödel's theorem.

I am, of course, not trying to make a case for fully disregarding formalisation. Our minds are extremely fallible, and a formalist approach to mathematics is, to some extent, indispensable; furthermore, there is a special beauty in exploring the inner workings of reason, and that can only be done in a formal framework. Nevertheless, we should not forget that formalisation, in spite of its undoubted importance, is still a tool, not an end. The art of mathematics goes far beyond the mere manipulation of symbols.

I will leave it there. An in-depth treatment of these issues is more suitable for a philosophy book. If you would like to get more insights on the philosophy of mathematics, I encourage you to read [1].

§4 ZFC Set theory

4.1 ZFC set theory. The standard axiomatisation of set theory is known as ZFC. The Z and the F stand, respectively, for the mathematicians Ernst Zermelo and Abraham Fraenkel. The C stands for the controversial axiom of choice, so, yes, there is an axiomatisation of set theory known as ZF that does not include this axiom.

The formal system ZFC in which we are going to formalise set theory uses first-order logic with equality and is built on the first-order language with set of symbols $S = \{=, \in, \varnothing\}$ and signature $\sigma(=) = -2$, $\sigma(\in) = -2$ and $\sigma(\varnothing) = 0$. Both $=$ and $\in$ are used with infix notation. If two elements x and y satisfy $x = y$ or $x \in y$, we will say, respectively that "x is equal to y" and that "x is an *element* of y". We can also say that "x is *contained* in y" to mean that $x \in y$. The constant $\varnothing$ will be referred to as the *empty set.* The elements of the domain of any model of ZFC are called *sets*.

For any two variables (sets) x and y, we introduce the notation $x \subseteq y$ as an abbreviation of the formula $(\forall a).\ a \in x \to a \in y$ for any variable a not occurring anywhere in the formula where $x \subseteq y$ is being used. If two sets x and y satisfy $x \subseteq y$, we say "x is *included* in y".

We will now present all the non-logical axioms of ZFC and analyse them from a semantic perspective.

(ZF1) Axiom of extensionality: $(\forall x)(\forall y).\ ((\forall a).\ (a \in x \leftrightarrow a \in y)) \to x = y$.

In layman's terms, the axiom of extensionality means that any two sets with the same elements are equal. In addition, using this axiom, (E4), 3.12 and the generalisation rule, one can easily deduce that

$$(\forall x)(\forall y).\ x = y \leftrightarrow ((\forall a).\ a \in x \leftrightarrow a \in y).$$

(ZF2) Axiom of the empty set: $(\forall x).\ \neg(x \in \varnothing)$.

The axiom of the empty set, as anyone should expect, simply states that $\varnothing$ has no elements. Furthermore, it follows from (ZF1) that any element verifying this property is equal to $\varnothing$.

(ZF3) Axiom of union: $(\forall \mathrm{X})(\exists \mathrm{U})(\forall x)(\forall a).\ (x \in \mathrm{X} \wedge a \in x) \to a \in \mathrm{U}$.

The axiom of union states that, given any set X, there exists a set U containing all the elements of the sets in X. Informally, this means that, given any collection of sets X, there exists a set that contains the union of all the sets in X.

(ZF4) Axiom of infinity: $(\exists \mathrm{X}).\ \varnothing \in \mathrm{X} \wedge (\forall x).$
$x \in \mathrm{X} \to (\exists y)(y \in \mathrm{X} \wedge (\forall a).\ a \in y \leftrightarrow (a = x \vee (\forall b)(b \in a \leftrightarrow b = x)))$.

I know what you are thinking and you are right; the axiom of infinity is a mess. There are simpler ways to present this axiom, but this is by far the most formal of all, and, you know, if we are going to do a formal treatment of set theory, let us do it properly! You and I are warriors, not soldiers.

If you take your time to analyse it, you will see that it postulates the existence of a set X having $\varnothing \in \mathrm{X}$ and verifying, for every $x \in \mathrm{X}$, $x \cup \{x\} \in \mathrm{X}$. Notice that we have not yet defined what $x \cup \{x\}$ means in ZFC, but we have already done an informal treatment of set theory and you should know that, with that, I (informally) mean "the set having as elements x and a set that only has x as an element".

You may wonder what the point of this axiom is. Turns out this will enable us to define the natural numbers and, from there on, the rationals, the reals...you name it!

(ZF5) Power set axiom: $(\forall x)(\exists \mathrm{P})(\forall a).\ a \subseteq x \to a \in \mathrm{P}$.

The power set axiom — which, after going through the axiom of infinity, looks ridiculously simple — simply establishes the existence, for any set x, of a set containing each subset of x, i.e., its power set.

(ZF6) Axiom of regularity: $(\forall x).\ \neg(x = \varnothing) \to (\exists y)(y \in x \land \neg(\exists z)(z \in y \land z \in x))$.

This axioms guarantees the existence in any non-empty set x of an element y containing no elements of x. To put it in perhaps clearer terms, it says that no set x can only consist of sets having elements of x.

The best way in which you can see why the axiom of regularity need be true is by trying to construct a set contradicting it.

(ZF7) Axiom schema of replacement: for any formula ψ in which there are only free occurrences of x and y and in which there are no quantifications over X and Y, $(\forall x)(\exists! y)\psi \to (\forall \mathrm{X})(\exists \mathrm{Y})(\forall y).\ y \in \mathrm{Y} \leftrightarrow (\exists x)(x \in \mathrm{X} \land \psi)$.

What the axiom of replacement tells us is that, if we are given a formula[2] $\psi(x, y)$ that, for each value of x, is true for one unique value of y — in other words, if $\psi(x, y)$ behaves like a function, — then, given any set X, there exists a set Y containing, exclusively, all the elements y for which there exists an $x \in \mathrm{X}$ verifying $\psi(x, y)$.

Putting it simpler terms, this means that if a “function” f taking any set x to a set y can be expressed as a formula $\mathrm{A}(x, y)$ that is true if and only if $f(x) = y$, then, for any set X, there exists a set containing the image of X under f. Of course, these formulas are not functions within set theory: a function within ZFC need be an object of ZFC and, therefore, a set, not a formula! We will later on define the concept of a function in ZFC in a precise and formal manner.

(ZF8) Axiom of choice: $(\forall \mathrm{X}).\ (\forall x)(x \in \mathrm{X} \to (\neg(x = \varnothing)\land$
$(\forall y).\ y \in \mathrm{X} \land \neg(x = y) \to \neg(\exists a).\ a \in x \land a \in y)) \to$
$((\exists \mathrm{S})(\forall s).\ s \in \mathrm{S} \to ((\exists x).\ x \in \mathrm{X} \land s \in x \land (\forall z)((\neg(z = a) \land z \in \mathrm{S}) \to \neg(z \in x))))$.

What this famous and controversial axiom of choice is telling us is something that is, well, obvious. It is simply stating that, given any collection X of non-empty sets that have no elements in common, there exists a set S containing one and only one element from every set in X. This assumption is extremely natural and so are the consequences that are derived from it — at least from my personal perspective — but the poor axiom of choice is rejected by some people (see [1] for reference). Please, give him some love and say with me “I choose choice!”.

By the way, if you were, by any chance, hoping that there could be a way to prove that either the axiom of choice or its negations would make ZF inconsistent and thus settle this debate of choice versus no choice once and for all...I have bad news. The axiom of choice is independent from ZF and, if ZF is consistent, so is ZF with the axiom of choice and so is ZF with the negation of the axiom of choice. So, at the end of the day, accepting or negating the axiom of choice is not a mathematical issue, but a philosophical one. This is one of those things that makes mathematics look more like a religion than like a science.

We shall now introduce a pretty solid collection of definitions and results. All those definitions and results that follow are done within ZFC unless otherwise stated.

4.2 Proposition. The following elementary properties of sets are true:

[2] We are using the informal notation $\psi(x, y)$ to represent an arbitrary formula ψ having two free variables x and y.

(i) Let X be a set. Both X and $\varnothing$ are subsets of X.
(ii) Two sets X and Y satisfy $X = Y$ if and only if $X \subseteq Y$ and $Y \subseteq X$.

Proof. (i) It is obvious that $X \subseteq X$. In regard to $\varnothing$, for it to be a subset of X, the formula $(\forall x)(x \in \varnothing \to x \in X)$ needs to be satisfied. Since, according to (ZF2), no set x can verify $x \in \varnothing$, the formula $x \in \varnothing \to x \in X$ is satisfied for any x and X. Thus, the empty set is a subset of any set.

(ii) Immediate from (ZF1) and from the definition of $\subseteq$.

■

4.3 Theorem (Schema of specification)**.** Let φ be a formula in the language of set theory with a single free variable. Given any set X, there exist the subset $Y \subseteq X$ of elements $x \in X$ verifying $\varphi(x)$. Such a set is described using *set-builder notation*: $Y = \{x \in X \mid \varphi(x)\}$.

Proof. If there are no elements $x \in X$ satisfying $\varphi(x)$, then $Y = \varnothing$, which certainly exists.

Let us then assume the existence of an element $y_0 \in Y$. We can consider the formula $\psi(x, y)$ given by

$$(\varphi(x) \wedge y = x) \vee (\neg\varphi(x) \wedge y = y_0).$$

It is clear that, for any set x, there exists a unique y satisfying $\psi(x, y)$: if $\varphi(x)$ holds, that y is x itself, whereas, if it does not, that y is y_0. Thus, we can apply (ZF7) to conclude that the collection Y' of all the elements y for which there exists an $x \in X$ verifying $\psi(x, y)$ is a set. Furthermore, this set is Y. Let us prove it by double inclusion, i.e., using 4.2(ii).

Given any $y \in Y$, the formula $\psi(y, y)$ holds and — since $Y \subseteq X$ and, therefore, $y \in X$ — we have $y \in Y'$, which shows that $Y \subseteq Y'$. Conversely, given any $y' \in Y'$, we there know to exist an $x \in X$ such that $\psi(x, y')$ holds. If $\varphi(x)$ is satisfied, then $y' = x$, so $y' \in Y$. If $\varphi(x)$ is not satisfied, then $y' = y_0$, which, by hypothesis, belongs to Y. In either case, $y' \in Y$, which shows that $Y' \subseteq Y$. ■

4.4. Do you remember when back in 1.1 we said that, in modern day set theory, we could identify predicates with sets — just as people wanted to do in the early days — provided we did it with care? The scheme of specification has just made that notion precise: given any set X we can identify each unary predicate φ with the set $\{x \in X \mid \varphi(x)\}$. What makes this approach different from the one that led to Russel's paradox is our requiring the "domain" over which we define φ to be a set complying with the axioms of ZFC, and that enables us to escape from any paradox.

We will shortly analyse some details regarding the schema of specification and how it avoids the paradoxes of primitive set theory.

4.5 Definition-Proposition. Let X and Y be any two arbitrary sets.

(i) There exists a set $\cup X$ containing, exclusively, all the elements of the sets contained in X. This set is known as the *union* of the elements of X.
(ii) There exists a set $\cap X$ containing, exclusively, the elements that belong to all the sets contained in X. This set is called the *intersection* of the elements of X.
(iii) There exists a set $X \setminus Y$ containing, exclusively, the elements of X that do not belong to Y. The set $X \setminus Y$ is known as the *difference* of X and Y. In particular, if $Y \subseteq X$, the set $X \setminus Y$ is said to be the *complement* of Y in X.

(iv) There exists a set $\mathscr{P}(X)$ containing, exclusively, all the subsets of X. This set is known as the *power set* of X.

Proof. According to (ZF3), there exists a set U that contains all the elements of the sets contained in X. Applying 4.3, it follows that

$$\cup X = \{u \in U \mid (\exists x \in X) u \in x\},$$
$$\cap X = \{u \in U \mid (\forall x \in X) u \in X\},$$

In regard to $X \setminus Y$, it suffices to use 4.3 again and consider $X \setminus Y = \{x \in X \mid x \notin Y\}$.

Lastly, regarding the power set, (ZF5) guarantees the existence of a set P containing every subset of X, so we just need to apply 4.3 once more and take

$$\mathscr{P}(X) = \{p \in P \mid p \subseteq X\}.$$

■

4.6 Proposition. Any element a of a set X is a set.

Proof. It suffices to apply 4.3 and 4.5(i) to conclude that $a = \cup\{x \in X \mid x = a\}$ is a set. ■

4.7 Proposition (Pairing)**.** Let a and b be sets. The set $\{a, b\}$ containing, exclusively, the elements a and b exists. In particular, if $a = b$, the *singleton* $\{a\}$ exists.

Proof. According to (ZF4), we there know to exist a set that contains, in particular, the element $X = \{\emptyset, \{\emptyset\}\}$. Thus, using 4.6, we know X to be a set. We can then consider the formula $\psi(x, y)$ defined by

$$(x = \emptyset \wedge y = a) \vee (x = \{\emptyset\} \wedge y = b) \vee (\neg(x \in X) \rightarrow x = y).$$

A direct application of the axiom schema of replacement (ZF7) shows that the set Y of elements y for which there exists an $x \in X$ verifying $\psi(x, y)$ is a set. Since $Y = \{a, b\}$, we have shown the set $\{a, b\}$ to exist. ■

4.8 Notation. Let X and Y be sets. By 4.7, $\{X, Y\}$ is a set. We define $X \cup Y = \cup\{X, Y\}$ and $X \cap Y = \cap\{X \cap Y\}$. If $X \cap Y = \emptyset$, we say that X and Y are *disjoint.* Furthermore, given any collection C of sets, we say that the sets in C are *pairwise disjoint* if, for any $X, Y \in C$ with $X \neq Y$, the sets X and Y are disjoint.

Let n be a natural number and let $X_1, \dots, X_n$ be sets. Through a recursive application of 4.7 and 4.5(i), we know the set X with elements $X_1, \dots, X_n$ to exist. This set is represented as $\{X_1, \dots, X_n\}$.

4.9. Now that we know singletons to exist, we can add a final touch on our discussion on the scheme of specification and Russel's paradox.

We can illustrate the robustness of our system with a simple example. If we consider *any* set X and try to recreate Russel's paradox by defining the set $R = \{x \in X \mid x \notin x\}$, we would have

$$R \in R \iff (R \in X \wedge R \notin R).$$

Does it look suspicious? It is actually harmless. Let A be any set. Applying the axiom of regularity (ZF6) to $\{A\}$, we know that $A \notin A$. Therefore, it is clear that $R = X$ and, therefore, that $R \notin X$. It then follows that both sides of the equivalence are false: everything fits nicely and we are paradox-free.

You may then wonder: and what if I take X to be a "universe" set containing all the sets in ZFC? Well, it is immediate from the axiom of regularity (ZF6) that such a set cannot exist in ZFC, so there is nothing to worry about.

4.10 Definition. Let x and y be sets. The *ordered pair* with *first coordinate* x and *second coordinate* y is the set

$$(x, y) = \{\{x\}, \{x, y\}\}.$$

Obviously, $(x, y) \neq (y, x)$ unless $y = x$, hence the name ordered pair. Moreover, if x_1, x_2, y_1, y_2 are some arbitrary sets, $(x_1, y_1) = (x_2, y_2)$ if and only if $x_1 = x_2$ and $y_1 = y_2$.

Let us assume that, for some sets X and Y, $x \in \mathrm{X}$ and $y \in \mathrm{Y}$. The ordered pair (x, y) belongs to the set $\mathscr{P}(\mathrm{X} \cup \mathrm{Y})$, so we can define the *binary cartesian product* of X and Y as the set

$$\mathrm{X} \times \mathrm{Y} = \{p \in \mathscr{P}(\mathrm{X} \cup \mathrm{Y}) \mid (\exists x)(\exists y).\ x \in \mathrm{X} \wedge y \in \mathrm{Y} \wedge p = (x, y)\}.$$

To put it simply, $\mathrm{X} \times \mathrm{Y}$ is the set of all ordered pairs (x, y) with $x \in \mathrm{X}$ and $y \in \mathrm{Y}$.

4.11. Just as every unary predicate (unary relation) on a set X was characterised by a subset of X, every binary relation on two sets X and Y will be characterised by a subset of X×Y. The ability to work with relations in ZFC will enable us to introduce functions as mere set-theoretical objects, and, in return, functions will enable us — among many other things — to define n-ary cartesian products and, therefore, to introduce n-ary relations in set theory.

4.12 Definition. Let X and Y be sets. A *binary relation* over the sets X and Y is a subset R of the cartesian product $\mathrm{X} \times \mathrm{Y}$. If a subset $\mathrm{R} \subseteq \mathrm{X} \times \mathrm{Y}$ is being regarded as a binary relation and not as a mere subset, it is customary to write $x\mathrm{R}y$ in lieu of $(x, y) \in \mathrm{R}$.

Any formula φ with a free variable can induce a binary relation according to 4.3, for it suffices to consider $\mathrm{R} = \{(x, y) \in \mathrm{X} \times \mathrm{Y} \mid \varphi((x, y))\}$. This kind of definition is often done in an implicit manner as "let R be the relation on X and Y that is satisfied if and only if $\varphi((x, y))$".

Very often, relations will be subsets of X×X. In this case, we say that R is a binary relation over X: it would be redundant to say that it is a binary relation over X and X.

Writing $x\mathrm{R}y$ might look a little bit confusing, that is why binary relations are often represented by fancy symbols such as $\sim$ or $\equiv$. In addition, some relations may use a more complex notational artefact than just the structure "element symbol element". As always, if there is something inexact and non-universal about mathematics, that is its notation!

4.13 Example. (i) Let X be any set. The empty relation $\mathrm{R} = \emptyset$ is a binary relation over X. It is, obviously, not satisfied by any $(a, b) \in \mathrm{X} \times \mathrm{X}$.

(ii) Let X be any set. The subset $\mathrm{R} = \{(a, b) \in \mathrm{X} \times \mathrm{X} \mid a = b\}$ is a binary relation over X. The relation R can equivalently be defined as the binary relation on X such that $a\mathrm{R}b$ if and only if $a = b$.

(iii) Let X and Y be sets. The set $\mathrm{R} = \mathrm{X} \times \mathrm{Y}$ is a binary relation over X and Y. Of course, any $x \in \mathrm{X}$ and $y \in \mathrm{Y}$ verify $x\mathrm{R}y$.

4.14 Definition. A *function* f from a set X to a set Y is a binary relation over X and Y such that, for every $x \in \mathrm{X}$, there exists a unique $y \in \mathrm{Y}$ verifying xfy. Instead of writing xfy, we will use $f(x) = y$. The fact that f is a function from X to Y is denoted by $f : \mathrm{X} \longrightarrow \mathrm{Y}$. In addition, $f(x) = y$ can also be written as $f : x \longmapsto y$.

Given $f : \mathrm{X} \longrightarrow \mathrm{Y}$, the set X is said to be the *domain* of f ($\mathrm{X} = \operatorname{dom} f$) whereas Y is known as the *codomain* of the function. The *image* $\operatorname{im} f$ of f is the subset of Y containing the elements $y \in \mathrm{Y}$ for which there exists an $x \in \mathrm{X}$ such that $f(x) = y$, i.e.,

$$\operatorname{im} f = \{y \in \mathrm{Y} \mid (\exists x \in \mathrm{X})(y = f(x)\},$$

where, as in 0-3.8, $(\exists x \in \mathrm{X})\theta(x)$ means $(\exists x)(x \in \mathrm{X} \to \theta(x))$.

We say that f is injective if, for every $y \in \operatorname{im} f$, there exists a unique x such that $f(x) = y$, or if, equivalently, for every $a, b \in \mathrm{X}$, $f(a) = f(b)$ implies $a = b$. In addition, f is said to be *surjective* if $\operatorname{im} f = \mathrm{Y}$. If a function is both injective and surjective, it is said to be *bijective.*

Given any $\mathrm{A} \subseteq \mathrm{X}$, the *image* under f of A is the set

$$f[\mathrm{A}] = \{y \in \mathrm{Y} \mid (\exists a \in \mathrm{A}).\ f(a) = y\}.$$

Notice how $\operatorname{im} f = f[\mathrm{X}]$. Similarly, given any $\mathrm{B} \subseteq \mathrm{Y}$, the *inverse image* of B under f is the set

$$f^{-1}[\mathrm{B}] = \{x \in \mathrm{X} \mid f(x) \in \mathrm{B}\}.$$

If f is injective, we can define a function $f^{-1} : \operatorname{im} f \longrightarrow \mathrm{X}$ mapping every $y \in \mathrm{Y}$ to $\cup f^{-1}[\{y\}] \in \mathrm{X}$, i.e., to the only element $x \in \mathrm{X}$ satisfying $f(x) = y$.

Given any set X, the *identity* function on X is the function $\mathrm{id}_{\mathrm{X}} : \mathrm{X} \longrightarrow \mathrm{X}$ taking any $x \in \mathrm{X}$ to itself: $f(x) = x$.

Given any two functions $f : \mathrm{X} \longrightarrow \mathrm{Y}$ and $g : \mathrm{Y} \longrightarrow \mathrm{Z}$, the *composition* of f with g is the function

$$\begin{aligned} g \circ f : \mathrm{X} &\longrightarrow \mathrm{Z} \\ x &\longmapsto g(f(x)). \end{aligned}$$

Lastly, given a subset $\mathrm{A} \subseteq \mathrm{X}$ and a function $f : \mathrm{X} \longrightarrow \mathrm{Y}$, we define the *restriction* of f to A as the function

$$\begin{aligned} f|_{\mathrm{A}} : \mathrm{A} &\longrightarrow \mathrm{Y} \\ a &\longmapsto f(a). \end{aligned}$$

4.15 Proposition. Let $f : \mathrm{X} \longrightarrow \mathrm{Y}$ and $g : \mathrm{Y} \longrightarrow \mathrm{Z}$ be functions.

(i) The function $g \circ f$ is injective only if f is injective.
(ii) The function $g \circ f$ is surjective only if g is surjective.

Proof. We will prove the contrapositive of each of the statements.

(i) If f is not injective, there exist $a, b \in \mathrm{X}$ such that $a \neq b$ and $f(a) = f(b)$; consequently, $g \circ f(a) = g \circ f(b)$. Thus, $g \circ f$ is not injective.

(ii) Analogously, if g is not surjective, no $y \in \mathrm{Y}$ satisfies $f(y) = z_0$ for some $z_0 \in \mathrm{Z}$. Thus, given any $x \in \mathrm{X}$, since $f(x) \in \mathrm{Y}$, we have $g \circ f(x) \neq z_0$.

■

4.16 Proposition. Let $f : X \longrightarrow Y$. The function f is bijective if and only if there exists a function $g : Y \longrightarrow X$ such that $f \circ g = \mathrm{id}_X$ and $g \circ f = \mathrm{id}_Y$.

Proof. If f is bijective, the function g we are looking for is $f^{-1} : \operatorname{Im} f = Y \longrightarrow X$. It is immediate that $f \circ f^{-1} = \mathrm{id}_X$ and that $f^{-1} \circ f = \mathrm{id}_Y$.

Conversely, since both $\mathrm{id}_X = f \circ g$ and $\mathrm{id}_Y = g \circ f$ are bijective, the result follows from 4.15. ■

4.17 Definition-proposition. We say that a set X is *inductive* if it verifies the formula $\Omega(X)$ given by

$$\varnothing \in X \wedge (\forall x)(x \in X \to x \cup \{x\} \in X),$$

where, for the sake of clarity, we have used $x \cup \{x\} \in X$ in lieu of

$$(\exists y).\ y \in X \wedge (\forall t).\ t \in y \leftrightarrow (t = x \vee (\forall s)(s \in t \leftrightarrow s = x)).$$

There exists a *minimal* inductive set ω, i.e., an inductive set that is included in every inductive set.

Proof. By (ZF4), we know an inductive set V to exist. The set ω can be constructed using 4.3 as

$$\omega = \{v \in V \mid (\forall X)(\Omega(X) \to v \in X)\}.$$

In plain English, ω is the set containing, exclusively, the elements that are common to all inductive sets. Informally, we could have defined it as the "intersection of all the inductive sets"; nonetheless, since we have not proven the existence of such thing as a set of all inductive sets, we have had to create this custom definition.

The set ω is clearly inductive. On the one hand, any inductive set X verifies $\varnothing \in X$, so $\varnothing \in \omega$. On the other hand, if $v \in \omega$, then v belongs to any inductive set X, and, by definition, so does $v \cup \{v\}$. Thus, $v \cup \{v\} \in \omega$.

Lastly, it is trivial that ω is included in any inductive set for the elements in ω belong, by definition, to any inductive set. ■

4.18. If set theory has any intention of becoming a foundational system for mathematics, it better let us work with natural numbers! Now, how could we possibly implement natural numbers in ZFC? How could we construct a set of natural numbers? The answer is, surprise surprise, we have already done it! The set ω is the set we have been looking for! Informally, if we let $0 = \varnothing \in \omega$, we can define

$$1 = \{0\} \in \omega, \qquad 2 = \{0, 1\} \in \omega, \qquad 3 = \{0, 1, 2\} \in \omega,$$

and so on. In general, given any natural number n representing a set $n \in \omega$, we define its successor as $s(n) = n \cup \{n\} = \{1, \dots, n\} \in \omega$.

In this construction, the set of natural numbers would be $\mathbb{N} = \omega \setminus \{\varnothing\}$.

I know. This looks weird. Having $4 \subseteq 5$ seems like an odd property; nonetheless, as you will later see, these oddities will not have any visible effect in our daily-life arithmetic. In fact, these very oddities will prove themselves very useful in enabling us to properly incorporate number systems into our beautiful foundational theory.

From now on, unless otherwise stated, we will use Arabic numerals in order to refer to their corresponding elements in ω. This is just notation.

Shortly, we will define the basic operations and relations in ω, but we need to go through a very significant result before we can get there.

4.19 Theorem (Principle of recursive definition)**.** Let X be a set with $a \in X$. Let $u : X \longrightarrow X$ be a function. There exists a unique function $f : \omega \longrightarrow X$ verifying $f(0) = a$ and $f(s(n)) = u(f(n))$.

Proof. We will prove this theorem by explicitly constructing f as a set of ordered pairs in $\omega \times X$.

For that purpose, let us consider the set F of all subsets $S \subseteq \omega \times X$ such that $(0, a) \in S$ and such that, whenever $(n, x) \in S$, then $(s(n), u(x)) \in S$. The set F is clearly non-empty for $\omega \times X \in F$.

We now take $f = \cap F$. It should be clear that $f \in F$ and that f is included in every element of f. Thus, if we showed f to be a function, we would have proved the result. The reasoning behind this is simple. If there existed another function g verifying the conditions of the theorem, $g \in F$ and, therefore, $f \subseteq g$; nonetheless, assuming that f is a function from ω, g can only be a function from ω if it is equal to f, because if it contained an additional ordered pair, we would have, for a certain $n \in \omega$ and $x, y \in X$ with $x \neq y$, both $(n, x) \in g$ and $(n, y) \in g$, which would mean that g would not be a function.

Let us then show f to be function! We can consider the set $N \subseteq \omega$ of elements $n \in \omega$ for which there exists a unique ordered pair in f with first coordinate n.

If we had $0 \notin N$, then there would exist an element of the form $(0, b) \in f$ with $b \neq a$. Nonetheless, then $f' = f \setminus \{(0, b)\} \in F$ with $f' \subset f$, which would contradict our hypothesis that $f = \cap F$. Analogously, If we had, for some $n \in \omega$, $n \in N$ but $s(n) \notin N$, then there would exist a unique $x \in X$ such that $(n, x) \in f$ and there would exist an $y \in X$ distinct from $u(x)$ such that $(s(n), y) \in f$. Were this the case, then it is trivial that $f \setminus \{(s(n), y)\} \in F$ and we would once again reach a contradiction.

Since ω is the smallest inductive set and we have shown that, for a set $N \subseteq \omega$, we have $0 \in N$ and, whenever $n \in N$, $s(n) \in N$, we can conclude *by induction* that $N = \omega$. This completes the proof. ■

4.20 Definition. Let $m \in \omega$. Applying the principle of recursive definition, we can define two functions s_m and p_m from ω to ω verifying, on the one hand, $s_m(0) = m$ and $s_m(s(n)) = s(s_m(n))$, and, on the other hand, $p_m(0) = 0$ and $p_m(s(n)) = s_{p_m(n)}(n)$.

This enables us to define *addition* $+$ and *multiplication* $\cdot$ in ω as

$$\begin{aligned} + : \omega \times \omega &\longrightarrow \omega \\ (x, y) &\longmapsto x + y = s_x(y), \end{aligned} \qquad \begin{aligned} \cdot : \omega \times \omega &\longrightarrow \omega \\ (x, y) &\longmapsto x \cdot y = p_x(y), \end{aligned}$$

where, of course, we have used infix notation.

Just to complete our definitions in ω, let us the relation $\leq$ in ω as $x \leq y$ if and only if $x \in y$ or $x = y$. Analogously, we define a relation $<$ such that $x < y$ if and only if $x \in y$. As was to be expected, if $x \leq y$ we say that x is *smaller than or equal to* y and if $x < y$, it is said that x is *smaller than* y.

4.21 Definition. Let X be a set. An *internal n-ary law of composition* on X is a function $\star : \times_{i=1}^{n} X \longrightarrow X$. The adjective "internal" is often dropped. Furthermore, for binary laws of composition, the adjective "binary" is often omitted too. Laws of compositions are commonly referred to as operations.

Infix notation is often used with laws of composition: thus, $*(x, y)$ is written as $x * y$. We say that a law of composition $*$ is

- *associative* if, for any $x, y, z \in X$, we have $(x * y) * z = x * (y * z)$,
- *commutative* if, for any $x, y \in X$, we have $x * y = y * x$,
- *left distributive* with respect to another law of composition $\Box$ if, for any $x, y, a \in X$, $a * (x \Box y) = (a * x) \Box (a * y)$,
- *right distributive* with respect to another law of composition $\Box$ if, for any $x, y, a \in X$, $(x \Box y) * a = (x * a) \Box (y * a)$,
- and *distributive* with respect to another law of composition if it is both left distributive and right distributive.

Laws of composition often use *additive notation* (i.e., use the symbol $+$ and infix notation) or *multiplicative notation* (i.e., use the symbol $\cdot$ and infix notation). In multiplicative notation, the use of $\cdot$ may be replaced by juxtaposition: thus, xy would be read as $x \cdot y$.

Clearly, the operations $+$ and $\cdot$ that we have defined on ω are laws of composition using, respectively, additive and multiplicative notation.

4.22 Lemma. Let $n \in \omega$. The following seemingly irrelevant statements are true:

(i) If $m \in n$, then $n \nsubseteq m$. In other words, no element of ω is a subset of any of its elements.
(ii) The set ω is transitive: if $m \in n$, then $m \subseteq n$.

Proof. (i) Let $N \subseteq \omega$ be the set of elements $n \in \omega$ such that, for any $m \in n$, we have $n \nsubseteq m$. It is clear that $0 \in N$ for $0 = \emptyset$. In addition, if $n \in N$, we can easily show that $s(n) = n \cup \{n\} \in N$.

The elements of $n \cup \{n\}$ are those of n and n itself. Can $n \cup \{n\}$ be a subset of n? For that to be the case, we would need to have $\{n\} \subseteq n$ and, therefore, $n \in n$, but, as $n \in N$ and $n \subseteq n$, it is clear that $n \notin n$. Can $n \cup \{n\}$ be a subset of $x \in n$? Since $n \subseteq n \cup \{n\}$, if this were the casee, n would also be a subset of x thus contradicting the hypothesis that $n \in N$. We can conclude that $s(n) \in N$ and, by induction, that $N = \omega$.

(ii) We proceed, again, inductively. Let $N \subseteq \omega$ be the set of elements n for which the result holds. It is out of question that $0 \in N$. In addition, if we assume $n \in N$, we can easily prove that $s(n) \in N$. The elements of $s(n)$ are n and all the elements of n. Clearly, since all the elements of n belong to $s(n)$, we have $n \subseteq s(n)$. Moreover, according to the inductive hypothesis, all the elements $x \in n$ are subsets of n, and, since $n \subseteq s(n)$, they must also be subsets of $s(n)$.

■

4.23 Theorem. Let us consider the interpretation of the formal system PA of Peano Arithmetic using $\mathbb{N}$ as domain and having $\{\emptyset\} = 1$ as $\underline{1}$; the set-theoretic relation $=$ as $\underline{=}$; the set-theoretic function s as $\underline{s}$, and the set-theoretic functions $+|_{\mathbb{N}\times\mathbb{N}}$ and $\cdot|_{\mathbb{N}\times\mathbb{N}}$ as $\underline{+}$ and $\underline{\cdot}$ respectively. This interpretation is a model.

Consequently, if ZFC is consistent, so is PA.

Proof. We need to show that, in the interpretation of PA that we have constructed, all the axioms are true. The axioms of equality are clearly verified, so we can get to work with the remaining non-logical axioms.

It is trivial that (PA1), (PA3), (PA4), (PA5) and (PA6) are true by how we have defined the elements of the interpretation. Furthermore, the principle of mathematical induction captured in (PA7) trivially holds by considering, for any property $\varphi(x)$, the set $\{x \in \omega \mid \varphi(x)\}$ and noting that ω is inductive.

The only axiom that remains to be shown to be true is (PA2). Let us assume that two elements $m, n \in \omega$ verify $s(m) = s(n)$. Under these conditions,

$$m \in s(m) = s(n) = n \cup \{n\} \quad \text{and} \quad n \in s(n) = s(m) = m \cup \{m\},$$

so either $m = n$ or both $m \in n$ and $n \in m$. In the latter case, we have $m \in n \in m$, but, applying 4.22(ii), this implies that $n \subseteq m$ and, therefore, yields a contradiction with 4.22(i). ■

4.24. The following statements about ω can be shown to be true in any model of ZFC in an identical way to their analogues in $\mathbb{N}$:

(ω1) $(\forall x \in \omega).\ \neg(s(x) = 0)$.
(ω2) $(\forall x, y \in \omega).\ s(x) = s(y) \to x = y$.
(ω3) $(\forall x \in \omega).\ x + 0 = x$.
(ω4) $(\forall x, y \in \omega).\ x + s(y) = s(x + y)$.
(ω5) $(\forall x \in \omega).\ x \cdot 0 = 0$.
(ω6) $(\forall x, y \in \omega).\ x \cdot s(y) = x \cdot y + x$.
(ω7) $(\forall \mathrm{A} \subseteq \omega).\ (0 \in \omega \wedge (\forall n \in \omega)(s(n) \in \omega) \to \mathrm{A} = \omega)$.

It goes without saying that (ω7) implies that any formula $\mathrm{P}(x)$ is true for any $x \in \omega$ if proved to hold for $x = 0$ and, given an arbitrary $n \in \omega$, for $x = s(n)$ assuming $\mathrm{P}(n)$.

4.25 Lemma. Assuming addition in ω to be associative, the following statements about ω are true:

(i) $(\forall x \in \omega).\ 0 + x = 0$.
(ii) $(\forall x, y \in \omega).\ s(x) + y = s(x + y)$.

Proof. (i) Let us show that $0 + n = 0$ for every $n \in \omega$ by induction on n. For $n = 0$, the result is immediate from (ω3). If we now assume this property to hold for an arbitrary $n \in \omega$, it follows from (ω3) and (ω4) that it is verified by $s(n)$ as

$$0 + s(n) = 0 + s(n + 0) = 0 + n + s(0) = n + s(0) = s(n).$$

Thus, the result is true by the principle of mathematical induction, i.e., by (ω7). Notice how we have also made an implicit use of the associativity hypothesis.

(ii) We shall prove that $s(x) + n = s(x + n)$ for any $x, n \in \omega$ using induction on n. The result $n = 0$ is a direct consequence of (ω3). Assuming it to hold for an arbitrary $n \in \omega$, we can deduce from (ω3) and (ω4) that

$$\begin{aligned} s(x) + s(n) &= s(x) + s(n + 0) = s(x) + n + s(0) = s(x + n) + s(0) \\ &= (x + s(n)) + s(0) = s((x + s(n)) + 0) = s(x + s(n)), \end{aligned}$$

and, therefore, the result is true by induction.

■

4.26 Proposition. Addition $+$ on ω is associative and commutative.

Proof. We shall first prove associativity: we will prove, by induction on $n \in \omega$, that $(x + y) + n = x + (y + n)$. For $n = 0$, the result is obvious since, according to (ω3),

$$(x + y) + 0 = x + y = x + (y + 0).$$

Let us then assume associativity to hold for an arbitrary $n \in \omega$. Regarding $s(n)$, the inductive hypothesis together with a repeated application of (ω4) yields

$$\begin{aligned}(x + y) + s(n) &= s((x + y) + n) = s(x + (y + n)) = x + s(y + n)\\ &= x + (y + s(n)).\end{aligned}$$

Now that we have associativity assured, let us deal with commutativity. We will show that any $x, n \in \omega$ verify $x + n = n + x$ using, once again, induction on n. The property is obvious for $n = 0$ as $x + 0 = x$ and $0 + x = x$. These equalities are direct applications of (ω3) and 4.25(i). Assuming as inductive hypothesis that $x + n = n + x$, we have

$$x + s(n) = s(x + n) = s(n + x) = s(n) + x.$$

The fact that $s(n + x) = s(n) + x$ was shown in 4.25(ii). ■

4.27 Lemma. The following statements about ω are true:

(i) $(\forall x \in \omega).\ 0 \cdot x = 0$.
(ii) $(\forall x, y \in \omega).\ s(x) \cdot y = x \cdot y + y$.

Proof. (i) As you should have expected, we proceed by induction on x. It is clear that $0 \cdot 0 = 0$ from (ω5). If we now assume the result to hold for an arbitrary x, it follows from (ω6) and (ω3) that

$$0 \cdot s(x) = 0 \cdot x + 0 = 0 + 0 = 0.$$

(ii) Once again, we proceed by induction on y. The result is immediate for $y = 0$ and, assuming it to hold for an arbitrary $y \in \omega$, we have

$$s(x) \cdot s(y) = s(x) \cdot y + s(x) = x \cdot y + y + s(x) = x \cdot s(y) + s(y),$$

which shows that the result holds for $s(y)$.

■

4.28 Proposition. Multiplication $\cdot$ on ω is distributive over addition, associative and commutative.

Proof. We will first prove left distributivity and use it in the proof of associativity and commutativity. Then, right distributivity will follow from commutativity.

We need to show that, for every $n, x, y \in \omega$, we have $n \cdot (x + y) = n \cdot x + n \cdot y$. As our notation suggests, we will proceed by induction on n. Using 4.27(i), the base case $n = 0$ for distributivity is trivial as $0 \cdot (x + y) = 0 = 0 \cdot x + 0 \cdot y$. Assuming left distributivity to hold for an arbitrary $n \in \omega$, we have

$$\begin{aligned}s(n) \cdot (x + y) &= n \cdot (x + y) + (x + y) = n \cdot x + n \cdot y + x + y\\ &= s(n) \cdot x + s(n) \cdot (y),\end{aligned}$$

where we have made implicit use of properties such as the associativity and commutativity of addition, 4.27(ii) and (ω6).

In regard to associativity, let us show that $x \cdot (y \cdot n) = (x \cdot y) \cdot n$ by induction on n. For $n = 0$, it follows from (ω5) that

$$x \cdot (y \cdot 0) = x \cdot 0 = 0 = (x \cdot y) \cdot 0.$$

Assuming associativity to hold for an arbitrary value of $n \in \omega$, we have, using left distributivity

$$\begin{aligned} x \cdot (y \cdot s(n)) &= x \cdot (y \cdot n + y) = x \cdot (y \cdot n) + x \cdot y = (x \cdot y) \cdot n + x \cdot y \\ &= (x \cdot y) \cdot s(n), \end{aligned}$$

where we have also used (ω6).

Lastly, regarding commutativity, let us prove that $x \cdot n = n \cdot x$ for any by induction on n. The result is trivial for $n = 0$ and, assuming it to hold for an arbitrary n,

$$x \cdot s(n) = x \cdot n + x = n \cdot x + x = s(n) \cdot x,$$

where we have made use of (ω6) and 4.27(ii). ■

4.29. What we have just done is quite significant. We have proven some facts about + and · without making any direct reference to the way in which they are defined: we have only used a collection of assumptions.

What we have seen in these proofs is an example of how the axiomatic method is used "in practice" within ZFC.

4.30 Definition. Let I and X be sets. A function $x : \mathrm{I} \longrightarrow \mathrm{X}$ may be regarded as a *family* of elements of X *indexed* by the set I. If that is the case, given $i \in \mathrm{I}$, we write x_i instead of $x(i)$ and we denote the function x by $\{x_i\}_{i \in \mathrm{I}}$ or by $\{x_i\}_i$ for short. On some occasions, the context might make it acceptable to also use the notation $\{x_i\}$.

Given a family $\{\mathrm{X}_i\}_{i \in \mathrm{I}}$ of sets, its union is defined as $\cup_{i \in \mathrm{I}} \mathrm{X}_i = \cup \operatorname{im} \mathrm{X}$. If $\mathrm{I} \neq \varnothing$, we define its intersection as $\cap_{i \in \mathrm{I}} \mathrm{X}_i = \cap \operatorname{Im} \mathrm{X}$. Lastly, the cartesian product of the family is the set $\times_{i \in \mathrm{I}} \mathrm{X}_i$ of families $\{x_i\}_{i \in \mathrm{I}}$ such that, for every $i \in \mathrm{I}$, $x_i \in \mathrm{X}_i$; or, in more symbolic terms,

$$\boxtimes_{i \in \mathrm{I}} \mathrm{X}_i = \left\{ x : \mathrm{I} \longrightarrow \bigcup_{i \in i} \mathrm{X}_i \;\middle|\; (\forall i \in \{1, \dots, n)(x_i \in \mathrm{X}_i) \right\}.$$

If the context allows for it, we may write $\cup_i$, $\cap_i$ and $\times_i$ instead of $\cup_{i \in \mathrm{I}}$, $\cap_{i \in \mathrm{I}}$ and $\times_{i \in \mathrm{I}}$. Naturally, a family $\{\mathrm{X}_i\}_{i \in \mathrm{I}}$ of sets is said to be pairwise disjoint if, for any $i, j \in \mathrm{I}$ with $i \neq j$, $\mathrm{X}_i \cap \mathrm{X}_j = \varnothing$. Notice how we are only requiring that $i \neq j$, not that $\mathrm{X}_i \neq \mathrm{X}_j$.

If the index set of a family $\{x_i\}$ is $\mathbb{N}$, the family is said to be a *sequence* of elements in X.

Of particular interest is the case where the index set is $\mathrm{I} = \{1, \dots, n\} \subseteq \mathbb{N}$. In this situation, a family $\{x_i\}_{i \in \mathrm{I}}$ is said to be an *n-tuple* and it is represented writing down its terms explicitly as $(x_1, \dots, x_n)$. Notice that tuples generalise ordered pairs. If $(\mathrm{X}_1, \dots, \mathrm{X}_n)$

is a tuple of sets, we can write

$$\bigcup_{i\in I} X_i = \bigcup_{i=1}^{n} X_i = X_1 \cup \cdots \cup X_n, \qquad \bigcap_{i\in I} X_i = \bigcap_{i=1}^{n} X_i = X_1 \cap \cdots \cap X_n,$$

$$\underset{i\in I}{\boxtimes} X_i = \overset{n}{\underset{i=1}{\boxtimes}} X_i = X_1 \times \cdots \times X_n.$$

Given any tuple of sets $(X_1, \dots, X_n)$, its cartesian product $X_1 \times \cdots \times X_n$ is, according to the definition, the set of all tuples $(x_1, \dots, x_n)$ having, for every $i \in \{1, \dots, n\}$, $x_i \in X_i$. The subsets of these cartesian products enable us to generalise the binary relations introduced in 4.12 to n-ary relations.

While it might seem awkward to have two distinct objects representing the same concepts (2-tuples vs ordered pairs and 2-ary relations vs binary relations), this redundancy has no noticeable consequences in practice. The different implementations of these concepts behave identically, so we might as well ignore their very nature. In addition, we will use their different denominations interchangeably.

4.31 Definition. Let X be a set. A binary relation $\sim$ defined over X is said to be an *equivalence relation* if it satisfies the following properties:

- Reflexivity: for every $a \in X$, we have $a \sim a$.
- Symmetry: for every $a, b \in X$, if $a \sim b$ then $b \sim a$.
- Transitivity: for every $a, b, c \in X$, if $a \sim b$ and $b \sim c$, then $a \sim c$.

Given any $a \in X$, we define its *equivalence class* as the set $[a] = \{x \in X \mid a \sim x\}$. The *quotient set* $X/\!\sim$ of X by $\sim$ is the set of all the equivalence classes in X.

4.32 Definition. Given a set X, a *partition* P of X is a collection of pairwise disjoint subsets of X such that $\cup P = X$.

4.33 Proposition. Let X be a set.

(i) If $\sim$ is an equivalence relation over X, the quotient set $X/\sim$ is a partition of X.
(ii) If P is a partition of X, then the equivalence relation $\sim$ defined in such a way that $x \sim y$ if and only if x and y belong to the same set in P is an equivalence relation.

Proof. (i) It is clear from reflexivity that, given any $a \in X$, we have $a \in [a]$. Thus, it is obvious that $\cup(X/\!\sim) = X$, so we just need to show that any two distinct equivalence classes are disjoint.

Let $a, b \in X$ and let us assume the existence of an element $x \in [a] \cap [b]$. Under these assumptions, we know that $x \sim a$ and $x \sim b$ and, therefore, by transitivity, $a \sim b$. Given any $x \in [a]$, we will have $x \sim a$, so, by transitivity, $x \sim b$ and, therefore, $x \in [b]$. This proves that $[a] \subseteq [b]$ and it can be shown analogously that $[b] \subseteq [a]$. Thus, two sets in the quotient set can only have a non-empty intersection if they are the same set, i.e., the elements of $X/\!\sim$ are pairwise disjoint.

Notice, by the way, how we have been constantly using the symmetry of equivalence relations throughout the proof.

(ii) Symmetry and reflexivity are obvious. Regarding transitivity, let $a, b, c \in X$ be such that $a \sim b$ and $a \sim c$. If we let $A \in P$ be the only subset of the partition to which a

belongs, as $a \sim b$, we know that $b \in \mathrm{A}$. In addition, since $b \sim c$, that must mean that $c \in \mathrm{A}$ and, therefore, that $a \sim c$. This proves that $\sim$ is an equivalence relation.

■

4.34 Definition. Let X be a set. A binary relation $\preceq$ defined on X is said to be a *partial order* if it verifies the following properties:

- Reflexivity: for any $a \in \mathrm{X}$, we have $a \preceq a$.
- Antisymmetry: for any $a, b \in \mathrm{X}$, if $a \preceq b$ and $a \neq b$, then $b \npreceq a$.
- Transitivity: for any $a, b, c \in \mathrm{X}$, if $a \preceq b$ and $b \preceq c$, then $a \preceq c$.

A partial order is said to be a *total order* if, for any $a, b \in \mathrm{X}$, we have $a \preceq b$ or $b \preceq a$.

A *strict partial order* $\prec$ is a binary relation that verifies antisymmetry, transitivity and *irreflexivity*, i.e., that for any $a \in \mathrm{A}$, $a \nprec a$. Any partial order $\preceq$ can be used to define a strict partial order $\prec$ as $a \prec b$ if and only if $a \preceq b$ and $a \neq b$. Conversely, any strict partial order defines a partial order in the obvious way.

Given a set X endowed with a partial order $\preceq$, an element $a \in \mathrm{A}$ is said to be a *minimal* element if there exists no $x \in \mathrm{X}$ such that $x \preceq a$. Analogously, a is said to be *maximal* if no $x \in \mathrm{X}$ exists such that $a \preceq x$. The element a is said to be a *minimum* in X if, for any $x \in \mathrm{X}$, $a \preceq x$; moreover, it is said to be a *maximum* if any $x \in \mathrm{X}$ satisfies $x \preceq a$. Notice how all these definitions need to be understood with respect to a particular partial order $\preceq$: an element may be maximal, minimal, a maximum or a minimum with respect to one partial order but not to another.

A total order $\preceq$ on a set X is said to be a *well-order* if any $\mathrm{A} \subseteq \mathrm{X}$ has a minimum. If a set X has a well-order $\preceq$, it is said that X is *well-ordered* (by $\preceq$).

Exercises

1) The inverse image of functions has some truly nice and lovely properties: when working with functions and set operations, the inverse image is the mathematical equivalent of the love of your life. To see what I mean, let $f : \mathrm{X} \longrightarrow \mathrm{Y}$ and let $\mathrm{A}, \mathrm{B} \subseteq \mathrm{X}$. Prove that

$$f^{-1}[\mathrm{A} \cup \mathrm{B}] = f^{-1}[\mathrm{A}] \cup f^{-1}[\mathrm{B}],$$
$$f^{-1}[\mathrm{A} \cap \mathrm{B}] = f^{-1}[\mathrm{A}] \cap f^{-1}[\mathrm{B}],$$
$$f^{-1}[\mathrm{Y} \setminus \mathrm{A}] = \mathrm{X} \setminus f^{-1}[\mathrm{A}].$$

Can you deduce similar properties for the image of f?[3]

2) Find a counter-example for the converses of the statements in 4.15.

3) Let us regard $\iff$ as a relation between formulas of a first-order formal system: the relation $\iff$ holds between two formulas A and B if and only if A $\iff$ B. Prove that this relation is reflexive, symmetric and transitive.

Analogously, show that the relation induced by $\implies$ is transitive and reflexive but not necessarily symmetric.

[3] That was a rhetorical question; of course you can. Please, do it.

4) Let P be a formula in a first-order language. Prove that, if $(\exists x)\mathrm{P}$ is true in an interpretation, there exists an assignment of values α in that interpretation such that $v_\alpha(\mathrm{P}) = 1$.

5) Let $\mathrm{A}_1, \dots, \mathrm{A}_m$ and $\mathrm{B}_1, \dots, \mathrm{B}_n$ be propositional forms for $m, n \in \mathbb{N}$. Prove that

$$\vdash (\mathrm{A}_1 \wedge \cdots \wedge \mathrm{A}_m) \to (\mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_m)$$

if and only if, for every $i \in \{1, \dots, n\}$, $\mathrm{A}_1, \dots, \mathrm{A}_m \vdash \mathrm{B}_i$.

6) The following statement is an alternative formulation of the axiom of choice: for every collection of sets X, there exists a *choice function*

$$f : \mathrm{X} \longrightarrow \bigcup \mathrm{X}$$

assigning, to every $x \in \mathrm{X}$, an element $a \in x$. Prove that this formulation is equivalent to the one we have used.

7) Show that the equality relation in any formal system using logic with equality verifies reflexivity, symmetry and transitivity.

8) Show that the axiom schema (P3) is independent in the formal system of propositional logic.

9) This exercise is a constructivist's nightmare. Use the axiom of choice to prove the existence of a sequence of natural numbers *without constructing a sequence.*

Appendices

§A Additional results in formal logic

A.1 Lemma. Let A and B be any propositional forms. The formulas

(i) $\neg A \to (A \to B)$,
(ii) $A \to \neg\neg A$,
(iii) $A \to (\neg B \to \neg(A \to B))$,
(iv) $(A \to B) \to ((\neg A \to B) \to B)$,

are theorems of the formal system P of propositional logic.

Proof. What follows are some tedious, long and boring mechanical proofs. It is hard to say that you will gain anything from carefully reading them other than convincing yourself that this lemma is true. You have been advised. Proceed at your own discretion.

(i) It is immediate that $A, (A \to B), (B \to C) \vdash C$, so a simple application of the deduction theorem I-2.8 reveals that

$$(A \to B), (B \to C) \vdash (A \to C).$$

This is known as the *hypothetical syllogism rule* (HS).

With HS in our toolbox, we can perform the following proof in the formal system P.

(1) [I-(P1)] $\neg A \to (\neg B \to \neg A)$,

(2) [I-(P3)] $(\neg B \to \neg A) \to (A \to B)$,

(3) [HS on (1) and (2)] $\neg A \to (A \to B)$.

(ii) No fancy artefacts are required for this proof. We just need some patience.

(1) [(i)] $\neg\neg A \to (\neg A \to \neg\neg\neg A)$,

(2) [I-(P3)] $(\neg A \to \neg\neg\neg A) \to (\neg\neg A \to A)$,

(3) [HS on (1), (2)] $\neg\neg A \to (\neg\neg A \to A)$,

(4) [I-(P2)] $(\neg\neg A \to (\neg\neg A \to A)) \to ((\neg\neg A \to \neg\neg A) \to (\neg\neg A \to A))$,

(5) [MP on (3), (4)] $(\neg\neg A \to \neg\neg A) \to (\neg\neg A \to A)$,

(6) [I-2.7] $\neg\neg A \to \neg\neg A$,

(7) [MP on (5), (6)] $\neg\neg A \to A$,

(8) [Instance of (7)] $\neg\neg\neg A \to \neg A$,

(9) [I-(P3)] $(\neg\neg\neg A \to \neg A) \to (A \to \neg\neg A)$,

(10) [MP on (8), (9)] $A \to \neg\neg A$.

(iii) For this proof, we need some background results. Firstly,we should notice that, for any formulas A and B, we have $\neg\neg A, (A \to B) \vdash \neg\neg B$. This follows from some simple applications of modus ponens taking into consideration that, as we showed in the previous deduction, $\neg\neg A \to A$ and $B \to \neg\neg B$ are both theorems of P. In addition, the application of the deduction theorem on this result yields the existence of a proof for

$$(A \to B) \to (\neg\neg A \to B). \tag{1*}$$

Furthermore, some other applications of the deduction of theorem and the modus ponens rule on $A, (A \to B) \vdash B$ reveal that

$$A \to ((A \to B) \to B) \tag{2*}$$

is another theorem of P.

With these preliminaries out of the way, we can now safely proceed to our proof.

(1) [(1∗)] $(A \to B) \to (\neg\neg A \to \neg\neg B)$,

(2) [I-(P3)] $(\neg\neg A \to \neg\neg B) \to (\neg B \to \neg A)$,

(3) [MP on (1), (2)] $(A \to B) \to (\neg B \to \neg A)$,

(4) [Instance of (3)] $((A \to B) \to B) \to (\neg B \to \neg(A \to B))$,

(5) [(2∗)] $A \to ((A \to B) \to B)$,

(6) [HS on (4), (5)] $A \to (\neg B \to \neg(A \to B))$.

By the way, observe that we have also shown

$$(A \to B) \to (\neg B \to \neg A) \tag{3*}$$

to be a theorem of P.

(iv) As an auxiliary result, we need to prove that, given any formula $A \in L_P$, the propositional form

$$(\neg A \to A) \to A \tag{$1\star$}$$

is a theorem in the formal system P. If we consider an arbitrary $B \in L_P$, this can be through from the following deduction.

(1) [(i)] $\neg A \to (A \to \neg B)$,

(2) [I-(P2)] $(\neg A \to (A \to \neg B)) \to ((\neg A \to A) \to (\neg A \to \neg B))$,

(3) [MP on (1), (2)] $(\neg A \to A) \to (\neg A \to \neg B)$,

(4) [I-(P3)] $(\neg A \to \neg B) \to (B \to A)$,

(5) [HS on (3), (4)] $(\neg A \to A) \to (B \to A)$,

(6) [Instance of (5)] $(\neg A \to A) \to ((\neg A \to A) \to A)$,

(7) [I-(P2)] $((\neg A \to A) \to ((\neg A \to A) \to A)) \to (((\neg A \to A) \to (\neg A \to A)) \to ((\neg A \to A) \to A))$,

(8) [MP on (6), (7)] $((\neg A \to A) \to (\neg A \to A)) \to ((\neg A \to A) \to A)$,

(9) [I-2.7] $(\neg A \to A) \to (\neg A \to A)$,

(10) [MP on (8), (9)] $(\neg A \to A) \to A$.

In addition, it is immediate that $\neg B, (\neg B \to \neg A), (\neg A \to B) \vdash B$, which, after some applications of the deduction theorem, shows that

$$(\neg B \to \neg A) \to ((\neg A \to B) \to (\neg B \to B)). \tag{2$\star$}$$

is another theorem of propositional logic.

We will now proceed to deduce B from $(A \to B)$ and $(\neg A \to B)$. This will, through some applications of the deduction theorem, yield the result we wanted to prove.

(1) [Hypothesis] $A \to B$,

(2) [(iii)(3$*$)] $(A \to B) \to (\neg B \to \neg A)$,

(3) [MP on (1), (2)] $\neg B \to \neg A$,

(4) [(2$\star$)] $(\neg B \to \neg A) \to ((\neg A \to B) \to (\neg B \to B)$,

(5) [MP on (3), (4)] $(\neg A \to B) \to (\neg B \to B)$,

(6) [Hypothesis] $\neg A \to B$,

(7) [MP on (5), (6)] $\neg B \to B$,

(8) [(1$\star$)] $(\neg B \to B) \to B$,

(9) [MP on (7), (8)] B.

Having shown that $(A \to B), (\neg A \to B) \vdash B$, the deduction theorem leads us to $\vdash (A \to B) \to ((\neg A \to B) \to B)$. This concludes the proof.

■

A.2 Lemma. Let $A \in L_P$ be any propositional and let i be any interpretation of propositional logic together with its induced valuation function v_i. Without loss of generality, let $x_1, \dots, x_k$ be the only distinct propositional variables appearing in A.

For every propositional form $\varphi \in L_P$, we define φ^i as φ if $v_i(\varphi) = 1$ and φ^i as $\neg\varphi$ if, otherwise, $v_i(\varphi) = 0$. Under these conditions,

$$x_1^i, \dots, x_k^i \vdash A^i.$$

Proof. We will prove this by induction on the number of connectives in A. If A has no connectives, then it will simply be the variable x_1 and, clearly, $x_1^i \vdash x_1^i$, so the base case is obvious.

Let us assume the result to hold for any formulas with n or less connectives and show it for an arbitrary formula A with $n + 1$. Necessarily, A will be of the form $\neg$X or X $\to$ Y for $X, Y \in F_P$. In either case, the result will hold, by hypothesis, for the subformulas X and Y.

If A is of the form $\neg$X, we may have $v_i(X) = 0$ or $v_i(X) = 1$. If $v_i(X) = 0$, then X^i will be $\neg$X, so, by the inductive hypothesis, $x_1^i, \dots, x_k^i \vdash \neg X$. Moreover, $v_i(X) = 0$ implies that $v_i(A) = v_i(\neg X) = 1$, so A^i will be $\neg$X and, therefore, we can conclude that $x_1^i, \dots, x_k^i \vdash A^i$.

If $v_i(X) = 1$, then X^i will be X and we will have $v_i(A) = v_i(\neg X) = 0$, so A^i will be $\neg\neg$X. In addition, by the inductive hypothesis, we know that $x_1^i, \dots, x_k^i \vdash X$. In conjunction with A.1(ii), this shows that $x_1^i, \dots, x_k^i \vdash \neg\neg X$ which translates into our desired $x_1^i, \dots, x_k^i \vdash A^i$.

Let us now assume A to be of the form X $\to$ Y. We will consider three different subcases: one for $v_i(X) = 0$, one for $v_i(Y) = 1$ and one for $v_i(X) = 1$ and $v_i(Y) = 0$.

If $v_i(X) = 0$, then X^i will be $\neg$X and we will obviously have $v_i(A) = v_i(X \to Y) = 1$, so A^i will be A. According to our inductive hypothesis, we know that $x_1^i, \dots, x_k^i \vdash \neg X$. Taking A.1(i) into consideration, it follows that $x_1^i, \dots, x_k^i \vdash (X \to Y)$, hence $x_1^i, \dots, x_k^i \vdash A$ as we wanted to show.

If $v_i(Y) = 1$, Y^i will be Y and we will also have $v_i(A) = v_i(X \to Y) = 1$. In this scenario, our hypothesis states that $x_1^i, \dots, x_k^i \vdash Y$, which, using I-(P1), leads us to $x_1^i, \dots, x_k^i \vdash X \to Y$.

Lastly, if $v_i(X) = 1$ and $v_i(Y) = 0$, we will have $v_i(A) = 0$, so A^i will be $\neg(X \to Y)$. By the inductive hypothesis,

$$x_1^i, \dots, x_k^i \vdash X, \qquad x_1^i, \dots, x_k^i \vdash \neg Y,$$

which, together with A.1(iii), leads to $x_1^i, \dots, x_k^i \vdash \neg(X \to Y)$ and completes our proof. ■

A.3 Theorem. The formal system P of propositional logic is semantically complete.

Proof. Let A be any tautology. We aim to prove that it is a theorem in P.

According to our previous lemma, letting $x_1, \dots, x_k$ be the only distinct propositional variables appearing in A, if we fix any interpretation i, we will have $x_1^i, \dots, x_k^i \vdash A^i$. Nonetheless, as A is a tautology, it will be true under any interpretation i and, therefore, A^i will be A.

We can pick any two interpretations i and j that make x_k true and false respectively and that are equal in all the remaining variables, this is, that satisfy $v_i(x_k) = 1$, $v_j(x_j) = 0$ and $v_i(x_l) = v_j(x_l)$ if $k \neq l$. According to our previous lemma, we will have

$$x_1^i, \dots, x_{k-1}^i, x_k \vdash A, \qquad x_1^i, \dots, x_{k-1}^i, \neg x_k \vdash A,$$

where we have used the fact that x_l^i is the same as x_l^j if $l \neq k$. Applying the deduction theorem, we are thus led to

$$x_1^i, \dots, x_{k-1}^i \vdash (x_k \to A), \qquad x_1^i, \dots, x_{k-1}^i \vdash (\neg x_k \to A).$$

If we now consider A.1(iv), it is immediate that

$$x_1^i, \dots, x_{k-1}^i \vdash A.$$

A recursive application of this reasoning eventually leads to $\vdash$ A. ■

A.4 Lemma. Let H be a consistent first-order system defined on a certain first-order language. There exists a consistent extension of the set of non-logical axioms of H (also known as an extension of H) that makes H syntactically complete.

Proof. Let us consider an infinite enumeration $\mathrm{A}_1, \dots, \mathrm{A}_n, \dots$ of all the formulas in the first-order language under consideration. Such an enumeration can indeed be constructed and I will leave the details for you. We define a sequence of extensions of H as follows. Firstly, the formal system H_0 will be H itself. Then, for every natural n, we will define H_n to be the extension of H_{n-1} including $\neg\mathrm{A}_n$ as an axiom if A_n is not a theorem in H_{n-1}. If it is a theorem, we define H_n to be H_{n-1}.

By I-3.14, we know that each of this extensions will be consistent. Thus, if we consider the extension H_∞ including as axioms all the axioms of all the H_n for every $n \in \mathbb{N}$, it is immediate that it will be consistent too. In addition, it can be easily seen that H_∞ will be a syntactically complete extension. Just assume that there is a sentence A such that neither A or $\neg$A are theorems in H_∞. Shouldn't one of them have been added as an axiom in a certain H_n? ■

A.5 Theorem. Any consistent first-order formal system H has a (countable!) model.

Proof. Let us consider the formal system H_0 obtained by adding an infinite sequence of symbols $b_0, \dots, b_n, \dots$ to the particular language it uses.

It is easy to see that the addition of these symbols has no effect in the consistency of the system. It suffices to show how a proof in the new formal system of any statement A can be transformed into one, in the old system, of the formula A′ obtained by replacing every constant $b_1, \dots, b_n, \dots$ in A by a new variable not occurring in A. Thus, should the new system be able to prove both A and $\neg$A, the old one would prove A′ and $\neg$A′, but that would be impossible for we have assumed H to be consistent.

We can now consider an enumeration $\mathrm{A}_1, \dots, \mathrm{A}_n, \dots$ of all the formulas in the language of H_0 with a single free variable — which we will denote, for each A_n, as y_n. From this point, we define a sequence of extensions of H_0 as follows. The first element of this sequence will obviously be H_0. Then, for every natural n, we will take H_n to be the extension of H_{n-1} incorporating the additional axiom S_n given by

$$\mathrm{A}_n(y_n \| c_n) \to (\forall y_n)\mathrm{A}_n,$$

where c_n is the first element in $b_1, \dots, b_n, \dots$ that does not appear in A and that does not belong to $\{c_1, \dots, c_{n-1}\}$.

Let us show that all these extensions will be consistent. Given any natural n, we will assume that H_n is not consistent and, therefore, that it can prove, for a certain formula A, both A and $\neg$A. According to the principle of explosion, this means that we will be able to prove $\vdash_{\mathsf{H}_n} \neg\mathrm{S}_n$. Proofs in H_n are deductions from S_n in H_{n-1}. Considering this fact together with the deduction theorem yields

$$\vdash_{\mathsf{H}_{n-1}} \mathrm{S}_n \to \neg\mathrm{S}_n.$$

Since $(\mathrm{A} \to \neg\mathrm{A}) \to \neg\mathrm{A}$ is a tautology, we can then deduce that $\neg\mathrm{S}_n$ will be a theorem in H_{n-1}, so we will have

$$\vdash_{\mathsf{H}_{n-1}} \neg(\mathrm{A}_n(y_n \| c_n) \to (\forall y_n)\mathrm{A}).$$

As both $\neg(\mathrm{A} \to \mathrm{B}) \to \mathrm{A}$ and $\neg(\mathrm{A} \to \mathrm{B}) \to \neg\mathrm{B}$ are tautologies, this means that

$$\vdash_{\mathsf{H}_{n-1}} \mathrm{A}_n(y_n \| c_n), \qquad \vdash_{\mathsf{H}_{n-1}} \neg(\forall y_n)\mathrm{A}_n.$$

As c_n does not appear in the axioms of H_{n-1}, we can safely replace each occurrence of c_n in the proof of $\mathrm{A}_n(y_n\|c_n)$ by a variable not appearing in the proof. Then, an application of the generalisation rule reveals that $\vdash_{\mathsf{H}_{n-1}} (\forall y_n)\mathrm{A}_n$, which would mean that H_{n-1} would not be consistent, thus contradicting our hypothesis and showing the consistency of all the extensions in the sequence.

As all these extensions are consistent, the formal system H_∞ defined as the extension of H_0 containing all the axioms A_n will be consistent too. We can now consider the (consistent) extension H^* of H_∞ obtained from a direct application of lemma A.4.

We will now show that every element in the domain of any model of H^* can be referenced directly from the language. Formally, this means showing that, given any formula A_n with a single free variable y_n, if — for every *closed term t* (term with no free variables) — $\mathrm{A}(y_n\|t)$ is a theorem, then so is $(\forall y_n)\mathrm{A}$.

This is fairly easy to do. If we assume $\mathrm{A}_n(y_n\|t)$ to be a theorem for every closed term t, then, in particular, $\mathrm{A}(y_n\|c_n)$ will be a theorem. Thus, applying modus ponens on S_n, one can prove $\vdash_{\mathsf{H}_\infty} (\forall y_n)\mathrm{A}$.

We will now define an interpretation of H^* that will be a model for it and, therefore, will also define a model for H. This interpretation will have as domain of discourse the set of closed terms of the language of H^* (recall that it is the extension of the language of H with the additional constants b_i). The interpretation of each constant b_i will be itself and the functions will behave in the obvious way. The interpretation of each n-ary predicate symbol P will be the predicate that, for any closed terms $t_1, \dots, t_n$ is true if and only if $\mathrm{P}(t_1, \dots, t_n)$ is a theorem in H^*.

We can then prove by induction on the number of connectives that any closed formula is a theorem in H^* if and only if it is true in the interpretation we have defined. Taking the generalisation rule into consideration, this can be trivially extended to non-closed formulas, which implies that our interpretation is, indeed, a model of H^*. From the existence of a model of H^*, the existence of a model for H can be trivially inferred.

Proving, by induction on the number of connectives, that closed formulas in H^* are theorems if and only if they are true is fairly straightforward and somewhat entertaining if you have a good understanding of the material. I will leave it to you. Keep in mind that — at least for the purpose of convincing yourself this result is true — you don't need to write a fully-fledged formal proof. A few diagrams and notes and a good dose of thinking will probably suffice. ■

To be continued…

Bibliography

[1] J. R. Brown, *Philosophy of mathematics: a contemporary introduction to the world of proofs and picrures*, Routledge, second ed., 2008.

[2] T. Gowers, J. Barrow-Green, and I. Leader, *The Princeton Companion to Mathematics*, Princeton University Press, 2008.

[3] K. Gödel, *On formally undecidable propositions of Principia Mathematica and related systems*, Dover Publications, 1992.

[4] P. R. Halmos, *Naive set theory*, Springer-Verlag, 1960.

[5] A. G. Hamilton, *Logic for mathematicians*, Cambridge University Press, 1978.

[6] R. Hammack, *Book of Proof*, third ed., 2018.

[7] K. Houston, *How to think like a mathematician: a companion to undergraduate mathematics*, Cambridge University Press, 2009.

[8] G. Hunter, *Metalogic. An introduction to the Metatheory of standard first order logic*, Palgrave Macmillan, 1971.

[9] A. N. Kolmogórov and A. G. Dragalin, *Lógica matemática. Introducción a la lógica matemática*, Editorial URSS, 2013.