1

Foundations and Paradoxes

In this chapter and the following, we shall learn lots of things in a short time.[1] Initially, some of the things we will gain knowledge of may appear unrelated to each other, and their overall usefulness might not be clear either. However, it will turn out that they are all connected within Gödel's symphony. Most of the work of these two chapters consists in preparing the instruments in order to play the music. We will begin by acquiring familiarity with the phenomenon of *self-reference* in logic – a phenomenon which, according to many, has to be grasped if one is to understand the deep meaning of Gödel's result. Self-reference is closely connected to the famous *logical paradoxes*, whose understanding is also important to fully appreciate the Gödelian construction – a construction that, as we shall see, owes part of its timeless fascination to its getting quite close to a paradox without falling into it.

But what is a paradox? A common first definition has it that a paradox is the absurd or blatantly counter-intuitive conclusion of an argument, which starts with intuitively plausible premises and advances via seemingly acceptable inferences. In *The Ways of Paradox*, Quine claims that "a paradox is just any conclusion that at first sounds absurd but that has an argument to sustain it."[2] We shall be particularly concerned not just with sentences that are paradoxical in the sense of being implausible, or contrary to common sense ("paradox" intended as something opposed to the δόξα, or to what is ἔνδοξον, entrenched in pervasive

[1] This chapter draws on Berto (2006a), (2007a), and (2007b) for an account of the basics of set theory and of logical paradoxes.

[2] Quine (1966), p. 1. Sainsbury's definition is: "an apparently unacceptable conclusion derived by apparently acceptable reasoning from apparently acceptable premises" (1995), p. 1.

and/or authoritative opinions), but with sentences that constitute authentic, full-fledged contradictions. A paradox in this strict sense is also called an *antinomy*.

However, sometimes the whole argument is also called a paradox.[3] So we have Graham Priest maintaining that "[logical] paradoxes are all arguments starting with apparently analytic principles … and proceeding via apparently valid reasoning to a conclusion of the form 'α and not-α'."[4]

Third, at times a paradox is considered as a set of jointly inconsistent sentences, which are nevertheless credible when addressed separately.[5]

The logical paradoxes are usually subdivided into the *semantic* and *set-theoretic*. What is semantics, to begin with? We can understand the notion by contrasting it with that of *syntax*. Talking quite generally, in the study of a language (be it a natural language such as English or German, or an artificial one such as the notational systems developed by formal logicians), semantics has to do with the relationship between the linguistic signs (words, noun phrases, sentences) and their meanings, the things those signs are supposed to signify or stand for. Syntax, on the other hand, has to do with the symbols themselves, with how they can be manipulated and combined to form complex expressions, without taking into account their (intended) meanings.

Typically, such notions as *truth* and *denotation* are taken as pertaining to semantics.[6] Importantly, a linguistic notion is classified as (purely)

[3] Beall and van Fraassen (2003), p. 119 claim that "a *paradox* … is an argument with apparently true premises, apparently valid reasoning, and an apparently false (or untrue) conclusion."

[4] Priest (1987), p. 9.

[5] This definition is taken as having some advantages over the previous ones by Sorensen (2003), p. 364.

[6] And *truth* is generally considered as the basic semantic notion. This is because, of the various syntactic categories, the dominant paradigm of contemporary philosophy of language puts (declarative) sentences at the core, and takes the meaning of sentences to consist mainly, if not exclusively, of their truth conditions. The celebrated motto comes from Wittgenstein's *Tractatus logico-philosophicus*: "To understand a proposition means to know what is the case if it is true." Since to understand a sentence is to grasp its meaning, the motto says that this amounts to understanding the conditions under which the sentence at issue is true. To know what "Snow is white" means is to know what the world must be like if this sentence has to be true. And it is true in the event that things in the world are as it claims them to be, that is, in the event that snow is actually white. Within this semantic perspective (which is therefore called "truth-conditional"), precisely the notion of truth is placed at center stage.

syntactic when its specification or definition does not refer to the meanings of linguistic expressions, or to the truth and falsity of sentences. The distinction between syntax and semantics is of the greatest importance: I shall refer to it quite often in the following, and the examples collected throughout the book should help us understand it better and better.

The set-theoretic paradoxes concern more technical notions, such as those of *membership* and *cardinality*. These paradoxes have cast a shadow over set theory, whose essentials are due to the great nineteenth-century mathematician Georg Cantor, and which was developed by many mathematicians and logicians in the twentieth century.

Nowadays, set theory is a well-established branch of mathematics. (One should speak of set *theories*, since there are many of them; but mathematicians refer mainly to one version, that due to Ernst Zermelo and Abraham Fraenkel, to which I shall refer in the following.) But the theory has also a profound philosophical importance, mainly because of the role it has had in the development of (and the debate on) the so-called *foundations* of mathematics. Between the end of the nineteenth century and the beginning of the twentieth, the great philosophers and logicians Gottlob Frege and Bertrand Russell attempted to provide a definitive, unassailable logical and philosophical foundation for mathematical knowledge precisely by means of set theory. When Gödel published his paper, the dispute on the foundations of mathematics was quite vigorous, because of a crisis produced by the discovery of some important paradoxes in the so-called naïve formulation of set theory.

In these initial chapters, therefore, we shall learn some history and some theory. On the one hand, we will have a look at the changes that logic and mathematics were undergoing at the beginning of the twentieth century, mainly because of the paradoxes: to know something of the logical and mathematical context Gödel was living in will help us understand why the Theorem was the extraordinary breakthrough it was. But we shall also learn some basic mathematical and set-theoretical concepts. Among the most important notions we will meet in this chapter is that of *algorithm*. By means of it, we should come to understand what it means for a given set to be (intuitively) *decidable*; what it means for a given set to be (intuitively) *enumerable*; and what it means for a given function to be (intuitively) *computable*. If this list of announcements on the subjects we shall learn sounds alarming, I can only say that the initial pain will be followed by the gain of seeing these separate pieces come together in the marvelous Gödelian jigsaw.

1 "This sentence is false"

I have claimed that the semantic paradoxes can involve different semantic concepts, such as *denotation*, *definability*, etc. We shall focus only on those employing the notions of truth and falsity, which are usually grouped under the label of the *Liar*. These are the most widely discussed in the literature – those for which most tentative solutions have been proposed. They are also the most classical, having been on the philosophical market for more than 2,000 years – a fact which, by itself, says something about the difficulty of dealing with them. The ancient Greek grammarian Philetas of Cos is believed to have lost sleep and health trying to solve the Liar paradox, his epitaph claiming: "It was the Liar who made me die / And the bad nights caused thereby."

One of the most ancient versions of semantic paradox appears in St Paul's *Epistle to Titus*. Paul blames a "Cretan prophet," who was to be identified as the poet and philosopher Epimenides, and who was believed to have at one time said:

(1) All Cretans always lie.

Actually, (1) is not a real paradox in the strict sense of a sentence which, on the basis of our *bona fide* intuitions, would entail a violation of the Law of Non-Contradiction. It is just a sentence that, on the basis of those intuitions, cannot be true. It is self-defeating for a *Cretan* to say that Cretans always lie: if this were true – that is, if it were the case that all sentences uttered by any Cretan are false – then (1), being uttered by the Cretan Epimenides, would have to be false itself, against the initial hypothesis. However, there is no contradiction yet: (1) can be just false under the (quite plausible) hypothesis that some Cretan sometimes said something true.

We are dealing with a full-fledged Liar paradox (also attributed to a Greek philosopher, and probably the greatest paradoxer of Antiquity: Eubulides) when we consider the following sentence:

(2) (2) is false.

As we can see, (2) refers to itself, because it is no. 2 of the sentences highlighted in this chapter, and tells something of the very sentence no. 2. Also (1) refers to itself, but does it in a different way from (2). This is what makes (1) not strictly paradoxical. Sentence (1) claims that all the members of a set of sentences (those uttered by Cretans) are false. In addition, it belongs to that very set, due to its being uttered by a Cretan. Therefore (1) can be simply false, under the empirical hypothesis that some sentence uttered by a Cretan, and different from (1), is true. This is also what makes it look so odd: it is unsatisfactory that a logical paradox is avoided only via the empirical fact that some Cretan sometimes said something true.

Some form of self-reference can be detected in (almost) all paradoxes, so that the phenomenon of self-reference as such has been held responsible for the antinomies. Nevertheless, lots of self-referential sentences are harmless, in that we seem to be able to ascertain their truth value in an unproblematic way. For instance, you may easily observe that, among the following, (3) and (4) are true, whereas (5) is false:

(3) (3) is a grammatically well-formed sentence.
(4) (4) is a sentence contained in *There's Something About Gödel!*
(5) (5) is a sentence printed with yellow ink.

In contrast, (2) is not harmless at all. Let us reason by cases. Suppose (2) is true: then what it says is the case, so it's false. Suppose then (2) is false. This is what it claims to be, so it's true. If we accept the Principle of Bivalence, that is, the principle according to which all sentences are either true or false, both alternatives lead to a paradox: (2) is true *and* false! To claim that something is both true and false is to produce a denial of the Law of Non-Contradiction. And this is how our *bona fide* intuitions lead us to a contradiction, via a simple reasoning by cases.

Other versions of the Liar are called *strengthened Liars*,[7] or also *revenge Liars* (whereas (2) may be called the "standard" Liar):

(6) (6) is not true.
(7) (7) is false or neither true nor false.

The reason why sentences such as (6) deserve the title of *strengthened* Liars is the following. Some logicians (including the best one of our times, Saul Kripke) have proposed circumventing the standard Liar (2) by dispensing with the Principle of Bivalence, that is, by admitting that some sentences can be neither true nor false, and that (2) is among them. Sentence (2) is a statement such that, if it were false, it would be true, and if it were true, it would be false. But we can avoid the contradiction by granting that (2) is neither. Such a solution has some problems with sentences such as (6), which appear to deliver a contradiction even when we dismiss Bivalence. In this case, the set of sentences is divided into three subsets: the true ones, the false ones, and those which are neither. Now we can reason by cases again with (6): either (6) is true, or it is false, or neither. If it's true, then what it says is the case, so it's not true. If it's false or neither true nor false, then it is not true. However, this is what it claims to be, so in the end it's true. Whatever option we pick, (6) turns out to be both true and untrue, and we are back to contradiction. This Liar thus gains "revenge" for its cousin (2).[8]

2 The Liar and Gödel

A sentence can refer to itself in various ways, so we can have various versions of (2). For instance:

(2a) This sentence is false.
(2b) I am false.
(2c) The sentence you are reading is false.

7 As far as I know, the terminology is due to van Fraassen (1968).

8 Some (e.g. Graham Priest, in his classic works on dialetheism) have conjectured that *any* consistent solution to the Liar faces the same destiny: for any version of the Liar paradox which is solved by the relevant theory, one can build a revenge paradox by exploiting the very notions introduced in the theory in order to address the previous one. On these issues, see Berto (2006a), Ch. 2, and Berto (2007b), Ch. 2.

The paradox can also be produced without any immediate self-reference, but via a short-circuit of sentences. For instance:

(2d) (2e) is true.
(2e) (2d) is false.

This is as old as Buridan (his sophism 9: Plato saying, "What Socrates says is true"; Socrates replying, "What Plato says is false"). If what (2d) says is true, then (2e) is true. However, (2e) says that (2d) is false ... and so on: we are in a paradoxical loop.

However, it seems that self-reference is obtained in all cases by means of an unavoidable "empirical," i.e., contextual or indexical, component. In fact, in the paradoxical sentences we have examined so far, self-reference is achieved via the numbering device, or via indexical expressions such as "I", "this sentence," and so on. Only factual and contextual information tells us that the denotation of (those tokens of) such expressions is the very sentence in which they appear as the grammatical subjects. This holds for the "looped Liar": suppose (2d) is as above, but (2e) now is "Perth is in Australia." Then (2d) is just true, and no paradox is expected. But it happens also with the immediately self-referential paradoxical statements above: for instance, if I uttered (a token of) (2a) by pointing, say, at (a token of) the sentence "2 + 2 = 5" written on a blackboard, there would be no self-reference at all, for "this sentence," in the context, would refer to (the token of) "2 + 2 = 5" (and, besides, I would be claiming something true). Ditto if I uttered (2c) referring to you, while you are reading the false sentence written on the blackboard.

Because of this, some (among which the Italian mathematician Giuseppe Peano, of whom I shall talk again later) have believed that the semantic paradoxes involve some non-*logical* phenomenon: they depend on contextual, empirical factors. Frank Ramsey, to whom the distinction between semantic and set-theoretic paradoxes is usually ascribed, depicted the situation thus by referring to the list of paradoxes examined in Russell and Whitehead's *Principia mathematica*:

> Group A [i.e., antinomies no. 2, 3, and 4 of the original list of *Principia*: among them, the Russell and Burali-Forti paradoxes, which I will introduce later] consists of contradictions which, were no provision made against them, would occur in a logical or mathematical system itself. They involve

> only logical or mathematical terms such as class and number, and show that there must be something wrong with our logic and mathematics. But the contradictions in Group B [i.e., antinomies no. 1, 5, 6, 7 of *Principia*: among them, the Liar] are not purely logical, and cannot be stated in logical terms alone; for they all contain some reference to thought, language, or symbolism, which are not formal but empirical terms.[9]

However, just after Ramsey had proposed the distinction, Gödel himself showed how to build, within a formal logical system, self-referential constructions with no empirical trespassers of any kind: self-referential statements whose content is as empirical and contextual as that of "2 + 2 = 4." To achieve this, Gödel used the language of mathematical logic as nobody had done before; and the apparatus he put to work is probably the most inspired aspect of the proof of the Theorem that bears his name.

Behind the Gödelian construction hide precisely the simple intuitions concerning the conundrum originated by the Liar which made the ancient Greeks lose their sleep. However, Gödel did not exploit those intuitions to engender a contradiction, via a sentence that claims of itself to be false, like (2), or untrue, like (6). He produced a sentence that walks on the edge of paradox, without falling into it. I shall talk of this mysterious Gödelian sentence at length: it is, in fact, the main character of the story I have begun to tell.

3 Language and metalanguage

The great Polish logician Alfred Tarski, and many after him, have held responsible for such semantic paradoxes as the Liars certain features of natural language, grouped under the label of "semantic closure conditions." Roughly, a semantically closed language is a language capable of talking of its own semantics, of the meanings of the expressions of the language itself. Less roughly, "a semantically closed language is one with semantic predicates, like 'true', 'false', and 'satisfies', that can be applied to the language's own sentences."[10] It is because English can mention

9 Ramsey (1931), pp. 36–7.
10 Kirkham (1992), p. 278.

its own expressions, and ascribe semantic properties to them, that we can have such sentences as (2) or (6): some expressions of our everyday language can somehow refer to themselves; "true" and "false" are perfectly meaningful predicates of English; and they can be applied to sentences of English.

In a Tarskian approach, the semantic paradoxes are due to a mixture of *object language* and *metalanguage*. Logicians and philosophers usually call "object language" the language we speak about, or we give a theory of, this being precisely the object of the theory. However, the theory itself will obviously be phrased in some language or other; and the language in which the theory is formulated can be labeled as a metalanguage, that is to say, a "language on a language."

That (object) language and metalanguage may be distinct is fairly clear. If you are studying a basic French grammar written in English, you will find that French figures in it mostly as the object language, whereas English is employed mainly as the metalanguage. But in our self-referential statements above, the two levels are mixed: these are English sentences talking of English sentences (specifically, of themselves). And this fusion, according to the Tarskian approach, gives rise to the paradox.

The Tarskian treatment maintains that the truth predicate cannot be *univocal*. A single surface grammar expression, "is true," has an ambiguous function for different languages, each of which is semantically open at the level of some deep logical grammar. Instead of a unique language, we would have a hierarchy, more or less with the following structure. For any ordinal n we have a language L_n, and n is the *order* of L_n. Let us begin with L_0, taken as our "basic" level language. The semantic concepts concerning L_0 cannot be expressed within L_0 itself, but must be expressed in a language, say L_1, which is its metalanguage. L_1 will contain predicates that refer to the semantic concepts of L_0 (and, in particular, by means of which we can provide a definition of *truth* for sentences of L_0: I shall come to the details of the Tarskian definition of truth in Chapter 9). However, L_1 is itself "semantically open": it cannot express its own semantic concepts. So a definition of truth for L_1 will be expressed in a language, L_2, which is the metalanguage of L_1; and so on.[11] The Tarskian solution parameterizes the semantic predicates along the hierarchy of the metalanguages: the metalinguistic

[11] See Sainsbury (1995), pp. 118–19.

"true" and "false" are now abbreviations for "true in the object language," "false in the object language." In particular, the standard Liar turns into "This sentence is false in the object language." Its place is in the metalanguage, and it is just *false* there, not paradoxical: since metalinguistic sentences do not belong in the object language at all, the Liar does not have the property it claims to have.[12]

In point of fact, however, Tarski proposed his hierarchy as a structure for the artificial languages of formal logic, and did not claim his strategy to be applicable to natural languages (though others after him have been less restrained on this). Tarski's prudence is easily understood. First, there is no evidence that the predicate "true" performs some ambiguous function along some hidden hierarchy of languages, metalanguages, metametalanguages, etc. This makes the proposal to apply the theory to ordinary English look like a form of *revisionism*: a suggestion to the effect that ordinary English be somehow regimented. If the idea came to Tarski's mind, he certainly found it unsatisfactory.[13]

Second, a decisive difference between such a hierarchy and English is that there does not seem to be *any* metalanguage for English. This becomes manifest if we accept the principle according to which ordinary language is, so to speak, "transcendental": anything that is linguistically expressible can be expressed within ordinary language – there is no limit to it. In Tarski's own words:

> A characteristic feature of colloquial language (in contrast to various scientific languages) is its universality. It would not be in harmony with the spirit of this language if in some other language a word occurred which could not be translated into it; it could be claimed that "if we can speak meaningfully about anything at all, we can also speak about it in colloquial language."[14]

[12] See Kirkham (1992), p. 280.

[13] "Whoever wishes, in spite of all difficulties, to pursue the semantics of colloquial language with the help of exact methods will be driven first to undertake the thankless task of a reform of this language. He will find it necessary to define its structure, to overcome the ambiguity of the terms which occur in it, and finally to split the language into a series of languages of greater and greater extent, each of which stands in the same relation to the next in which a formalized language stands to its metalanguage. It may, however, be doubted whether the language of everyday life, after being 'rationalized' in this way, would still preserve its naturalness and whether it would not rather take on the characteristic features of the formalized languages" (Tarski (1936), p. 406).

[14] Tarski (1956), p. 164.

We need not enter into subtle issues in the philosophy of language, though. Two important things to be kept in mind in following this book are (a) the idea of the distinction between (object) language and metalanguage, and (b) the basic intuition behind the kind of self-reference taking place when we can see a certain linguistic expression as talking of *itself*, and as ascribing to itself some features and properties.

4 The axiomatic method, or how to get the non-obvious out of the obvious[15]

We have seen that the Liar has been around since the ancient Greeks. We also need to start from ancient Greece to understand what the *axiomatic method* is, and why this method has enjoyed an almost spotless reputation throughout Western thought. We need to refer, in fact, to Euclidean geometry – a theory we all know from elementary school, and which has been the paradigm of axiomatization for centuries.

In his *Elements of Geometry*, the Greek mathematician Euclid introduced some simple geometrical definitions (such as "A point is that which has no part"), and the celebrated five postulates, or axioms, that bear his name (for instance, the first says, "Any two points can be joined by a straight line"; the fourth says, "All right angles are congruent"). In the axiomatic approach, axioms are sentences accepted without a proof, as principles of deduction – principles from which we can infer other sentences, via purely deductive reasoning. The sentences with which the various deductive chains come to an end are called the *theorems* of Euclidean geometry. Such deductive chains are the *proofs* of Euclidean geometry; the closures of such chains, i.e., the theorems, are what are properly said to have been demonstrated, or proved, from the axioms.

Axioms, proofs, theorems: this beautifully simple and powerful pattern of knowledge has always fascinated scholars. Beginning with a small amount of what Quine would have called "ideology," that is, with a few intuitive notions and the initial postulates, Euclidean geometry delivered a large amount of theorems by means of deductive procedures which appeared to be fairly clear and rigorous. And the axioms

[15] This comes from the Leibnizian motto: *spernimus obvia, ex quibus tamen non obvia sequuntur*.

were considered – keep this in mind for the following – as (manifestly) *true*. In the so-called classical conception of axiomatic systems, axioms were taken as evident, if not trivial, truths. This is exactly how we wanted them, so that we could accept them without further argumentation. The chain of deductions and inferences has to come to an end somewhere, and what better place to append it than the obvious? The proofs of Euclidean geometry being valid proofs, truth could go downstairs from the axioms to often more complex and less evident theorems. This, after all, is the fundamental virtue of valid deductive reasoning: transmitting truth from premises to conclusions.

For these reasons, numerous philosophers (from Descartes to Spinoza and Kant) took Euclidean geometry as a paradigm of rigorous knowledge. They sometimes even tried to export the model and its successful features to other compartments of science, so as to raise them to a comparable level of certitude and precision. The case of "export" we are most interested in takes the stage in the next paragraph.

5 Peano's axioms ...

A closer ancestor of Gödel's results is constituted by the amazing developments of mathematics in the nineteenth century. Some of them had to do with the so-called arithmetization of analysis, which allowed the reduction of higher parts of mathematics to elementary arithmetic, that is, to the theory of natural numbers (the positive integers, including zero: 0, 1, 2, ...). Thanks to the work of mathematicians like Weierstrass, Cantor, and Dedekind, other kinds of number were referred to rational numbers (the numbers representable as ratios of integers) and, via these, to the natural numbers.

The two mathematical results we are most interested in, however, are (a) the aforementioned theory of infinite numbers and sets due to Cantor, and (b) an axiomatic achievement: the formulation of the axioms for arithmetic due to Dedekind and Peano. While the axiomatization of geometry dated back to the ancient Greeks, an analogue account for arithmetic became available only at the end of the nineteenth century, when Dedekind provided the recursive equations for addition and multiplication (which will be met and explained in a later chapter), and immediately afterwards Peano proposed the famous axioms for arithmetic that bear his name.

Only three notions appear in Peano's axioms – three notions taken as primitive and fundamental: *zero*, (natural) *number*, and (immediate) *successor*, the (immediate) successor of a number being the one that follows it immediately in the ordering of the naturals, i.e., 1 is the successor of 0, 2 is the successor of 1, and so on. In the *Arithmetices principia, nova methodo exposita* Peano employed the three basic notions to formulate the following five principles:

(P1) Zero is a number.
(P2) The successor of any number is a number.
(P3) Zero is not the successor of any number.
(P4) Any two numbers with the same successor are the same number.
(P5) Any property of zero that is also a property of the successor of any number having it is a property of all numbers.

Peano's fifth axiom, (P5), is usually called the (mathematical) *induction principle*. I'll come back to it repeatedly in the following (as we will see, "induction" here has little to do with inductive reasoning; on the contrary, it is a typical procedure of deductive sciences).

6 ... and the unsatisfied logicists, Frege and Russell

Peano's axiomatization of arithmetic, just like Euclid's for geometry, was still considered by some scholars to be an inadequate account. They complained especially about the insufficient logical rigor in the proof chains. In fact, in his *Elements* Euclid had formulated some so-called "common notions" which looked like general rules of logical inference ("Two things identical to a third one are identical to each other," for instance). However, Euclid's language lacked the rigor of modern logical languages. To establish that a given deductive chain is valid (that is, that it will never lead us from true premises to a false conclusion), one has to look at the meanings of (some of) the words and phrases used to express it. But ordinary language expressions, as we noted when we mentioned Tarski's position on natural language, are often vague, equivocal, or both. Because of this, at least since Leibniz's *Characteristica universalis*, philosophers have been envisaging artificial, formal languages to serve as antidotes to the deficiencies of natural language, and in which rigorous science could be formulated: languages whose syntax was to be absolutely precise, and whose expressions were to have completely precise and univocal meanings.

Now, some of Euclid's proofs appeared to include notions captured neither by the explicit definitions, nor by the postulates; and they certainly adopted principles of logical inference which were not listed among the common notions. As for Peano, he had already introduced a formal notation in 1888 (in fact, one including symbols which are nowadays embedded in the canonical logical and set-theoretical notation). However, his axiomatization of arithmetic lacked a rigorous specification of the logical principles employed in the deductions from the axioms. Peano's proofs were rather informal, and the task of establishing the correctness of the deductive passages was often simply left to the reader. By contrast, since the introduction to his 1879 *Ideography* – the text whose publication is considered the founding act of modern logic – Gottlob Frege had begun to show how arithmetical claims could be proved by means of precise, explicitly stated logical rules. On the one hand, arithmetical proof sequences had to be translated into an artificial symbolic language. On the other hand, the logical rules operating in the proofs had to be made rigorously explicit. Frege provided a first precise characterization of what we nowadays call a *formal system* – a notion to which I will return again and again, and whose richness will be explored little by little as this book develops.

Once "higher" mathematics had been reduced to the natural numbers, and secure (or so it seemed) logical rules to reason on them were available, one might have gained the impression that mathematical knowledge had reached safe ground. Infinitesimals and irrational numbers had a problematic status. For a long time, mathematicians and philosophers had had qualms concerning the consistency of mathematical analysis; but everyone considered the good old integers to be reliable guys. However, Frege had a deeper ambition: that of providing a *foundation, on pure logic, for arithmetic itself*. Such an ambition was shared, between the end of the nineteenth century and the beginning of the twentieth, by Bertrand Russell. It was Russell, in fact, who brought Frege's work to the attention of a wider audience; and "logicism" was the name given to the project, precisely because of its aiming at a rigorous logical foundation of arithmetic. Both Frege and Russell believed there to be no theoretical distinction between the two domains. The notions of zero, (natural) number, and (immediate) successor, taken by Peano as primitive for arithmetic, and its fundamental principles as captured by Peano's axiomatization, were to be defined and deduced in their turn from still more fundamental and purely logical principles. Specifically, they were to be obtained precisely from the principles of set theory, which at the time was considered a limb of logic.[16]

7 Bits of set theory

To understand what the logicist program consisted of – and, most importantly, what major obstacle it stumbled upon – we need to swallow some of the medicine of set theory. What's a set, to begin with? In the first instance, a set is just a collection of objects. In the following, I will usually refer to sets by means of capital Latin letters: A, B, C, ...[17] The fundamental and primitive relation at issue in set theory is that of an object *belonging to* a set, or *being a member of* a set. This is expressed by the symbol "$\in$" (and non-membership is expressed by "$\notin$"). I will write, then, such things as "$x \in$ A" ("$x \notin$ A"), to mean that a given object x is (is not) a member or an element of set A, that is, it belongs (does not belong) to the set.

One can sometimes specify a set simply by providing a complete list of its elements, which is usually written thus: $\{x_1, \ldots, x_n\}$ is the set whose

[16] Nowadays the border between logic and arithmetic is much more precise, mainly because of Gödel's Theorem. When logicians talk about "logic," with no further qualification, they usually refer to the so-called *elementary* or *first-order* logic: the predicate calculus with quantifiers and identity, familiar to anyone who has attended a course in basic logic. The so-called "higher-order" logic has quite a different status, and fuzzy borders with set theory. We shall come back to these notions and acquire more familiarity with them in the following; but we should bear in mind that the sharp distinction between elementary logic and set theory is due to their quite different features, and that such differences emerged mainly thanks to Gödel's work.

[17] Sometimes taken as variables ("For any set A ..."), sometimes as constants, that is, as names for specific sets ("Russell's set R ...," "the set N of natural numbers ..."). Context will disambiguate.

members are indeed x_1, ..., x_n. For instance, one can specify the set whose sole elements are Frege, Juliette Binoche, and the city of Melbourne, thus:

{Gottlob Frege, Juliette Binoche, Melbourne}.

Notice that what is properly listed are the elements' names, not the elements themselves (similarly, when one lists the players of AC Milan, one doesn't actually put in a row Kaka, Pirlo, Shevchenko,..., but writes down their names).

The only things that matter about sets are their members. It is irrelevant in what order (the names of) the members are listed, or whether they occur more than once. This means that, for instance, the following sets:

{Juliette Binoche, Melbourne, Gottlob Frege}
{Gottlob Frege, Melbourne, Melbourne, Gottlob Frege, Juliette Binoche, Juliette Binoche}

are still the same set as the one introduced above.

When we deal with sets having an infinity of elements, such as the set of natural numbers (which I shall call "N"), we cannot in practice specify (the names of) all these elements in a list: we could never finish the job.[18] Therefore sets are often introduced via a *condition*: one specifies some feature, or shared property, or characteristic enjoyed by all and only the elements of the set at issue. A standard notation is the following:

$$\{x \mid \ldots x \ldots\},$$

to be read in English as: "The set of all the x, such that ... x ...." For instance, the set of odd numbers is $\{x \mid x \text{ is an odd number}\}$.

Now, Frege had based his logicist program, aimed at defining numbers via set-theoretic notions, on a version of what is nowadays called

[18] Which doesn't mean that one cannot think about the elements of an infinite set, such as N, as arranged in an infinite list with a first element, a second one, etc., as we do when we write: "0, 1, 2, 3, ..." – on the contrary, the theoretical possibility of arranging the naturals in one such list will prove of great importance in the following.

"naïve set theory." This is built upon two fundamental principles (supposedly) capturing our intuitive conception of set.[19] The *Extensionality Principle* spells out the sufficient conditions for identity between sets. In the canonical notation, it goes like this:

(EP) $\forall x(x \in y \leftrightarrow x \in z) \rightarrow y = z.$

Reading this in English, it says: "If y and z have precisely the same elements, they are the same set" (x, y and z being variables ranging over objects and/or sets).[20] This captures the aforementioned idea that only members matter to the identity of sets.[21]

A consequence of (EP) is that all the sets to which nothing belongs, that is, all sets with no elements, are the same. If there are no winged horses and no unicorns, it can be conjectured that the set of unicorns and the set of winged horses exist; they both lack members. (EP) says that x and y are the same set iff no member of the former is a member of the latter and vice versa – and this is always the case when they are empty. We can therefore talk of *the* empty set, which is usually labeled with the symbol "$\varnothing$".

[19] The formulation provided below is not Frege's, who used what we nowadays would call a higher-order language. This is a complication we can skip here, but, as announced, I shall have something to say on higher-order languages and theories in the following.

[20] One of the features that render set theory important for mathematics is the fact that sets can be members or elements of sets in their turn. In this sense, sets are not just collections of objects, but objects themselves – taking "object" to mean "something which is capable of set membership." As we will see quite soon, Frege and Russell based their logicist approach on the possibility of reducing numbers to sets by considering them as sets of sets. On the other hand, in many contemporary theories of sets (often called "pure" theories), *only* sets figure in the domain of the theory; there are no *Urelemente*, objects that aren't sets. To put it otherwise, within the range of the things the interpreted theory talks about there are no such things as Gottlob Frege, Juliette Binoche, or the city of Melbourne, and then sets, but only sets. In this case, the only objects the variables of the theory, such as x, y, and z, range over, are sets. These specifications, in any case, are not essential in order to understand what follows.

[21] Consequently, different properties can originate the same set. To recycle the classical Fregean example: the property of being an animal with a heart and the property of being an animal with kidneys are intuitively different (*having a heart* does not seem to be the same thing as *having kidneys*). However, they correspond to a unique set, for any animal with a heart also has kidneys, and *vice versa*. This is the mark of the difference between such "intensional" entities as properties, and sets.

8 The Abstraction Principle

The second fundamental principle of (what was later to be called) naïve set theory was the *Abstraction* or *Comprehension Principle* (AP). It can be expressed, via the notation with set abstracts introduced above, thus:

(AP1) $x \in \{y \mid P(y)\} \leftrightarrow P(x)$.

Translating into English, this says: "x is a member of the set of the *P*s iff x is (a) P" (e.g., Jeffery Deaver is a member of the set of writers if and only if Jeffery Deaver is a writer). Otherwise, it can be formulated without set abstracts, thus:

(AP2) $\exists y \forall x (x \in y \leftrightarrow \alpha[x])$,

where α is a *metavariable* for formulas of the formal language, that is, a placeholder for any of those formulas.[22] Specifically, "$\alpha[x]$" can be replaced by any formula with one free variable x (perhaps occurring more than once), expressing a property or condition on x.[23] (AP2) thus says something like: "There's a (set) y such that, for all x, x is a member of y iff x has the property (satisfies the condition expressed by) $\alpha[x]$." (AP) was supposed to express the idea of set included in Cantor's celebrated definition:

> By an "aggregate," we are to understand any collection into a whole M of definite and separate objects *m* of our intuition or of our thought.[24]

The basic insight was that *any* "collection into a whole" is a set, which means that any property P in (AP1), or any condition $\alpha[x]$ in (AP2), is taken as defining one. When we think of something as a thing of a certain sort (and taking "sort" in a broad sense, not in the strict sense of

[22] In the following I will use Greek letters, sometimes as (meta-) variables for formulas ("for some formula α …"), sometimes as names of famous sentences, such as "the Gödel sentence γ …." Again, context will disambiguate.

[23] With the proviso: y must not be free in $\alpha(x)$.

[24] Cantor (1895), p. 85.

sortal concepts), at the same time we appear to think of it as being one of a group, which is itself a thing of a certain sort. More specifically, given any multiplicity with some characterizing condition, the Abstraction Principle seems to guarantee that there exists a set of all and only those objects, and that the set is itself an object. "Object" should be taken as meaning, more or less: something we can refer to as a unity, which is the subject of attributions and predications, and has properties. All this is intuitive if anything is.

But intuition betrays us. In fact, (AP) originated the major crisis in the foundations of mathematics – and the logical milieu in which Gödel grew up. Before we turn to this, we need a few more set-theoretic notions which will prove useful in the following.

9 Bytes of set theory

Another basic relation in set theory is that of *inclusion*, usually expressed by the symbol "$\subseteq$". A set, say A, is said to be included in a set B if and only if each element of A is also an element of B. In this case, A is also called a subset of B. For instance: the set constituted by (all and only) the Germans is a subset of the set of Europeans, since all Germans are Europeans. The set of (all and only) AC Milan midfielders is a subset of the set of AC Milan players, and so on.

Never confuse $\in$ and $\subseteq$, that is, membership and inclusion. Juliette Binoche is a member of the set of Frenchmen, and since all Frenchmen are European, the set of Frenchmen is included in the set of Europeans. But the set of Frenchmen does not *belong to* the set of Europeans, for it is a set, not a European (the set of Frenchmen, being a set, is an abstract object: it is not a person, therefore it cannot be a European, even though its members, the Frenchmen, are Europeans). The membership relation can hold between objects and sets; the inclusion relation can hold between sets. This does not rule out sets having other sets as members, as I have already said above. The fact that one can have sets whose members are sets, that is, sets of sets, made set theory mathematically significant from Frege and Russell's viewpoint; their basic insight, as we shall see soon, was to define natural numbers precisely as sets of sets.

Another notion we will need in the following is that of *ordered n-tuple*: an ordered couple (or pair), triple, etc. I will write "$<x_1, \ldots, x_n>$"

to signify an ordered n-tuple of objects $x_1, \ldots, x_n$. n-tuples are said to be ordered because, unlike sets, the order in which the elements are listed matters: if x is different from y, $<x, y>$ isn't the same thing as $<y, x>$, whereas $\{x, y\}$ is the same thing as $\{y, x\}$.[25] For instance, the triple <Gottlob Frege, Juliette Binoche, Melbourne> is different from the triple <Juliette Binoche, Melbourne, Gottlob Frege>.

One can define the *Cartesian product* of sets via the notion of ordered n-tuple; let us label it with the spot "·". Given two sets A and B, their Cartesian product A · B is the set constituted by all and only the ordered couples whose first element is a member of A and whose second element is a member of B. Generalizing, given n sets $A_1, \ldots, A_n$, their Cartesian product $A_1 \cdot \ldots \cdot A_n$ is the set of all the n-tuples $<x_1, \ldots, x_n>$ such that $x_1 \in A_1$, $x_2 \in A_2, \ldots$, and so on. One can build the Cartesian product of a set with itself, and in this case one talks of a *Cartesian power*: given a set A, its Cartesian square, A^2, is the set of all the ordered couples of elements of A. Generalizing, the n-ary Cartesian power of a given set A, A^n, is the set of all the n-tuples of elements of A.

Never confuse Cartesian powers with *power sets*. The power set of a given set A, usually written as "P(A)," is the set of all subsets of A. For instance, given the set {Gottlob Frege, Juliette Binoche}, its power set is the following:

> {∅, {Gottlob Frege}, {Juliette Binoche}, {Gottlob Frege, Juliette Binoche}}.

This is because (a) the empty set is included, by definition, in any set, and (b) each set is a subset of itself (although not, as is sometimes specified, a *proper* subset).

10 Properties, relations, functions, that is, sets again

In the following, I shall talk quite often of properties, relations, and functions – mainly, although not only, of properties of numbers, relations

[25] The notion of ordered n-tuple can itself be defined in terms of sets, by means of a procedure due to K. Kuratowski. One can define $<x,y>$ as the set $\{\{x\}, \{x,y\}\}$, and $<y,x>$ as the set $\{\{y\}, \{x,y\}\}$.

between numbers, and numerical functions. In mathematics, a function is just a correspondence between one or more numbers called *arguments* (the input), and a unique number called the *value* (the output) given by the function for those arguments (the set of arguments of a function is usually called its *domain*, and the set of values associated with those arguments is usually called its *range* or its *image*).[26] Sometimes people use "function" as a synonym for "operation," and I will follow this usage. For instance, such an operation of elementary arithmetic as addition is a function: when I add 2 to 3 I have a unique value, that is their sum 2 + 3, that is 5, which corresponds to the two numbers 2 and 3 taken as arguments.

We can now begin to put to work our bits and bytes of set theory: it turns out, in fact, that the notions of (numerical) property, relation, and function can be captured set-theoretically. For instance, the property of being a mammal, or that of being an odd number, can be associated with, or considered as, the sets of objects which enjoy the properties (the set of mammals, the set of odd numbers) and which constitute their extension. Therefore in the following I will often say, indifferently, that an object (an animal, a number) has a certain property, or that an object belongs to the set of all and only the objects having the property (the set of mammals, the set of odd numbers).

Similarly, an n-ary relation can be expressed set-theoretically as a set of ordered n-tuples: those between which the relation holds. For instance, one can take the binary or two-place relation between father and son (the relation … *is the father of* …) as the set constituted by all and only the ordered pairs whose first member is the father of the second. The binary relation … *is greater than* …, holding between numbers, can be identified with the (infinite) set of all and only the ordered pairs of numbers, $<m, n>$, such that m is greater than n. The ternary or three-place relation … *is halfway between* … *and* … can be considered as the set of ordered triples such that the first element of the triple is halfway between the second and the third (<Milan, the north pole, the equator>, for instance); and so on.

Also functions can be characterized by means of set-theoretical notions: there exists an easy inter-definability between functions and

[26] I'll most often talk of *total* functions, that is, functions that are defined for all arguments, that is, functions that assign a value to all of them. Functions that are undefined for some arguments are called *partial* functions.

sets. With each n-ary relation – say R (which, as I have just said, can be considered set-theoretically as a set of ordered n-tuples) – one can associate a function with n arguments – say c_R – called its *characteristic function*. This is a function, whose range is $\{1, 0\}$, such that:

$$\begin{array}{ll} \text{If } < x_1, \ldots, x_n > \in \text{R, then}: & c_R(x_1, \ldots, x_n) = 1 \\ \text{Otherwise}: & c_R(x_1, \ldots, x_n) = 0\ . \end{array}$$

That is to say: the characteristic function c_R for the relation R is the function such that, if R holds between the objects $x_1, \ldots, x_n$ (in this order), then c_R gives 1 as its value for the arguments $x_1, \ldots, x_n$ (in this order); if, on the other hand, the relation does not hold between those objects, then c_R gives 0 as its value. Properties are just a special case: given a set M, corresponding to a property, its characteristic function is the unary function c_M, such that:

$$\begin{array}{ll} \text{If } x \in \text{M, then}: & c_M(x) = 1 \\ \text{Otherwise}: & c_M(x) = 0. \end{array}$$

That is, it is the function which maps the given argument x to 1 if x has the property at issue (that is, if x belongs to M), and to 0 if x does not have the property at issue (does not belong to M).

Conversely, with each n-ary function f one can associate an n+1-ary relation G_f, usually called its *graph* relation. This is the relation holding exactly between the arguments of the function and its values, that is, a relation such that:

$$< x_1, \ldots, x_n, x_{n+1} > \in G_f \text{ iff } f(x_1, \ldots, x_n) = x_{n+1}.$$

When all's said and done, discourses on properties and relations on the one hand, and functions on the other, are reducible to each other. We can talk only in terms of sets by explaining away functions in terms of their graph relations; or, conversely, we can talk only in terms of functions by explaining away sets (that is, properties and relations) in terms of their characteristic functions.

If you are beginning to wonder what is the purpose of all this apparatus, remember: we're still just setting up the instruments to perform the Gödelian symphony!

11 Calculating, computing, enumerating, that is, the notion of algorithm

Talking about setting up instruments, we shall now take on board another group of definitions, whose importance for our Gödelian piece of music cannot be overestimated. All the notions we will meet in this section are related to that of *algorithm*. Roughly, this is the name given in mathematics to a mechanical (also called *effective*) procedure which, when applied to a number or to a sequence of numbers, terminates after a finite number of steps, providing some information on the number or sequence of numbers.[27] Such a procedure has to be specifiable as a finite series of totally explicit, simple, and deterministic instructions. The instructions must tell you what is to be done at each step of the procedure, so that no creativity, ingenuity, or free choice is required. The procedure is labeled "mechanical" to evoke the idea that a machine, such as a computer, could carry it out: in fact, the connection is so close that I will often use computers as intuitive examples when explaining algorithmic notions.

Now a (say unary) function f is said to be (*effectively*) *computable* when there is some algorithm that in principle allows to calculate its value for each of its arguments – that is, when it is possible to specify a series of instructions (for instance, in the form of some computer software) following which one can, in principle, determine mechanically and effectively the output $f(x)$ for each input x. Generalizing to functions with n arguments is straightforward. For instance, addition is a two-argument function which is computable in this sense: at school we learn algorithms, that is, mechanical procedures, to calculate, given two numbers m and n, their sum $m + n$.

A set M is called *decidable* (sometimes also *computable*) if, for every x, some algorithm provides a positive or negative answer to the question "$x \in$ M?" – that is, the algorithm allows one to decide, in principle, if x belongs to M or not (for instance, one can set up a computer so

[27] Sometimes the term "algorithm" is also used in the literature to name procedures that do not necessarily terminate. As we shall see, discourses concerning algorithms can be applied also to domains that are not, so to speak, "immediately" numerical – thanks to Gödel's work. We will learn these things little by little, in any case.

that, once asked "$x \in$ M?", after a finite amount of time it will answer with a "Yes" or a "No").

Equivalently, one will claim that a property is decidable in the case when some algorithm can establish whether a given object x has that property or not. For instance, we have arithmetical algorithms to decide whether a given natural number has the property of being divisible by 2, whether it has the property of being prime (i.e., it belongs to the set of prime numbers, the numbers – bigger than 1 – divisible only by themselves and by 1), and so on. Also in this case, generalizing to relations is straightforward: an n-ary relation R is decidable if and only if, for each n-tuple $<x_1, \ldots, x_n>$, there is some algorithm for deciding ...; and so on. The terminology is extended to predicates, that is, to those linguistic entities that denote decidable sets (properties, relations).

One could describe the notions of computable function and decidable set (property, relation) in more verbose and exhaustive ways. But when the chips are down, it remains the case that we are dealing with intuitive and, in this sense, slightly vague notions. One may wonder, in fact, which operations and instructions are intrinsically simple, and which combinations of such operations or instructions are admissible in order to preserve the mechanical nature of the procedure. Later in the development of this book, we will see that the mathematical market offers different theories taken as delivering precise and systematic accounts of the notion of algorithm (computable function, decidable set). We will mainly focus on one of them: the theory of *recursive functions*. For the time being, however, we can stick to our quick, intuitive characterization. This should be enough to immediately understand that *a set* (a property, a relation) *is decidable if and only if its characteristic function is effectively computable*. If there exists an algorithm to decide, for any x, whether x belongs to the given set M or not, then that algorithm allows us to figure whether the value of the corresponding characteristic function $c_M(x)$ is 1 or 0. Conversely, a function is (effectively) computable if and only if its graph relation is decidable. Discourses concerning the computability of functions and the decidability of sets (properties, relations), therefore, can be phrased only in terms of functions, or only in terms of sets (so I will sometimes talk only of sets, sometimes only of functions, and sometimes of both kinds of things).

A set M is called *enumerable* or *effectively enumerable* (sometimes *computably enumerable*, or also *semi-decidable*) when there is a mechanical procedure such that, given some object x, *if* x belongs to M, the procedure will deliver a "Yes" as its output after a finite amount of time; but if x does *not* belong to M, an answer may not be forthcoming (extending to relations is easy also in this case). This means that, with such sets as M, there is a mechanical procedure that generates all the elements of the set (the objects having the relevant property, etc.). For instance, one can program a computer to compute, and print out one after the other, the (names of the) members of the set – not taking into account the limitations due to time, resources, available memory, etc.

The notions of decidable and of (computably) enumerable or semi-decidable sets are not coextensive: as we shall see, there exist (computably) enumerable or semi-decidable sets that are not decidable (and some such set will play a major role in the development of the Gödelian symphony). The two notions are, nevertheless, closely connected in two important ways.

(1) The first connection is the following: *every decidable set is enumerable*. If we have an algorithm, that is, an effective procedure, to decide in a finite number of steps whether a given x is a member or element of a given set M or not, then certainly we also have a mechanical procedure to generate all the (names of) the elements of M one after the other (once again, the generalization to relations is obvious). We can make the point quite simply by resorting, again, to the example from computers: if we can program a computer so that, for each x, it can decide in a finite amount of time whether $x \in$ M or not, then we can certainly program it in such a way that (a) the computer prints (the name of) x when, once the computing process is done, it turns out that x actually belongs to M, and (b) the computer discards x if it turns out that x doesn't belong to M. This way, the computer will print in sequence, and therefore enumerate in a list, all the elements of M.

(2) To understand the second link between decidable and (computably) enumerable sets, let us consider that, given any set M, by the (set-theoretic) *complement* of M is meant the set of all and only the x such that $x \notin$ M (let us label such a complement set "–M").[28] Now

[28] Mainstream set theories usually make or presuppose a distinction between the absolute complement of M and its complement relative to a pre-specified set (say U) of which M is a subset. The relative complement of M with respect to U is the set of all elements of U that are not elements of M. The distinction between absolute and relative complement is required because in most of the current theories an absolute complement of a given set M, that is, the set of just anything not belonging to M, cannot be admitted. This is one of the consequences of the set-theoretic paradoxes we are about to meet.

the second connection goes like this: *a set is decidable if and only if both the set and its complement are enumerable*. First, if a given set M is decidable, then its complement –M obviously is too; therefore both sets are computably enumerable because of connection 1 (that is, any decidable set is enumerable). Second, if both a set M and its complement are enumerable, this means that (a) there exists a mechanical procedure to produce all the elements of M in succession – a procedure which sooner or later will let us know (with a "Yes," or by printing its name, etc.) if a given x belongs to M, although it will remain silent if x does not; and (b) there exists a mechanical procedure to produce all the elements of –M in succession – a procedure which sooner or later will let us know if a given x does not belong to M, although it will say nothing if x belongs to it. Therefore a computer will always be able to decide, given some object x, whether $x \notin$ M or not, by combining the procedures (a) and (b), perhaps by alternating one step of the former with one step of the latter. The computer, that is, applies a step of the procedure which enumerates the element of M: if x shows up, x belongs to M and we are home already. Otherwise, it applies a step of the procedure which enumerates the elements of –M, and if x shows up now, then $x \notin$ M. If x didn't show up either way, the computer applies a second step of the first procedure; then a second step of the second one; and so on. Eventually, x will show up as belonging to M or to –M, and the computer will let us know, or print, its positive or negative answer.

I have used several times the expression "in principle." When the general notions of algorithm, computable function, etc., are specified, one typically disregards the factual and practical limitations of the calculus. Pragmatic considerations concerning the time required to perform the calculus successfully, the energy expended in the process, the amount of memory a computer would need to carry it out, or the paper needed to print the output, and so on, are not taken into account: the general, abstract notion of algorithm or effective procedure is, in this sense, deliberately "idealized." Even the systems of symbols actually adopted in the calculus are irrelevant – which does not mean, of course, that calculating with certain notations may never turn out in practice to be more convenient than with certain others. The symbols which designate numbers are called *numerals*, and should not be confused with the numbers themselves. A numeral is not a number, but a linguistic sign designating a number – and the same number, of course, can be designated by different linguistic signs. For instance, in our ordinary written language we can designate the number four with such numerals as "4" or "four." But we could adopt different notational conventions: for instance, we could express the number *n* by using *n* strokes, so the number four as | | | | (this is the so-called *tally* notation). Now, it is certainly easier for us to add 34 to 8 than XXXIV to VIII. But translating from one notational system to another is itself an effective procedure, so the fact that numbers (arguments, values, etc.) are presented in some numeric notation or other changes nothing from the viewpoint of the idealized notion of algorithm.

12 Taking numbers as sets of sets

Back to history now. We have seen how Peano axiomatized arithmetic via the three basic notions and the five axioms which now bear his name. You should remember that the notion of (natural) number was taken as a primitive, intuitive one. But the logicist enterprise pursued by Russell and Frege required numbers themselves to be reduced to something even more fundamental and basic: specifically, it required numbers to be definable in terms of sets, and of properties of or relations between sets. The basic idea was to the effect that numbers be taken as properties of sets, and therefore (given that, as we have seen, properties can be reduced to sets in their turn) as sets of sets.

To grasp the basic intuition, we have to buy the following definition: two given sets A and B are said to have the *same cardinality*, or to be *equinumerous*, when it is possible to put their elements in a one-to-one connection – that is, it's possible to pair each element of A with one and only one element of B, so that nothing is left "unpaired." In this case, one also claims that there is a one-to-one correspondence or mapping between the two sets, or a *bijective* function, or bijection, pairing each element of A to each element of B. For instance, assume that A is the set of all husbands, and B is the set of all wives, and assume also that the only admissible marriage prescribes monogamy. Then we know that A and B have the same number of elements, even when we don't know exactly what that number is, that is, how many married couples there are. The reason is precisely that we know we can pair one-to-one each husband to the respective wife, so that no husband remains without wife, and no wife without husband: there is a one-to-one mapping between the two sets.

Now we can define the *number of* a given set as the set of all sets equinumerous to it. Any number is then characterized as a property of sets which have the same cardinality, and therefore as a set of equinumerous sets. For instance: which feature is shared by the set of Hercules' labors, the set of Apostles, and the set of months in a year? *Twelve* immediately comes to mind, and the number twelve can be seen as the property of all and only the sets which are equinumerous to those sample sets, and therefore as a set of sets. When we claim that the months in a year are twelve, we are attributing a property not to the months taken individually (January, February, etc.) but to their set: it's a property of that set, and of various others (such as the set of Apostles, etc.); so it's a set of sets. This is how discourses on numbers can be reduced to discourses on sets of sets. And the notions of set, and of belonging to a set, unlike that of number, appear to be definitely primitive ones, irreducible to anything still more fundamental.

This reduction of numbers to sets was just the first half of the story for the logicist project. The second half was to consist in obtaining all of mathematics from the primal set-theoretic notions, taken in those times as logical concepts, by means of purely logical inferences. Logicists aimed at deriving mathematics from logic via deductive chains going from the premises of symbolic logic, down to geometry, through finite and infinite arithmetic.

But the logicists' confidence was doomed to be deeply shaken.

13 It's raining paradoxes

Russell discovered that set theory, which was supposed to provide the foundational machinery for arithmetic (and therefore for the whole of mathematics), actually provided a devastating contradiction. In fact, set theory quickly found itself caught in a storm of contradictions. Let us see how.

All the things I have been saying so far on sets will be needed in the remainder of this book; and almost all the set-theoretic notions presented so far are nowadays customary mathematics. But at least one of the principles we have met brings trouble. This is the Abstraction or Comprehension Principle (AP): no matter exactly how one phrases it, (AP) appears to grant that to any property or condition there corresponds a set. This seemingly intuitive and obvious idea produces various set-theoretic paradoxes in the so-called naïve (version of) set theory. These paradoxes struck the foundations of mathematics as an earthquake at the beginning of the twentieth century.

The simplest and most celebrated of the set-theoretic paradoxes hit Frege on June 16, 1902, in the form of a letter sent by Russell – a letter which belongs to the history of contemporary thought. The paradox exploits a bunch of beautifully simple insights. First, it is intuitive that some sets do not belong to themselves – do not include themselves as members or elements. The set of Frenchmen, for instance, is not French itself (it's not a person of French nationality, for it is a set, an abstract object). Therefore such a set does not belong to the set of Frenchmen. The set whose only member is Juliette Binoche is not Juliette Binoche (a French actress, thank God, is not a set), so it does not belong to itself either. But, still intuitively speaking, it seems that some sets *do* belong to themselves: the set of all sets with more than one element, for instance, has more than one element, so it should be a member of itself. The set of anything but Juliette Binoche is not Juliette Binoche (thank God), so it should also be a member of itself. (Don't you scent in these self-membered sets the fragrance of self-reference? Recalling the Liar, you should also smell the danger.)

The sets which don't belong to themselves are often called "normal." This leads us naturally to consider the set of all normal sets, which is usually called "R" after Russell:

$$\mathrm{R} = \{x \mid x \notin x\}.$$

Translating into English: "R is the set of all and only those things x that are not members of themselves." This set of non-self-members originates the paradox that caused Frege's sorrow – with the decisive help of the Abstraction Principle. Since the schematic $\alpha[x]$ in (AP2) stands for any condition or property, we can take precisely the property of *not being a member of oneself*, $x \notin x$, and we get:

$$\exists y \forall x (x \in y \leftrightarrow x \notin x).$$

Translating into English: "There is a (set) y, to which any (set) x belongs if and only if x does not belong to itself." So there exists a set and, by the Extensionality Principle, *the* set, corresponding to such a condition, i.e., y is precisely R:

$$\forall x(x \in \mathrm{R} \leftrightarrow x \notin x).$$

Translation: "For all x, x belongs to R if and only if x does not belong to itself." Now, R in its turn is something about which we can speculate, given any property or condition, whether it has that property or satisfies that condition, or not. This is the case also with the property of not being a member of oneself. Since what holds for any x holds for R, we have:

$$\mathrm{R} \in \mathrm{R} \leftrightarrow \mathrm{R} \notin \mathrm{R},$$

that is, R belongs to itself if and only if it doesn't. This is "*the* contradiction," as Russell called it (it has the shape of a biconditional, but we easily get an explicit contradiction of the form $\mathrm{R} \in \mathrm{R} \wedge \mathrm{R} \notin \mathrm{R}$ via elementary logical steps). It follows via simple reasoning from the Abstraction Principle, which was assumed to be a quite obvious basic principle of set theory. Logicists discovered that if (AP) is assumed without restrictions, allowing any property or characterizing condition to deliver the corresponding set, then the consideration of some properties, such as non-self-membership, leads straightforwardly to paradoxes.

14 Cantor's diagonal argument

Russell's paradox is a simplified variant of a paradox deducible within naïve set theory and known to Cantor since 1899, even though it was published only in 1932. This begins with consideration of the *universal* set, which is most often indicated (after Peano) as V. V is usually characterized by means of a condition anything is expected to satisfy, such as self-identity:

$$\mathrm{V} = \{x \mid x = x\}.$$ [29]

[29] See e.g. Fraenkel, Bar-Hillel, and Levy (1973), p. 124.

This is just the set of everything. But V can be taken as the set of all *sets* if we take a pure theory of sets, that is to say, if we assume that the domain described by the theory does not include *Urelemente*, objects that are not sets. It is not difficult to have V afford us a contradiction, via a line of reasoning just slightly more complex than that involved in Russell's paradox. To obtain this new paradox, one has to consider *Cantor's Theorem*: the fundamental theorem due to Cantor, claiming that the power set P(A) of any given set A (that is, as we know, the set of all subsets of A) has larger cardinality than (so is "bigger than") A: P(A) > A.

The key to the theorem lies in Cantor's ingenious construction called the *diagonal argument*. Cantor initially used the argument to show that the set of natural numbers is not the largest infinite set, for it is exceeded by the set of real numbers (informally, the numbers represented by an infinite decimal expansion, such as the famous π: 3.141593...). Before Cantor, mathematicians were aware of the fact that the real numbers are somehow more numerous than the naturals,[30] even though there are infinitely many natural numbers. But to make full sense of the intuition one has to clarify the idea that one infinity can be "greater" than another, and therefore not "equinumerous" with it. It was precisely the idea of equinumerous sets, that is, of sets whose elements can be paired one-to-one via a bijective correspondence, that provided the required clarification.

Cantor's diagonal argument begins by assuming, for the sake of a *reductio ad absurdum* (that is, a refutation of the assumed thesis), that there is such a one-to-one mapping between natural and real numbers. An infinite set whose members can be paired one-to-one with the naturals is said to be *countably infinite*, or *(d)enumerably* infinite, or denumerable. The elements of such a set can theoretically be arranged in a list – an infinite one, of course, and therefore one that we could never finish writing down in practice, but such that (the name of) every member of the set will appear sooner or later in the list, an acceptable list being such that each member appears as the *n*th entry for some finite *n*.

Now, let us assume that the set of real numbers is enumerable. This means that we could have a list of all the (numerals for) real numbers,

[30] Or than the rational numbers, which, despite having the property of *density* (that is, the property that between any two rationals there sits another – infinitely many others, in fact) are "as many as" the naturals.

which would look like an "infinite square" or matrix, such as the following:

1/3	=	0.	**3**	3	3	3	…
1/2	=	0.	5	**0**	0	0	…
√0.1	=	0.	3	1	**6**	2	…
√0.5	=	0.	7	0	7	**1**	…
…	=	…	…	…	…	…	…

Of course, we could list the real numbers in a different order, thereby having different squares, but this is irrelevant. What matters is that, were such a list possible, we would have enumerated them all. Then we could easily have each item correspond one-to-one to a natural number. But consider the real number – call it r – whose decimal expansion is as follows: the first decimal digit is equal to the decimal digit of the first number in the list, increased by 1 (when we have a 9, we always turn it into a 0); the second digit equals the second digit of the second number in the list, increased by 1; … ; the nth digit equals the nth digit of the nth number in the list, increased by 1; and so on. We are interested in the bold-faced decimals in the square (which make us see why Cantor's argument is labeled "diagonal"). So the number at issue is $r = 0.4172\ldots$. Now r cannot be in the list: for it differs from the first number in the list at least in the first decimal; it differs from the second at least for the second decimal; …; from the nth at least for the nth decimal; and so on. The list does not include r, and so is incomplete, against the initial assumption. The procedure holds for whichever way one tries to constrict the real numbers in a list: we can always produce an element (whose identity, certainly, will vary according to the way in which the list is constructed) that cannot appear as an item in the list. All in all, the set of real numbers is not enumerable: it is actually larger than the set of natural numbers.

But this is just the beginning. Given *any* infinite set, Cantor's Theorem in its general version tells us that we can always have a larger infinity: we just have to consider the power set of the infinite set we began with.[31]

[31] Here is a general account. I have said that two sets x and y have the same cardinality (let's write: $x \cong y$), or are equinumerous, iff there is a one-to-one correspondence between them: a function mapping each member of x to a member of y, and such that

Now consider the universal set V, i.e., the set of all (pure) sets, and take its power set P(V). Since all members of P(V) are sets, P(V) is a subset of V. But V is itself a subset of P(V). Therefore, P(V) = V. So there is a one-to-one correspondence between V and P(V) (namely, identity), and P(V) $\cong$ V. But Cantor's theorem rules this out for any set, so we have:

$$P(V) \cong V \wedge \neg(P(V) \cong V).$$

Even more rapidly: given Cantor's theorem, P(V) is bigger than V. This is inconsistent with the fact that V is, by definition, the most

---

each member of y is mapped to by a single member of x. Then, x is *bigger than or equal to* y ($x \geq y$) iff there is a subset of x which has the same cardinality of y; and x is bigger than y ($x > y$) if $x \geq y$ but it is not the case that $x \cong y$. Now Cantor's Theorem says that there are no functions from x *onto* its power set P(x) (the set of all subsets of x), that is to say, P(x) has a greater cardinality than x: P(x) > x. It is easy to see that P(x) $\geq x$; the tricky part consists in showing that it is not the case that P(x) $\cong x$. Cantor's proof begins by assuming – again, for the sake of a *reductio* – that there is a one-to-one function ϕ from x to P(x), so that they have the same cardinality. Now the diagonal argument goes as follows: consider the set z of all the elements of x that are not members of the set assigned to them by ϕ – so $z = \{y \in x \mid y \notin \phi(y)\}$. z is a member of P(x), since it is a subset of x. So there must (by supposition) be an element w of x, such that $z = \phi(w)$. The question is: is w a member of z, i.e., $\phi(w)$, or not? We have:

$$w \in \phi(w) \leftrightarrow w \in \{y \in x \mid y \notin \phi(y)\} \leftrightarrow w \notin \phi(w).$$

Given the Law of Excluded Middle, either w is a member of $\phi(w)$ or not, hence:

$$w \in \phi(w) \wedge w \notin \phi(w).$$

The contradiction so deduced leads us to conclude that there cannot be such a one-to-one mapping. Roughly, the strategy is always the same: given a group of objects of a certain kind, the diagonal construction allows us to define an object that cannot be in the group "by systematically destroying the possibility of its identity with each object on the list. The new object may be said to 'diagonalise out' of the list" (Priest (1995), p. 119), as happened with our allegedly complete enumeration of the real numbers with respect to the number r, which was so constructed as to be distinct from each item in the list.

inclusive of all sets: V would have to be bigger than itself! And this is Cantor's paradox.[32]

Another paradox I shall introduce very quickly is Burali-Forti's. This is important both historically, since it was the first to be discovered, and theoretically, since its proof is a direct one (no excluded middle is required). The paradox concerns ordinal numbers. Cantor's initial idea was that ordinals should index *well-ordered* sets. A well-ordered set is a set such that each of its non-empty subsets has a least element (following von Neumann's later idea, an ordinal can correspond to the set of the preceding ordinals: so if 0 is $\varnothing$, 1 is $\{0\}$; 2 is $\{0, 1\}$; 3 is $\{0, 1, 2\}$; and so on). Now consider the set Ω of *all* ordinals. One can give independent arguments for both $\Omega \in \Omega$ and $\Omega \notin \Omega$. By construction, Ω is itself well-ordered, so since any well-ordered set has an ordinal number, Ω must have an ordinal too. However, this ordinal must be greater than any member of the set, and therefore it cannot be in the set.[33]

15 Self-reference and paradoxes

I have dealt with the details of the proofs quite quickly and informally, but this rapid presentation of the most famous set-theoretic paradoxes might nevertheless look a bit technical. What matters to us is that these paradoxes were at the origin of the crisis in the foundations of mathematics at the beginning of the twentieth century. The correspondence with Dedekind shows that Cantor was aware of the paradoxes – particularly of the fact that the universal set was an anomaly with respect to the diagonal argument. I should also add that he wasn't too bothered. Being a religious man, he had a certain tendency to see the whole situation with a mystical eye:

[32] Thus Fraenkel, Bar-Hillel, and Levy (1973), p. 7. Cantor's paradox is nothing but Russell's, once one chooses as ϕ the identity function (therefore, sometimes scholars speak of a unique Cantor–Russell paradox). In this case, $\phi(w)$ just *is* w, i.e., $\{y \in x \mid y \notin y\}$, and the thing goes like:

$$w \in w \leftrightarrow w \in \{y \in x \mid y \notin y\} \leftrightarrow w \notin w.$$

[33] See ibid., p. 8.

> I have no doubt at all that in this way we extend ever further, never reaching an insuperable barrier, but also never reaching any even approximate comprehension of the Absolute. The Absolute can only be recognized, never known, not even approximately.[34]

But Russell (who was not a man of faith) saw the implications of Cantor's theory of infinite sets and numbers going down in flames. I have claimed that the distinction between semantic and set-theoretic paradoxes, due to Ramsey, came after the publication of Russell and Whitehead's *Principia mathematica*. Russell believed that all the logical paradoxes had their root in some form of circularity, or self-referentiality, which he had named "reflexiveness":

> In all the above contradictions … there is a common characteristic, which we may describe as self-reference or reflexiveness. The remark of Epimenides must include itself in its own scope. If *all* classes, provided they are not members of themselves, are members of [R], this must also apply to [R]; and similarly for the analogous relational contradiction … In the case of Burali-Forti's paradox, the series whose ordinal number causes the difficulty is the series of all ordinal numbers. In each contradiction something is said about *all* cases of some kind, and from what is said a new case seems to be generated, which both is and is not of the same kind as the cases of which *all* were concerned in what was said.[35]

Russell's solution came with his *theory of logical types*, which he proposed in various papers, and incorporated in the *Principia*. The theory developed a rigid hierarchy of types of objects: individuals, sets, sets of sets, sets of sets of sets … (something quite similar to the Tarskian hierarchy of metalanguages).[36] What belongs to a certain logical type can be (or not be) a member only of what belongs to the immediately superior logical type. The membership relation can hold, or fail to hold, only between an individual and a set of individuals; or between a set of individuals and a set of sets of individuals; and so on. The construction allows any set to contain only things of one order: it allows only sets composed, so to speak, of objects that are homogeneous with respect to the hierarchy. Therefore, there is no set of all sets, or of all ordinals, etc. Such sets

[34] Hallett (1984), p. 42.

[35] Russell and Whitehead (1910–25), pp. 61–2.

[36] Of course, the historical succession is reversed with respect to my exposition. Tarski presented his theory of truth and his hierarchic approach several years after Russell.

would have to be constituted by members of totally heterogeneous kinds, that is, by things belonging to different levels in the hierarchy. As is clear, the whole construction aims at ruling out self-referential expressions such as "$x \in x$", or "$x \notin x$". These are now rejected as ill-formed: they are taken as simply meaningless. For instance, Russell's paradox disappears because a set can neither be, nor not be, a member of itself. For the same reason, Cantor's paradox disappears because there can be no set V of all sets: we cannot even *say* within the theory that a set contains all sets.

Other rigorously axiomatized theories of sets, developed at the same time as Russell's as well as later, are based upon a general principle that has come to be called the principle of the *limitation of size*, and which, roughly, prohibits some very comprehensive sets. The mainstream and most popular axiomatic set theory nowadays is that proposed by Zermelo, developed and modified by Fraenkel, known as **ZF**, or as **ZFC**,[37] depending on whether or not it includes a set-theoretic axiom called the *Axiom of Choice* (but we can skip the details). Other axiomatic theories, such as those proposed by von Neumann, Bernays, and Gödel himself, introduce a distinction between *sets* and *classes*.[38] It is claimed that classes, as the extensions of some very comprehensive predicates, cannot be taken as full-fledged mathematical objects capable of set membership. Things can be members of sets or classes, and sets can be members of classes, but classes can be members of nothing. Most of these theories, despite avoiding the known paradoxes, have to abandon intuitively plausible or philosophically important sets, such as the total set V (an exception is Quine's system **NF**, which however is not particularly popular among mathematicians). In all the mainstream accounts, the Abstraction-Comprehension Principle has to go, and its work is now carried out as far as possible by weaker and more limited principles. But we don't need to go into the details of axiomatic set theories, for now we have more or less all the set-theoretical notions required to understand the rest of this book.

Before we begin to play the Gödelian symphony, however, we have to take into account another chapter in the story of the crisis in the foundations of mathematics, concerning a development indispensable to an understanding of Gödel's position: the advent of Hilbert's Program.

[37] In the following I will use boldfaced capital letters to label famous formal systems and theories.

[38] The terminology is not uniform: sometimes in this context sets are also called *classes*, and classes in the strict sense are called *proper classes*.