The Multiverse: a Very Short Introduction
© Jeremy Butterfield: Draft 30iii2022: please do not cite

Chapter 3: All the logically possible worlds

In this Chapter, I will proceed in four stages. At the end of the last Chapter, I said that although logicism failed, it was a major reason why early twentieth-century philosophy placed logic centre-stage. Since then, logic has remained central, and this Chapter's first stage will be to present some details about this. We will then be ready to discuss the philosophers' multiverse, in three further stages.

In the second stage, I will urge that in everyday life, and technical science, and philosophy, we are up to our necks in modality. This word is philosophers' jargon for the topic of necessity, possibility and impossibility. That is: in order to state what we believe to be true, whether in everyday life or in technical science, we need to accept non-actual possibilities. Once we see this, it becomes clear how, by about 1970, philosophy was ripe for the proposal that there is a multiverse of all the logically possible worlds.

In the third stage, I will sketch some of the benefits for philosophy, of adopting an explicit framework of a set of possibilities. The prototypical example of such a set---a cautious prototype, in Chapter 1's spectrum of attitudes---is the set of instantaneous possible states of some physical system, as postulated by some physical theory. As I shall explain, this set is called the state-space of the theory. But more ambitiously, one might accept maximally specific possibilities for the cosmos as a whole. These are the possible worlds. So one envisages a set W of all the possible worlds. The exact nature or status of these worlds thus becomes this Chapter's main concern.

One can take either a cautious or a confident attitude to them. The most confident attitude says: 'They are all equally real; the non-actual worlds are just not "hereabouts", in much the same way that for a person in England, all the other countries e.g. France and Australia, are equally real, but just not hereabouts.' Agreed, that is hard to believe. And indeed: almost no philosopher does believe it. But the great philosopher David Lewis, who thought hard and deeply about possible worlds, believed it. The doctrine is called modal realism.

Lewis argued for it at length; (especially in his book, On the Plurality of Worlds (1986)). He did not claim to have a knock-down i.e. irrefutable argument. As we discussed in Chapter 1, in philosophy such arguments cannot be expected. Rather, he argued that modal realism was on balance better than the rival, cautious, conceptions of possible worlds.

But he also agreed that most of the philosophical benefits of using a set of logically possible worlds do not require his modal realism. They can also be had while adopting much more cautious conceptions of what the worlds are. So in the third stage, I will show how various philosophically important concepts and doctrines can be made precise in terms of the framework of possible worlds. There are many such concepts and doctrines. But I will restrict my examples to ones we will need in later Chapters.

Finally in the fourth stage, I turn to the outstanding question: What exactly is a possible world? This question is compulsory, for cautious conceptions of possible worlds as much as confident conceptions, in particular Lewis' modal realism. Several possible answers are defended in the philosophical literature. But to avoid anti-climax, I announce now, at the outset, that I will not settle on one answer. So the Chapter will end inconclusively, and perhaps disappointingly. For I will leave this question hanging, without endorsing any answer. But there is some consolation: the following Chapters will not depend on my having endorsed an answer. Besides, the next Chapter might help. For it will suggest a new answer, derived from quantum physics.

Chapter 3.1: The legacy of logicism: the endeavour of reduction

Since 1900 logic has been central to philosophy, in two main ways: which I take up in this Section and the next. The first way amounts to the legacy of logicism. Although logicism failed because (as we discussed at the end of Chapter 2) set theory is not really the same as logic, logicism nevertheless engendered two broader visions which have persisted. They both involve the idea of reduction; and they are the topic of this Section.

The first vision is about pure mathematics. And it has not merely persisted after the demise of logicism. It has been, in effect, proven to be true, through detailed work by various mathematicians from about 1890 to 1920. (The second vision, discussed below, is about philosophy, especially about what the task (or at least, one task) of philosophy should be. Philosophy being controversial, this vision remains of course unproven.)

This first vision is easily stated. As logicism developed, it became clearer that its task, of proving that all of pure mathematics was really logic, amounted to two sub-tasks: first, show that all of pure mathematics, e.g. arithmetic, the calculus, geometry etc., can be written in terms of a (paradox-free) theory of sets; second, show that this theory of sets is really logic in disguise. So even if---as agreed---we cannot do the second sub-task, i.e. set theory is not logic, we can still complete the first. And this was indeed achieved, by the collective work of various mathematicians.

Thus by about 1910, there was a vision, endorsed by many opponents of logicism as well as by its advocates, that set theory is a universal framework in which to formulate all of pure mathematics. More precisely, the vision says: a paradox-free set theory adequate for formulating all mathematics can be written in a formal language, with precise vocabulary, rules of grammar, and of inference (as discussed at the end of Chapter 2).

Indeed, the requisite formal language is very simple. It has exactly one basic predicate, representing the relation of set-membership. This is always written with the Greek letter epsilon, ε. So in set theory, 'x ε y' means that x (itself a set) is an element of the set y.

Besides, the rules of grammar, and of inference were also very simple. They were the rules proposed for predicate logic that had been invented by Frege in 1879. Here, 'predicate logic' comprises the logical behaviour of both (i) 'and', 'or' and 'not' (called 'propositional logic', or 'Boolean logic') and (ii) 'every' (similarly: 'any', 'all'), 'some' and 'none'. (Here, 'or' is understood inclusively, as synonymous with 'and-or'. So 'Bill is tall or blond' is true if Bill is both tall and blond.)

Thus predicate logic is concerned with valid patterns of argument whose validity turns on the placing of these words within the argument. Here are two examples, (1) and (2); example (2) also uses some propositional logic.

(1): Premise: 'Some As are Bs'. Premise: 'All Bs are Cs'.
So, Conclusion: 'Some As are Cs'.

(2): Premise: 'Some As are Bs'. Premise: 'All Bs are Cs or Ds' (meaning: 'any B is a C or is a D'; not 'all Bs are Cs, or all Bs are Ds').
So, Conclusion: 'Some As are Cs or Ds' (where 'or' is again understood as inclusive).

Thus the vision had three parts. The first part is about set theory; the second about pure mathematics (apart from set theory); and the third part about how to show that the second part can be understood as included in the first part: as follows. I will label the parts (A), (B) and (C).

(A): There is a paradox-free formulation of set theory in a formal language with just one basic predicate, '… ε …', representing set-membership; and whose rules of grammar, and of inference, are just those of predicate logic. So in this language, the only allowed inferences are those that depend on the words listed in (i) and (ii) above, like my examples (1) and (2). Indeed this formulation of set theory is an axiomatization: all the theorems, all the truths of set theory to be appealed to, follow by these allowed inferences from a few initial axioms.

(B): Take all the accepted truths of pure mathematics, apart from set theory: the truths of arithmetic, of the calculus, and of geometry and the other traditional areas of mathematics. Here, 'accepted truths' means: claims accepted as proved by mathematicians. Of course, the usual formulations of these truths, in textbooks etc., are enormously varied, in that the different areas have their own special vocabularies. Arithmetic has the numerals '1', '2',…, the ratios (rational numbers), '2/5', '42/9', …, the signs + and x for addition and multiplication. Geometry has nouns for geometric objects, e.g. 'point', 'line', 'triangle', and predicates for relations between them, e.g. 'intersects', 'is perpendicular to'.

(C): Despite the sparse simplicity in (A), and the variety and complexity in (B), it is possible to give an explicit definition of each of the special vocabulary items, in each of the many areas of mathematics in (B), in terms of sets, in such a way that: once we add these definitions to the sparse and simple set theory (A), each of the claims of (B)---now understood, using the added definitions, as claims about certain sets---can be derived within (A), using only (A)'s strictly limited rules of inference.

Here of course, (C) is the punch-line. It offers you re-interpretations of your traditional familiar mathematical words, e.g. the numerals '1', '2',…, '2/5',… 'intersects', 'is perpendicular to', in such a way that all the mathematical claims you accept, if thus re-interpreted, follow, by simple and compelling rules of inference, about 'and', 'or', 'all, 'some' etc., from the axioms of a simple and compelling set theory. In short: (C) shows a way to interpret (B), i.e. the truths in (B), as really "already there" in (A).

Philosophers and logicians call this a <u>reduction</u> of (B) to (A). So (C) is the claim that each of the traditional areas of mathematics (and so also: the grand conjunction of all their accepted claims) can be <u>reduced</u> to set theory. So set theory is called the <u>reduction-basis</u>.

Of course, the definitions offered of the traditional familiar words must be judiciously chosen. For if you define these words in terms of sets wholly at random, it will only be by the greatest coincidence that your beloved mathematical truths, e.g. '2+2=4', 'there are infinitely many primes', 'all equilateral triangles are equiangular', turn out to be theorems of set theory. Very probably, your haphazard definitions will render these claims as false statements of set theory; or even as not a grammatical sentence about sets at all.

On the other hand, needing to choose judiciously does not mean that there is only one choice that would work. For example, here is a great variety in which set to choose as the interpretation of the numeral '1'. But having made a choice, your choices for the other numerals, '2', '3'…, and so for other number-expressions like '2/5' etc. for the rationals, are heavily constrained. For they need to "align" or "mesh" with your choice for the numeral '1', if your accepted truths are to follow as theorems of set theory.

So let me sum up this vision. It is (C) that was achieved---proven true---by various mathematicians from about 1890 to 1920. It is a very remarkable achievement. Indeed, it is undoubtedly one of the greatest transformations in the entire history of mathematical thought.

Nowadays, this achievement is, as the saying goes, hidden in plain sight. Both research articles and pedagogic writings (textbooks) usually start by invoking the framework of set theory (almost always informally, without mentioning axiomatization), and then proceed informally, in natural language augmented with mathematical symbols. They never mention that the proofs of all the text's theorems can be formulated without loss, using the very limited rules of inference endorsed by the predicate logic.

(Of course, becoming hidden in plain sight is often the fate of major changes. They become ubiquitous, entrenched---and unnoticed. Another example in the history of mathematics is the adoption of Arabic in place of Roman numerals. The advantages for addition and multiplication are so great that we hardly ever think of adding or multiplying Roman numerals, and so we forget how cumbersome it would be.)

But in the early twentieth century, this vision, and its achievement, had a large impact on philosophy. It led to what at the start of this Section I labelled as 'the second vision' bequeathed by logicism: a vision about what the task (or at least, one task) of philosophy should be. So to this, I now turn.

As discussed in Chapter 1, much of philosophy has throughout the centuries been about "conceptual house-keeping". That is: scrutinizing concepts to see if they are in order, and if so, giving an account or even an analysis of them; (and if they are misleading, rejecting or maybe revising them). One even sees this at the beginning of Western philosophy, in Plato. Socrates besets the people whom he accosts in the agora (market-place), with requests for definitions (analyses) of virtue, courage etc. And much of philosophy since---about many diverse topics, such as virtue, free will, knowledge, causation, number or necessity---can be read as aiming to give an account of the concept in question; and maybe even an analysis of it.

Here, 'giving an account' means describing how the concept relates to other kindred concepts (e.g. one implies the other, or one tends to cause the other); and stating what are the important accepted truths involving the concept (and of course, kindred concepts). And 'giving an analysis' means something more specific and ambitious: defining the concept in terms of previously understood concepts (and so displaying their logical connections), in such a way as to recover the accepted truths involving the concepts. And here, 'to recover' means, ideally at least: to derive, i.e. deduce, from other accepted truths invoking the previously understood concepts.

Thus we return, in the more general context of philosophy, to the above idea of reduction. If the scrutinized concept or concepts are considered to be in order, then we can aim, ideally, to deduce the accepted truths invoking them, viz. (B) in the above labelling, by adding to a previously understood and accepted body of doctrine (A), some judiciously chosen definitions, analyses, of (B)'s concepts in terms of (A)'s.

The second vision is now clear. Seeing mathematicians' achievement of reducing all of traditional pure mathematics to the sparse and simple framework of set theory and predicate logic, philosophers conceived the task of similarly reducing accepted bodies of doctrine about other matters: in particular physical theories, or even everyday propositions about the empirical world.

Of course, philosophers differed about the details of the proposed task. Russell with the programme (ca. 1910 to 1920) that he called 'logical atomism' proposed to analyse all our everyday empirical knowledge, as did Carnap with his Aufbau programme (1928). But their contemporary Reichenbach aimed in his 1920s work "only" to axiomatize Einstein's relativity theories. But these programmes had much in common. In particular, they agreed on the answer to the immediate question, 'What is the previously understood and accepted body of doctrine to which you propose a reduction should be made?'. Namely, a staunchly empiricist answer: propositions about sensory experience.

Thus the programmes of Russell's logical atomism, and somewhat later, the logical empiricism of Carnap, Reichenbach and others in Vienna and Berlin, should be seen as modelled on the successful set-theoretic (though not logicist) reduction of pure mathematics.

Chapter 3.2: Logic as a toolbox of formal systems: modal logics

Clearly, the reduction programmes of Russell and Carnap were very ambitious. Everyday empirical knowledge is a vast open sea. It far outstrips a single knowing mind; its content shades continuously into technical science; and we have no agreed chart for it, i.e. no agreed taxonomy breaking it down into parts appropriately (e.g. logically) related to one another. Besides, we have no agreed language in which to talk about the reduction-basis, i.e. sensory experience. In my jargon above: there is no uncontroversial 'previously understood and accepted body of doctrine'. So unsurprisingly, these programmes failed. As the Bible warns us: pride comes before a fall (Proverbs 16:18).

But programmes with a much more modest aim---for example, axiomatizing a single physical (not: pure mathematical) theory, using predicate logic and a basic vocabulary that was small, but not required to be solely about sensory experience---fared much better. A single physical theory, such as Newton's theory of gravity or Einstein's special relativity, is pretty well-defined. The textbooks largely agree in how they present it to us, and in what its special vocabulary is. And in axiomatizing it we do not need to reach for some other vocabulary, e.g. solely about sensory experience, and for some doctrine using that vocabulary, to serve as a reduction-basis. Rather, the axioms we seek will be the reduction-basis. Nor was it just philosophers like Reichenbach who undertook such efforts. Mathematicians, including great ones like Hilbert and von Neumann, did so too.

Thus arose a more modest and flexible conception of the role of logic in philosophy, which has persisted till today. Namely, as a resource, a toolbox, for formalizing various bodies of doctrine, without necessarily axiomatizing them or reducing them to another body of doctrine. Of course, the bodies of doctrine are to be chosen because of their philosophical interest. They use concepts central to everyday life and thought (like my list above: virtue, free will, knowledge etc.) and-or science (like space, time, matter, causation). And so this conception goes along with philosophers' traditional endeavour of conceptual analysis.

Nowadays, there are countless such examples of "logic in action". (Indeed there are even logics of action, as well as logics of concepts that seem more amenable to a logical treatment, such as knowledge.) We already saw one example of this in Chapter 2. It was about what it is rational to believe---what principles should govern what we believe?---in addition to the indisputable requirement that we should believe the deductive consequences of what we already believe. Thus since the mid-twentieth century, philosophers have developed formal systems prescribing how to change your beliefs when you get evidence (often called 'inductive logics').

For the purposes of this Chapter, the most important example is of course: the logic of modality. (Recall that 'modality' is jargon for the topic of necessity, possibility and impossibility.)

Aristotle himself initiated this, by discussing such principles as that necessity implies truth. That is: if a proposition is necessary (must be true), then it is in fact true. And similarly, truth implies possibility: if a proposition is in fact true, then it is possibly true. (For the actual situation counts as one of the possibilities. Here, we set aside the conventional rule of conversation whereby calling something 'possibly true' connotes its being in fact false.) The natural way to think of such principles is that the phrase 'It is necessary that …' has an empty slot or argument-place … into which a sentence 'P' can be inserted, to produce a sentence 'It is necessary that P'. So there is a valid argument: 'It is necessary that P; therefore P'. Similarly, 'P; therefore it is possible that P' is a valid argument. And as I mentioned in Chapter 2: to these valid arguments, there correspond conditional propositions that are themselves necessary. Namely: 'if it is necessary that P, then P'; and 'if P, then it is possible that P'.

Medieval logicians developed the logic of modality. But as we have seen in Chapter 2, philosophy in the modern period, i.e. from the seventeenth century, neglected logic up until the late nineteenth century. And then, although logicians like Frege and Russell took logic to be a collection of necessary truths, they showed no interest in studying the logic of modality, i.e. studying the logical behaviour of phrases like 'It is necessary that …', and 'It is possible that …'. Thus the logic of modality lay dormant until spear-headed in about 1915 by the Harvard philosopher, Clarence Lewis; (usually cited as 'C.I. Lewis': no relation of David Lewis---about whom, more shortly).

C.I. Lewis was the first person to write down formal logics of modality, called 'modal logics'. They build on the logics we noted in the previous Section. Thus recall that propositional logic comprises the logical behaviour of (i) 'and', 'or' and 'not'; while predicate logic adds to this the logical behaviour of (ii) 'every', 'some' and 'none'. C.I. Lewis proposed adding to any system of propositional logic a new symbol, which I write as 'N(…)' (for 'necessary') which accepts a sentence 'P' in its argument-place … to make another sentence, 'N(P)': which we read as 'it is necessary that P'. It follows that 'It is possible that …' does not need a separate treatment. For recall that 'not' makes a sentence from a sentence: 'not-P' is true if P is false, and vice versa. (Any piece of language that makes a sentence from a sentence, like 'N(…)' and 'not', is called a sentence-operator.) So 'It is possible that P' can be rendered as 'not-necessarily-not-P'. That is: as 'not-(N(not-P))'.

So far, so straightforward. But building such a system of modal logic soon leads to interestingly controversial issues. For sentence-operators can be iterated. So what should we say about 'NN', in particular in relation to 'N'? One may well be content that the argument 'N(N(P)); therefore N(P)' is valid, whatever our choice of proposition P. (For it is itself an instance of our previous valid form 'N(P); so P'.) But what about the converse argument, i.e. 'N(P); therefore N(N(P))'? Is this second form of argument valid, for all choices of P?

On such questions, C.I. Lewis himself took a liberal view. He developed various systems of modal logic, that obeyed various sets of principles, while sharing those I began with: that 'It is necessary that P; therefore P' is a valid argument, and 'It is possible that P' is rendered as 'not-necessarily-not-P'.

Matters become even more controversial when one considers how 'N' should behave in relation to the 'every', 'some' and 'none' of predicate logic. Thus suppose 'is F' is some predicate, e.g. 'is red' or 'is a horse'; and suppose that 'For every object, it is necessary that the object is F' is true. Does it follow---is it valid to infer---that 'It is necessary that for every object, it is F', i.e. 'It is necessary that every object is F'?

There is good, though I think not compelling, reason to deny this. For the premise is naturally read as about all actually existing objects: and as saying of each of them that it is necessarily F, i.e. that however the world happened to be, the object would be F. Notice here how natural it is to say 'world' i.e 'possible world'. But the conclusion is naturally read as: however the world happened to be, every object in that world would be F. So if we envisage that the world could contain objects that it actually does not contain, then the way is open to denying that the inference is valid. For we can admit the premise, that all actually existing objects must be F, but insist that there could be yet other objects: some of which, in some worlds, are not F.

On the other hand, this reason is not compelling. For it seems tenable that the actual world is "privileged" among all possible worlds, in being "the ultimate resource" for objects. That is: any object that possibly exists, actually exists. So the idea is: "no newcomers are allowed into view, as my mind's eye goes from the actual world to some other possible world."

So the interplay between modality and the notion of object is controversial; and the controversy shows up in questions about which principles combining the modal operators with the 'every' etc. of predicate logic we should accept. These controversies were pursued by C.I. Lewis and others (including Carnap) in the mid-twentieth century. They were also much clarified and enlivened in the 1960s by the work of David Lewis, Kripke, Kaplan and others: all of whom emphasized the semantics of modal logic. This semantics explicitly invoked possible worlds, and so made vivid the central question of this Chapter: what exactly are possible worlds? And as we shall see in the next Section but one, this semantics also led to detailed proposals about the semantics of natural languages.

This completes this Chapter's first stage: a summary of logic's role in philosophy up to about 1970, especially the development of modal logics. As I announced in the Chapter's preamble, we are now ready to discuss the proposed multiverse of possible worlds, in three further stages.

Chapter 3.3: Up to our necks in modality

In this Section---which is the second stage of the Chapter---I will argue that in order to state what we believe to be true, whether in everyday life or in technical science, we need to accept non-actual possibilities. Then the main question for the rest of the Chapter will of course be: exactly what does this commitment involve?

Let us begin with our beliefs in everyday life. Consider some belief of yours that is true. It can be utterly mundane, e.g. that grass is green. Then the negation of what you believe, the proposition that grass is not green, is false. It represents a non-actual possibility: but what exactly is that?

There is a temptation to dismiss this question, saying that after all, grass could not fail to be green, thanks to its genetic make-up encoding that it produce chlorophyll, i.e. the green pigment essential for photosynthesis. But this dismissal is unconvincing. For suppose we agreed that grass must be green, and also that there is no need to accept the impossibility, grass not being green, as some sort of ghostly non-fact---we can take the impossibility to be nothing at all. Nevertheless, there are surely countless everyday propositions that are in fact false but could be true. Suppose I stay at home tonight: then the false proposition 'I go to the cinema tonight' surely could be true. (For this example, it does not matter whether I have free will, i.e. whether I could freely choose to go to the cinema. The example only needs that 'I go to the cinema' could be true.) So there is a way the world could be that makes this proposition true. So accepting that it might have been true commits us to such 'ways', i.e. to non-actual possibilities---in some sense. Besides, some of these propositions that are in fact false but could be true are among our beliefs---yours and mine. So we cannot duck out of the issue by just focusing on true beliefs. For any of our false beliefs that could be true has as its content, i.e. what it represents about the world, a non-actual possibility.

This discussion may seem suspiciously abstract. Let me make it more vivid by giving two main ways in which our beliefs invoke non-actual possibilities. The first way concerns deliberation and decision. Suppose a person hesitates between two options for action, deliberating which to do, and then does one. The options could, again, be utterly mundane: for example, which of two keys to try so as to unlock a door. We cannot understand the process of deliberation, what the person thinks, purely in terms of the one actual course of events that ends in, say, trying the bigger key. To explain the process of deliberation and the eventual action, we need to attribute to the person beliefs, some of which are about non-actual possibilities. For suppose the bigger key is the wrong one. So the proposition 'the bigger key fits' is false. It represents a non-actual possibility: but the person believes this proposition and acts on it.

Examples like the choice of key are the bread-and-butter of a discipline, decision theory, that lies at the interface of philosophy with economics and psychology. We shall meet decision theory again in the next Chapter: for it has a surprising application in support of Everettian quantum theory, i.e. the quantum multiverse. But for the moment, I will just state decision theory's general description of a deliberating person, so as to bring out its ubiquitous use of non-actual possibilities.

Decision theory assumes that a deliberator has:

(i) various degrees of belief, i.e. subjective probabilities, about various possible states of the world, i.e. degrees of belief in propositions about the world;

(ii) desires of various strengths that various such propositions be true; and

(iii) a set of options for action, which are again taken as propositions---propositions that the person can at will make true (like trying the bigger key).

Decision theory then formulates principles that prescribe which option for action is best for the deliberator. A common idea of these principles is that the best option has the highest 'score'. Here, a 'score' is defined as the weighted-average strength of the desired propositions (ii), where the average is computed with degree-of-belief weights given by (i). (Of course, one should also allow for first-equal scores: then the best option is any of the options with the highest score.)

This common idea can be made precise in various ways, that in some cases disagree about which option they prescribe. But we need not discuss these disagreements and the ensuing controversies in decision theory. For us it is enough that, as the common idea shows: when a person decides and acts, they are up to their necks in modality.

Turning from decision making to technical science: it also is up to its neck in modality. Of course, decision theory itself counts as science. But let me stress examples in physics. For again, this will help us prepare for the next Chapter.

Mention of subjective probabilities prompts an obvious suggestion. Namely, chance: Chances are objective probabilities that are made true by the subject-matter rather than by the state of mind of a person thinking about it. As I mentioned in Chapter 1, the standard example is radioactivity: e.g. the chance of this Uranium atom decaying in the next hour. Again, the very concept of chance commits one to non-actual possibilities, going beyond the one actual course of events. For chance requires a range of future alternatives: in my atom example, just two---atom decayed after an hour, and atom undecayed.

But even without probabilities, physics endemically invokes non-actual possibilities. This occurs in every physical theory: from the most elementary, such as Newtonian mechanics, to the most advanced, like quantum theory and general relativity. (And it occurs in speculative theories, like Chapter 5's cosmological theories and string theory, as much as in well-established theories.) To explain this, I will introduce some physics jargon, which will also be useful in later Chapters; and then consider the simplest theory, Newtonian mechanics (which is familiar from Chapter 2).

Any physical theory describes a certain kind of object by ascribing to it numerically measurable properties like position or momentum (i.e. mass times velocity) or energy. In the jargon of physics, the objects are called systems; their properties like position etc. are called quantities (also: 'magnitudes', but I will not use this word); and the amounts or degrees of such properties that are ascribed are called values (almost always real numbers). (In the jargon of philosophy, the quantities are determinables, and each of their values is a determinate. The standard philosophical example of a determinable is colour; of which scarlet is a determinate.) Then a state of a system, according to a physical theory that describes the system, is a list, or conjunction, stating what are the system's values for the various quantities that apply to it.

The state of course changes over time, as the values of the various quantities go up or down. So the state is also called the instantaneous state. A physical theory gives descriptions of these changes. In most theories (including all that this book will discuss), the theory provides an equation stating exactly how the state (the values of all the system's quantities) changes over time. This is the system's equation of motion. Typically, it fixes the rate of change of some chosen quantity (or quantities) of interest, as a function of the values of that quantity, and usually also other quantities, at some initial time. Given those other values, and thereby the rate of change of the chosen quantity, one then solves the equation so as to find the value of the chosen quantity at later times. In short, one predicts the future values of that quantity, on the basis of some present state.

Newtonian mechanics is of course the archetypal case. Imagine that a small solid object, say a sphere, is our system of interest. In Newtonian mechanics, this is usually called a 'body'. If we know the forces that are now, and that will later, be exerted on the sphere (say by other bodies, e.g. gravitational forces or electric forces), and we also know the sphere's present position and momentum: then the equation of motion for its position can be solved. That is: the position at later times (and so also the momentum at later times) can be calculated.

Agreed, two qualifications are needed. We already glimpsed the first, in Chapter 2. When two bodies collide, what happens is very complicated. They usually distort each other, or even break up, so that describing what happens often outstrips the resources of Newtonian mechanics---for example, because the collision generates heat. So let us set aside collisions: for example, by imagining the sphere is in empty space, a vacuum, and is far away from all other bodies.

Secondly, even apart from collisions, the sphere's motion can be influenced by motions internal to the sphere, for example if it is spinning or is not completely rigid. So mechanics often idealizes the situation, by imagining the sphere is so small and rigid as to be effectively extensionless: a point-particle (also called a point-mass). The instantaneous state of such a point-particle, sufficient for solving the equation of motion, is indeed just its position in space (so three real numbers, for its x-, y- and z-coordinates) and its momentum (again, three real numbers, for its mass times its speed in each of the x-, y- and z-directions). That is: the state of a point-particle is given by an ordered set of six real numbers: a 6-tuple. So for a mass m, we could write this as: $(x, y, z, mv_x, mv_y, mv_z)$.

Here, what matters most is not these qualifications, but the fact that Newtonian mechanics explicitly postulates the set of all possible instantaneous states of the sphere. And similarly for other systems that the theory describes.

For the simplest possible system, a point-particle, that means: the set of all 6-tuples of real numbers. Unlike the set of triples of real numbers, which we of course visualize as familiar three-dimensional Euclidean space, this space cannot be visualized. We should instead think of its structure as follows: at each point of physical space, i.e. at each possible position of the point-particle, we have attached a separate copy of the set of all triples of real numbers. This copy represents all possible triples of momenta in the three spatial directions, that a point-particle at that location in space could possess. It is a dizzying idea.

Besides, when we consider more and more complicated systems, the set of all possible instantaneous states rapidly becomes very intricately structured. Even for two point-particles, which we label '1' and '2', with masses m_1 and m_2, we need 12-tuples of real numbers: which we could write as $(x_1, y_1, z_1, m_1v_x, m_1v_y, m_1v_z; x_2, y_2, z_2, m_2w_x, m_2w_y, m_2w_z)$. (Here, I use 'v' for speeds for the first particle, and 'w' for speeds for the second particle.) But to set aside collisions, the two triples representing particle positions, (x_1, y_1, z_1) and (x_2, y_2, z_2), must be different. So the structure of this set is: for every pair of distinct positions throughout physical space, we attach to each position in the pair, a copy of the set of all triples of real numbers, representing all possible triples of momenta for a particle located there.

And so it goes. To write down Newtonian mechanics, we need to mention these sets of possible instantaneous states, endowed with their intricate structures. Although these sets are of course not physical space, nor located in physical space---one would naturally call them pure mathematical entities, albeit usefully applicable to physical systems---they are called 'spaces': more specifically, state-spaces. Calling a structured set a 'space' (and its elements 'points') is ubiquitous in mathematics: the rationale is that often the structure is suggested by our visual intuitions about physical space, or even by precise geometric ideas like distance.

Similarly for all other physical theories: both the classical theories developed between 1700 and 1900 of light, electricity and magnetism, and of heat; and their twentieth-century descendants, which adapted their ideas and techniques to quantum theory and relativity theory. All these theories postulate, for each system they describe, an intricately structured set of all possible instantaneous states of the system. This set is again called a state-space; though of course the quantities involved will in general not be position and momentum, as in our example of Newtonian mechanics. (Chapter 4 will give more details about the state-space for a quantum system.)

In short: the theories simply cannot be written down without describing this space. Thus I rest my case that physics is, as the catchphrase goes, up to its neck in modality. In each case, the system concerned is like a toy-model of the universe, i.e. a very simple way the world could be, according to the theory. For example, according to Newtonian mechanics, a system of two point-particles is a toy universe: the instantaneous possibilities for such a universe are the points in the two-particle state-space. And similarly for a system of five, or seventeen, or any number of point-particles. Each is a toy Newtonian universe, whose possibilities are the points of the corresponding state-space.

We can also now readily see how useful the idea of a state-space is: again, in any of these theories. A sequence of instantaneous states is a possible history of the system. (Here, 'history' means not just the system's past states, but includes future states, so that a history is an entire "life-story" of the system.) We can think of this as a curve in the state-space. Then the structure of the state-space, especially its geometric structure like distance between points, helps us to understand the behaviour of these curves, i.e. these possible histories. For example, that two curves converge represents the two histories becoming more similar, i.e. the two systems' values for quantities becoming closer.

In particular, we can now state the idea of determinism. There are various precise formulations, but the general idea is of course that the state at one time determines the state at other times. So one common formulation is that any state in the state-space determines the sequence of states for all future, and indeed all past, times. In terms of curves in the state-space: through any point of the state-space, there is a unique curve to the future, and indeed to the past.

Again, Newtonian mechanics is the archetypal case; (setting aside collisions, as I did above). Namely: given the forces that are exerted on a sufficiently rigid body (or a point-particle), not just at the given time (`now') but throughout the past and the future, any state in the state-space has passing through it a unique history or curve. (To be precise: this claim assumes not only that the forces are given throughout time, but that they satisfy some "good behaviour" properties.) So we say that Newtonian mechanics is a deterministic theory.

But I stress that many other theories are also deterministic: and not just non-quantum theories---in the next Chapter, the Everettian version of quantum theory will be deterministic. (I will return to determinism later in this Chapter, when I discuss an important philosophical notion: supervenience.)

Chapter 3.4: A philosopher's paradise

So much by way of arguing that our everyday and our scientific beliefs commit us to non-actual possibilities. In the next few Sections, I describe how a set of possibilities gives us a framework for formulating many philosophically important ideas and doctrines. (I will restrict my examples to ones we will need in later Chapters.)

By and large, the benefits of such a framework can be had, even with only a cautious or modest conception of the possibilities: for example, as the state-space of a physical theory, so that the system concerned is a "toy-universe". And even if, more confidently and ambitiously, one accepts a vast set of possibilities for the cosmos as a whole, the possible worlds, still the benefits can be had, by and large, without addressing the question 'What exactly are the possible worlds?': which I will turn to only at the end of the Chapter.

So in effect, the next few Sections are an advertisement for using the framework of possible worlds---whatever exactly they are. Thus Lewis called this framework 'a philosopher's paradise'; and I concur. (The phrase deliberately echoes the achievement I lauded above, of formulating all of pure mathematics as set theory. For the mathematician Hilbert called set theory 'Cantor's paradise', after Georg Cantor who was the main inventor of set theory.) But again: Lewis allowed---and I agree---that most of the benefits of using possible worlds do not require his modal realism. After all, he called it 'a philosopher's paradise', not 'a modal realist's paradise'.

I will take as my first example of how useful possible worlds are, semantics. More specifically: a scheme called intensional semantics, which is inspired by ideas of Frege and Carnap about how words gets their references in the world. It is clearest to start with Frege, who expresses the ideas without regard to modality or possible worlds. It was Carnap, and later writers like Montague and Lewis himself, who adopted possible worlds. (As I mentioned at the end of the first Section, Frege and his contemporaries like Russell did not think about modal logic.)

To make this semantics vivid, and to reflect the intentions of its proponents, I shall explain it with examples from natural language, and so invoke possible worlds representing the cosmos as a whole. So we will be envisaging a vast set W of all the logically possible worlds.

But as I suggested above, readers too cautious for such examples, and the worlds they invoke, could---and I say: should---still accept the scheme's ideas for some modest fragment of language with correspondingly modest possible worlds. On this cautious or modest approach, the obvious cases are: the languages and claims of physical theories; and their state-spaces. In such cases, the possible worlds will be the instantaneous states; or if change over time is a topic, the possible worlds will be the system's possible histories (curves through state-space).

Chapter 3.5: Paradise, Part I: Intensional semantics

Frege's idea is that the meaning of a word---whether a proper name like 'Plato', 'Copenhagen', 'Denmark', or a predicate like 'is red', 'walks', or 'has a heart'---has two main aspects.

The more obvious one is the referent: the object or objects in the world that, as we say, it denotes or refers to. For my examples of proper names, these are, respectively: a human being, a city, a country. By this, Frege means the "concrete" object---a human body, a conurbation etc.---located in space and time, and in all its myriad complexity: not some feature of the object, nor someone's ideas or beliefs about it. So one and the same object, i.e. referent, can be referred to in diverse ways. We say: 'Plato is the most famous pupil of Socrates', 'Copenhagen is the capital of Denmark' etc.

It is these ways of referring that are the second aspect of meaning, according to Frege. He calls it 'the mode of presentation' by means of which the word presents the referent to us (i.e. draws our attention to the referent). The idea is clearest for definite descriptions, like 'the capital of Denmark'; especially those that do not include a proper name: for example, 'the tallest human alive today'. Suppose it happens to be a tall German doctor called Gustav Lauben, who lives in Hamburg. Then clearly, 'the tallest human alive today' presents Lauben to us in a different way than, say, 'the best-known doctor in Hamburg'.

Frege's jargon for these modes of presentation is: sense. So these two definite descriptions have different senses. This is of course why 'the tallest human alive today is the best-known doctor in Hamburg' conveys useful information, going far beyond saying that a certain person is identical with themselves. Frege also argues that proper names have senses, though they are vaguer and more idiosyncratic than the senses of definite descriptions. Thus for me, 'Plato' might have the sense: the most famous pupil of Socrates; while for you, it has the sense: the teacher of Aristotle. (Agreed: for Frege to fully explain along these lines how we refer by saying 'Plato', we must each associate a sense with 'Socrates' and 'Aristotle', respectively. In fact, the name 'Gustav Lauben' is from Frege's example, used to expound this very topic. It occurs in one of his most famous essays, called 'The Thought: a logical inquiry'.)

Similarly, says Frege, for predicates. A predicate has instances: the objects it is true of. Frege says that these objects taken together, i.e. the set of them, is the referent of the predicate. But the same objects could be the referent of another predicate. One standard example assumes that all and only those animals that have a heart (i.e. a pump for a circulatory system for nutrients) have a kidney (to remove waste products). That assumption can of course be questioned, depending on the meanings of 'heart' and 'kidney'. But let us accept it. Then the predicates 'has a heart' and 'has a kidney' have the same set of instances: what Frege calls the 'referent'. But indeed, the predicates present the referent in different ways; and intuitively, they have different meanings. Thus Frege says they have different senses. And again: that is why 'an animal has a heart if and only if it has a kidney' conveys useful information, going beyond saying that a certain set of animals is self-identical.

Frege then puts these assignments to words, of senses and thereby of referents, to work in a compositional semantics. That is: he gives an account of how the senses and referents of individual expressions combine to determine (i.e. to uniquely specify) the senses and referents of the composite expressions in which they occur. Think for example of how the senses of 'tall', 'human' 'alive' etc. combine to fix the sense of 'the tallest human alive today'. Similarly: just as a proper name and a predicate combine in a simple sentence, such as 'Plato is a teacher', 'Dr Lauben walks', so also their senses combine to make a proposition. And this proposition is true if the referent of the name is in (i.e. is an element of the set that is) the referent of the predicate; and otherwise, it is false.

Frege even extends the ideas of referent and sense to propositions and to sentences. Thus he proposes that the truth-value, 'true' or 'false', of the proposition is the referent of the sentence. His rationale is, in part, that this promises a smooth treatment of compound sentences, like a conjunction 'P and Q' or a disjunction 'P or Q'. Thus the truth-value of 'P and Q' is true if and only if both sentences are true. Here, we think of 'and' as a sentence operator, '… and …', that accepts two sentences 'P' and 'Q' into its two slots or argument-places, to produce a sentence 'P and Q'. Similarly, '…or …' is two-place sentence operator. Both operators are truth-functional in the sense that the truth-value of the resulting sentence (P and Q, P or Q) is completely determined by the truth-values of the pair of input sentences. Thus 'and' is associated with a function, sending the pair '(true, true)' to 'true', and each of the other three pairs, viz. '(true, false)', '(false, true)' and '(false, false)', to 'false'.

Here I use 'function' in the mathematical sense: a rule that sends each appropriate "input" (usually called an 'argument' of the function) to an "output" (usually called the 'value' of the function for the given argument). By the way: we will see shortly that we need an innocuous generalization of the idea of a function, viz. to allow that for some arguments, the rule produces no output, no value. It is simply silent: this is called a partial function.

Thus Frege can propose that the referent of 'and' is this function, from pairs of referents of sentences as "inputs", to referents of sentences, i.e. 'true' and 'false', as "outputs". And similarly for disjunctions: the referent of 'or' (in our inclusive and-or sense) is the function taking three of the pairs of truth-values, i.e. all except '(false, false)', to 'true'. Functions like this, that map truth-values, or pairs of them, or even triples etc., to truth-values are called truth-functions. So the Fregean referent of 'and' is a truth-function; and so is the Fregean referent of 'or'.

So far, so Frege. I have not mentioned possible worlds at all; and I have invoked the actual world only "in the background", namely as making true a sentence such as 'Plato is a teacher'.

But here enter Carnap, and his followers like Lewis and Montague. They show how these Fregean ideas, about a sense being a mode of presentation of a referent, and using functions in a compositional semantics, can be smoothly developed in a framework of possible worlds.

For example: the capital of Denmark is in fact Copenhagen. But it might not have been. It could have been Aarhus, or Odense. Following Carnap, we understand this as: in some possible worlds, but not the actual one, the capital is Aarhus; while in yet others, it is Odense. Thus the referent of the definite description, 'the capital of Denmark', varies from world to world. (On the other hand, proper names like 'Denmark' seem, at least usually, to have the same referent in the various worlds.) Similarly, of course for predicates. Plato might not have been a teacher; and so the referent of 'is a teacher', i.e. the predicate's set of instances, varies across the worlds. And so on, e.g. for 'has a heart'.

All this can be neatly formulated in terms of functions. Since at a possible world, 'the capital of Denmark' picks out a city (in Denmark), we can say that the sense of 'the capital of Denmark' is a function whose arguments (inputs) are possible worlds, and whose value (output) for a given world as argument is the city in that world which is the seat of government for Denmark. And for a proper name like 'Denmark' with, we may suppose, the same referent across the worlds, we can say that the sense is again a function from worlds as arguments to objects, viz. countries, within the "argument-world". It is just that for a proper name, this function is constant: it always outputs the same value. And again, similarly for predicates. For example, the sense of 'is a teacher' is a function from worlds as arguments to the set of teachers within the "argument-world".

Here, I should make two clarifications. The first, (1), is rather technical, and not important for us. But the second, (2), is philosophically important.

(1): Agreed: we need to allow that at some (presumably vastly many) worlds, there is no country Denmark; or there is such a country, but it has no capital (seat of government). So at many worlds, a description such as 'the capital of Denmark', or a name such as 'Denmark' simply has no referent. Similarly for predicates: at a possible world with no animals with circulatory systems, the predicates 'has a heart' and 'has a kidney' will have no instances. (Here again, I assume, so as to make the point as simply as possible, that we take the meanings of 'heart' and 'kidney' to require a circulatory system.)

But this pervasive scarcity, across all the worlds, of referents causes no trouble. We simply use the idea mentioned above of a partial function. That is, we say that the sense of a word (a proper name, a predicate etc.) is a partial function: worlds are the arguments, but for many arguments, the function produces no output, no value. Agreed: as a result, the sense of a compound expression (such as a definite description) in which the given expression occurs will also in general be a partial function. Besides, one will need some sensible rules about e.g. what should be the truth-value (referent, for Frege) of a sentence at a world that contains no referent of some proper name within the sentence. But there are such sensible rules: and we need not consider them here.

(2): Beware of the preposition 'at'! That is to say: it is tempting to think that in this semantics, a phrase such as 'the referent of 'the capital of Denmark' at a given world' means: the city that within the world is called by some speakers in that world 'the capital of Denmark'. That is not so. The semantics being provided is a semantics for our language (in my examples, English), as we actually speak it. What language is spoken by people within a possible world is in general not relevant to the semantics of our language; and in particular, it is not relevant to how facts about a possible world make various sentences of ours true at that world.

Here, I say 'in general' because I agree: some of our sentences, albeit rather long and contrived ones, are indeed about a language that could be spoken---if you like, a variant of English. So according to our possible world semantics, such sentences are about a language that is spoken by people within a possible world. For example, one such long and contrived sentence is: 'People could have spoken a variant of English that used the name 'Denmark' for Sweden (but without other changes); in which case their sentence 'Stockholm is the capital of Denmark' would be true in their language.' But these contrived sentences do not alter the general point: that these semantics---though it invokes other worlds, in some of which people use our words but with different senses---are a semantics for our language: that is, for our language as we actually speak it, with our senses.

Though this point is straightforward, it is important. For as we will see at the end of the Chapter, the erroneous temptation goes along with a wrong answer to our main philosophical question: what exactly are the worlds?

To finish this exposition, note that just as Frege put his assignments to words, of senses and thereby of referents, to work in a compositional semantics referring only to the actual world: so also the scheme proposed by Carnap, Lewis et al., with senses as partial functions, gives a compositional semantics, in which senses get composed together according to the syntactic structure of the composite linguistic expressions. In particular, the sense of a whole sentence i.e. the proposition it expresses, is naturally taken as a function from worlds to the two truth-values, 'true' and 'false'. But which worlds get sent to 'true', and which to 'false', depends on the senses of the sentence's parts.

Thus consider the sentence 'Plato is a teacher'. Its sense sends a world to 'true' provided that the sense of 'Plato' takes the world to an object (in the world) that is in (i.e. is an element of the set that is) the output of the sense of 'is a teacher' for that world as input. And similarly for compound sentences. The sense of a conjunction 'P and Q' sends a world to the output of the truth-function that is Fregean referent of 'and', for inputs that are the truth-values at that world of 'P' and 'Q', i.e. are the outputs of the senses of 'P' and of 'Q'. (Another way to think of this is to identify a proposition with the set of worlds in which it is true. Then the sense of 'and' is precisely the operation of intersection on sets of worlds.)

Finally, a note about jargon. You will ask: Why is this scheme called 'intensional semantics'? The answer is that Carnap suggested using 'intension' instead of 'sense', and 'extension' instead of 'referent'. So a more informative, but long-winded, label would have been 'semantics by intensions and extensions'; but the single adjective 'intensional' was adopted. In any case, Carnap's jargon has become widespread. In particular, it is well-nigh universal usage to call the set of instances of a predicate its 'extension'. This usage is adopted even by those who are wary about intensional semantics; and setting aside talk of possible worlds, even by those who are wary about Frege's notion of sense as applied (e.g. by Frege himself) to just the one actual world.

So much by way of sketching intensional semantics, especially as it applies to names, definite descriptions and predicates. In the next two Sections, we will see how it can be readily extended to treat two further topics. First: modality, so as to give semantics for expressions like 'It is necessary that …'; and second: counterfactual conditionals, i.e. if-then sentences, whose antecedent (after the 'if') is contrary to fact, i.e. is actually false.

Chapter 3.6: Paradise, Part II: Modality and laws of nature

Intensional semantics, with its set W of logically possible worlds, also treats sentences with modal locutions, like 'It is necessary that …', 'It is possible that …', or the corresponding adverbs, 'Necessarily, …' and 'Possibly, …'. (As we discussed in Section 2 of this Chapter, a sentence P is to be inserted in the place marked by dots …. . So these are sentence operators: they make a sentence as "ouput" from a sentence P as "input". But unlike 'and', 'or' and 'not' (discussed in Section 5), they are not truth-functional.)

The main idea will of course be the intuitive (and Leibnizian) one that we already discussed. Namely: 'Necessarily, P' is true at a world w if and only if P is true at all the worlds w in W; and 'Possibly, P' is true at a world w if and only if P is true at some world w in W.

This idea gets developed in various ways. For example, one considers what is the sense or intension of 'Necessarily, …', analogous to the sense of 'and' being the operation of intersection on sets of worlds. And (as I mentioned in Section 2) one considers how this operator relates to other logical words like 'some' and 'all'.

But for this book's purposes, the development that matters is about restricting the set of worlds that a sentence operator, 'L(…)' say, requires one to check (for the truth there of the argument/input proposition P) in order for 'L(P)' to be true. That is, we need to consider notions that invoke a subset of W, not the whole of W. (So it would not be appropriate to call these operators 'Necessarily, …', 'Possibly, …' etc. But philosophers still use the word 'modality'. That is, they call these restricted notions of modality.)

One philosophically important example of such a restriction is the idea of a law of nature. (In Chapter 1, this was an example of a concept that is philosophically contentious; but that contentiousness will not undermine any points here.)

Thus someone might say: 'it is logically possible for you to fly to the Sun in less than 8 minutes, but it is not physically possible, i.e. it is not compatible with the known laws of physics'. Or they might say: 'it is physically possible for you to fly to the Sun in 8 hours (namely, by going at one sixtieth of the speed of light), but it is not practically possible, i.e. it is not compatible with present technology---supplies of rocket fuel, funding etc.'

Such examples prompt the idea of abstracting from the laws of physics, or another science, that we happen to know (or at least: that we believe we know). So for the example of flying to the Sun, the idea is to go beyond what I dubbed 'known laws of physics'. After all, 'known', as I used it above, is a weasel-word. Agreed, our confidence that a person cannot travel faster than light is a central claim of an extraordinarily well-confirmed theory (Einstein's relativity theory) which we can hardly imagine being overturned by future physics. Nevertheless: strictly speaking, 'known' implies 'being true'. And we must accept that all laws as at present formulated, even the laws of relativity theory, are fallible.

Thus such examples suggest that we have a notion of a law of nature. That is, roughly speaking: the notion of a proposition that: (i) is perhaps not formulated by us---and might never be formulated by humans---but that: (ii) is true about the cosmos (the actual one!), and is deeply informative about the way the cosmos "works". This last phrase is intended to set aside the countless true propositions we never have and never will formulate, that are dull, maybe arcane, matters of happenstance: as it might be, that all the children living on my street have prime-number birthdays.

Philosophers differ about how to make precise the phrase, 'deeply informative about the way the cosmos works'. (One suggestion, which I myself like, is by David Lewis, building on ideas of John Stuart Mill and Frank Ramsey.) But we do not need the details. We just need the idea that the laws of nature are an elite minority of the countlessly many true propositions about the cosmos, and that the conjunction, L say, of all these laws is thus an elite proposition that is deeply informative about the way the cosmos works.

The Humean tradition (cf. Chapters 1.4, 2.5) then suggests: although this conjunction L is true, it is contingent, i.e. not necessarily true. For example, consider the classical theory of electricity and magnetism, formulated by Maxwell in the late nineteenth century. This theory, embodied in Maxwell's famous equations, is extraordinarily successful. But it is in fact *not* true: for we live in a quantum world. But this theory *could* have been true. That is: there are logically possible worlds that are exactly and accurately described by the theory.

A note for afficionados: To make this more precise and more convincing, let me keep matters simple by imagining that there is no massive or charged matter---for agreed, matter is quantum. So I imagine just some configuration of electric and magnetic fields, propagating across spacetime, e.g. the spacetime of special relativity (called 'Minkowski spacetime'), obeying Maxwell's equations. That is indeed logically possible: physicists call it 'a solution of Maxwell's equations (in vacuum)'.

Thus with the set W containing all the logically possible worlds, we conclude with Hume that the conjunction L of all the actually-true laws of nature is contingent. That is: the set of worlds where L is true is a subset of W. Then 'physical possibility' corresponds to being true in some or other world that is in this subset of worlds.

Now we can easily make sense of our opening example. It was the sentence: 'it is logically possible for you to fly to the Sun in less than 8 minutes, but not physically possible, i.e. not compatible with the known laws of physics'. We assume that 'No object can move faster than light' is indeed a contingent law of nature, i.e. a conjunct in the long conjunction L. Then 'it is logically possible, but not physically possible, for you to fly to the Sun in less than 8 minutes' is indeed true: there is a logically possible world---but not a world making L true---in which you fly to the Sun in less than 8 minutes.

(Some jargon. Nomos is the Greek word for 'law'. So the restriction of modality to what conforms to the laws of nature is sometimes called nomic modality (also: nomological modality).

Chapter 3.7: Paradise III: Counterfactual conditionals

My next example of the philosophers' paradise is counterfactual conditionals: that is, propositions of the form, 'If P were so, then Q would be so'. Here the phrase 'were so' signals that the antecedent P is actually false ('contrary to fact': hence the name). In terms of possible worlds, P is false at the actual world.

The discussion will have two stages. The first is uncontroversial: I will report these propositions' curious logical behaviour, which was noticed by philosophers and logicians in the 1950s and 1960s. The second is more controversial: I will report the proposal, made by Lewis and Stalnaker ca. 1968 (independently of each other), that we should understand this behaviour in terms of degrees of similarity between possible worlds. This will amount to a generalization of the last Section's idea of restricted modality. For the proposal will be that to formulate what makes a counterfactual true---in the jargon: to give the truth-condition of a counterfactual---we must invoke, not a single subset of the set W of all worlds, but a collection of such subsets, defined in terms of similarity.

So first: the curious logical behaviour. A conditional connective, 'if…, then…', is a sentence operator. Like 'and', it accepts two sentences as "inputs", and "ouputs" a third sentence. The intuitive idea of a conditional, of 'if…, then…', suggests several logical principles which one expects the connective to obey.

One example is transitivity. This is the principle that, writing the connective as ->, the following inference is valid, for any propositions P, Q and R: 'P -> Q; Q -> R. So, P -> R'. One naturally says: surely, any 'if, then' will obey transitivity. For it accords with the idea that the truth of a conditional goes along with an argument being valid, or anyway in some sense good or plausible; and such arguments can be concatenated to give valid or good arguments.

Another example is strengthening the antecedent. This is the principle that given a true conditional, adding a conjunct to the antecedent (making it logically stronger) yields another true conditional. That is: one expects the following inference is valid, for any propositions P, Q and R: ' P -> Q. So, (P and R) -> Q'.

But many examples show that the counterfactual conditional violates these principles; (and several others that are, at first sight, equally plausible for any conditional connective).

Here is one example showing that counterfactuals do not obey transitivity: an example from the Cold War in 1950s USA. The first two statements below are true. Or at least we can take them to be, in some conversational context that determines what possibilities are relevant or likely. But the third is false; or at least we can take it to be.

1) If J. Edgar Hoover were Russian, he would be a Communist. (The idea here is: Hoover's ambitious but conformist temperament is retained under the supposition that he grows up in Russia.)

2) If J. Edgar Hoover were a Communist, he would be a traitor. (The idea here is: under the supposition that Hoover is a Communist, we still imagine him as an American citizen living in the USA, indeed perhaps as head of the FBI.)
3) If J. Edgar Hoover were Russian, he would be a traitor. (The reason this is false, or at we can take it to be, is exactly as in 1): Under the supposition that Hoover grows up in Russia, his ambitious but conformist temperament is retained, and so there is no reason to think he is a traitor to the Communist one-party state.)

And here is an example showing that counterfactuals do not obey 'strengthening the antecedent'. The first statement below is true, the second false. (Or again, at least we can take them to be true and false respectively, in some conversational context.)

4) If I were to strike this match on the side of the matchbox, it would ignite.
5) If I were to strike this match on the side of the matchbox and the matchbox was wet, it would ignite.

How should we explain such strange logical behaviour? There is a natural proposal, due to Lewis and Stalnaker, for what 'If P were so, then Q would be so' means; and this proposal explains the logical behavior. Namely: Lewis and Stalnaker propose 'If P were so, then Q would be so' means 'In the world or worlds that are most similar to the actual world while making P true, it is also true that Q'. (As I mentioned; this specification of meaning in terms of what would make a proposition true is called a 'truth-condition'.)

Lewis and Stalnaker differ about the details of this proposal. The main difference is that Stalnaker proposes that for any world w (in particular, the actual world) and any proposition P that is not true at w, there is a unique world that is most similar to w while making P true; whereas Lewis proposes, more cautiously, that relative to any world w (in particular, the actual world), other worlds are ordered by their similarity to w, but in this ordering, two worlds might well be equally similar to w. This makes the proposal readily visualizable, if we think of worlds as dots on the page that are closer together, the more similar the worlds are. Thus Lewis envisages that around any world w, we can draw a sequence of concentric circles that have (the dot representing) w as their common centre. As we go out from w, we successively include worlds that are more and more dissimilar to w.

But these differences of detail do not affect the main point: that if we accept the proposed truth-condition for 'If P were so, then Q would be so', then the strange logical behaviour is readily explained. For different antecedents, i.e. different counterfactual suppositions P, will ``carry us" to different worlds making P true, at which we then ask whether Q is true. So suppositions that are more "outlandish" will carry us to worlds more dissimilar to the actual world, that are represented by dots in a bigger circle. And these explanations are readily visualized.

Here is how it goes for our two examples. (There are analogous, and equally visualizable, explanations of various other curious logical behaviours.) The failure of transitivity in the Hoover example is due to it being more dissimilar to the actual world to imagine Hoover growing up in Russia, than his being a Communist within the USA.

ADD A DIAGRAM HERE

In an analogous way: counterfactuals do not obey 'strengthening the antecedent' because strengthening the antecedent, from 'P' to 'P and R', can make the antecedent carry us to worlds more dissimilar to the actual world than does P (i.e. more dissimilar to the actual world than the most similar P-world(s)). Indeed, in everyday life we make sure that matchboxes stay dry, and so the antecedent of 5) above is more "outlandish" than the antecedent of 4).

? ADD ANOTHER DIAGRAM HERE

<u>Chapter 3.8: Paradise IV: Supervenience, physicalism and determinism</u>

My last example of 'the philosophers' paradise', i.e. the uses of possible worlds, is the notion of <u>supervenience</u> (also known as 'determination'). This notion is important in many philosophical discussions. But for this book's purposes, I need to describe how it is useful for formulating (and so also assessing) just two ideas: physicalism and determinism. As we will see, for both supervenience in general, and for its application to these two ideas, one can be either confident or cautious in the sense of Chapter 1.3.

Supervenience is a relation describing how one set of properties and relations determines another. In philosophy, 'attribute' is sometimes used as an umbrella term for 'properties and relations'; but I shall just say 'properties', for short. All the properties in each set are about some single subject-matter or topic. Because of this kinship between the properties, it is common to call the sets 'families'. So philosophers talk of supervenience as a relation between families of properties (or attributes). Thus the objects in the subject-matter can be described by which properties they have; and the set of properties amounts to a taxonomy or classification-scheme for the objects.

For example: think of botany as a subject-matter or topic. It classifies objects as plants or non-plants, and then further classifies the plants as daffodils, roses etc. So botany can be presented as the family of properties: being a plant, being a daffodil, being a rose etc. And the set of all the botanical facts just is the classification of all appropriate objects (especially the plants) with respect to this family: the assignment of each object to its botanical pigeonhole.

Similarly for other subject-matters: including larger, more encompassing ones such as, for our example, biology. We can think of the set of all the biological facts as the classification of all appropriate objects (all organisms) in terms of all the many biological properties.

So supervenience is to be a relation between sets of properties. Or in alternative jargons: a relation between subject-matters or taxonomies or classification-schemes. What relation? The answer is: the classification of any of the objects using one set of properties implies how it is classified by the other set. This is as supervenience's synonym, 'determination', suggests: the classification of an object using one set of properties determines ('fixes': in the sense of 'makes rigid', not 'repairs') its classification by the other set.

By taking facts as given by what properties objects have, we can also put this in terms of facts. Thus supervenience is: all the facts about one family of properties, one subject-matter, F1 say, are fixed by all the facts about another family F2. In other words: specifying all the facts about F2 involves, *ipso facto*, specifying all the facts about F1. We say: F1 supervenes on F2. We also say: F2 subvenes F1.

Here is an example that is standard, since it is uncontroversial. At least, it is uncontroversial by the standards of philosophy. The objects in question are pictures. Then it is very plausible that the aesthetic properties of pictures---the classification of them as being beautiful, being well-composed, having a dark palette etc.---supervene on their pictorial properties, i.e. the properties about how exactly the paint, and of what kind, is distributed on the canvas or paper: being an oil painting, having magenta in the top left square centimetre etc. So the idea is: any two pictures that match in all (all, not just some) of their pictorial properties must also match in all their aesthetic properties (again: all, not just some). That is: two pictures that are replicas of each other as regards pictorial properties, must also be replicas as regards aesthetic properties. So both are beautiful, or both are ugly; and both are well-composed, or both are badly composed; and so on. The pictorial properties subvene or determine the aesthetic properties. To put it the other way around: pictures cannot differ from one another as regards some aesthetic property, without also differing as regards some pictorial property. In a slogan: no aesthetic difference without a pictorial difference. This is what it means to say that aesthetic properties supervene on pictorial properties.

There are various major topics in philosophy where the question, whether a certain family of properties or subject-matter supervenes on a certain other one, is central. One is the relation between mind and matter (sometimes called the 'mind-body relation'). Do mental properties of sentient animals (like seeing yellow in the top-left of the visual field, or feeling hungry, or hoping for a sunny day) supervene on their natural scientific properties, i.e. their panoply of physical, chemical and biological properties? That is: If two animals matched as regards all their physical, chemical and biological properties, must they also match as regards seeing yellow in the top-left of the visual field, and also as regards feeling hungry?

Saying 'Yes' to this question is often called materialism. The idea is that all the facts about matter, as explored by the sciences of physics, chemistry and biology, fix all the facts about an animal, even the facts about its mental life. Nowadays, this is very widely endorsed. But agreed: in the nineteenth century, it was reasonable to deny it. It is only with the cumulative successes of physiology, molecular biology and neuroscience in describing mental states that the idea of special mental causes (of at least some such states) has died away.

Analogous comments apply to the dependence, nowadays evident, of biology on chemistry, and of chemistry on physics. That is: in the nineteenth century, it was reasonable to believe in what were called 'vital forces': causal factors occurring only in living organisms that "rode free" from their underlying chemical and physical descriptions. But the successes of physiology, e.g. its explanations in physico-chemical terms of the nerve impulse, muscle-contraction and vision, put paid to vital forces. And until 1930 or even later, it was reasonable to believe that chemical phenomena, in particular chemical bonding, would not be explicable by the physics of atoms. But since 1930, quantum theory has achieved ever more precise descriptions and explanations of chemical phenomena: in a way that was impossible---indeed, provably impossible---according to the earlier classical physics.

Thus arises the doctrine of physicalism. This is a strengthening of materialism, that gives physics a pre-eminent role, as compared with the other sciences. So as a claim of supervenience, physicaliam amounts to: all the facts described by chemistry, biology and the other sciences, in particular psychology, are determined (fixed) by the panoply of all the physical facts.

Obviously, this sketch of materialism and physicalism as claims of supervenience leaves a lot to be made precise. What exactly are the sets of objects being described by the two (or more) subject-matters or taxonomies, of which one is said to supervene on the other? And what exactly are the sets of properties (and relations) defining the subject-matters, or taxonomies. For example, what exactly is the set of physical properties? As you would expect, different philosophers give different answers to these questions: influenced, usually, by different judgments about which precise concepts of e.g. 'physics' or 'all sentient animals' make for a supervenience thesis that is---not obviously true, but---debatable enough to be worth assessing for truth. And in debating such formulations, there are choices about whether to be confident or cautious in the sense of Chapter 1.3. For example: would you be confident or cautious about a firm distinction between physical properties and other ones?

But we do not need to go in to the details of these debates. Here, I only want to describe one main way that possible worlds help us be precise in formulating such supervenience claims. This concerns what one might call the 'modal range', or 'modal extent' of the claim.

For consider the actual world, i.e. the actual cosmos spread throughout all space and all time, past present and future. (We adopted this use of 'actual world' at the start of Chapter 1.) And suppose we take a supervenience claim as being about actual objects. So we say that two actual objects matching exactly as regards all the properties in a set (family) F2 also match as regards all the properties in another family F1.

Then there is likely to be trouble. For the families of properties F2 we are concerned with are bound to be rich, i.e. to make fine distinctions. Recall our examples: all pictorial properties; all material properties as described by physics, chemistry and biology; all physical properties. In all these examples, the idea is that the family F2 has to be rich, in order to have a chance of fixing all of F1. And there is the rub. For any reasonably rich taxonomy (family of properties), the actual objects are very likely to be a very varied set. Agreed: two actual objects often match for some property; both have it or both lack it. But it may well be that no two actual objects match for <u>every</u> property in F2. And if so, the supervenience claim restricted to actual objects---'if they match in this way, then they also match for all of F1'---loses its force, or content. For the antecedent 'they match in this way' is never true. (Philosophers and logicians call this 'vacuous truth'.)

The answer lies in recognizing that the basic idea of supervenience is modally involved. It is not an assertion about a case of actual matching for all of F2: since for most of philosophy's interesting supervenience claims, there are no such cases. Rather, it is about <u>trans-world matching</u>. Thus we again see the theme of Section 3: that throughout our thought and language, both everyday and scientific, we are up to our necks in modality.

To take an example, physicalism says: if there were a replica of this actual cat that is now seeing yellow in the top-left of its visual field, and this replica was 'physically perfect' i.e. utterly matched all the actual cat's physical properties, then the replica would also see yellow in the top-left visual field.

Besides, the usefulness of possible worlds for formulating supervenience claims is not limited to providing possible objects: e.g. cats that are atom-for-atom replicas of some actual cat. There is also the question of whether the supervenience claim being considered, e.g. materialism or physicalism, is propounded as contingent or as necessary. And if it is propounded as contingent, that means in a possible worlds framework: true in some possible worlds but not all. And this prompts the further question: across exactly what set of worlds is supervenience claimed? For example, for physicalism: across exactly what set of worlds must an atom-for-atom replica of some actual cat utterly match all the actual cat's mental properties?

Again, we do not need to take a view about the answer. What matters for us are two points, both of which echo some previous themes. First: most philosophers do indeed formulate materialism and physicalism as logically contingent claims, not necessary ones. This is of course because the success, since 1800, of the natural sciences, and especially of physics, in describing and explaining phenomena lying outside their original scope---as illustrated above---was undoubtedly contingent. It did not have to be so. We might have discovered vital forces underpinning metabolic processes, or phototropism in plants, or what-not. And we might have discovered distinctive chemical forces that explained bonding, chemical valences etc., independently of the electron orbitals around atoms' nuclei. This contingency---this happenstance of a "one-way street" for 200 years, from the other sciences towards chemistry and then on to physics---makes it very natural to formulate materialism and physicalism as contingent claims.

Second, answering the question 'Across exactly what set of worlds is supervenience claimed?' leads us back to the idea of a law of nature. For one natural answer is: 'the set of worlds that share with the actual world their laws of nature---the nomically possible worlds'. (This answer is natural, but by no means compulsory. For as we discussed in Chapter 2 and Section 3.6: one might well be cautious, rather than confident, about the very idea of a law of nature.) For example, a physicalist might say: 'I accept the idea of a law of nature, and believe they are all contingent. And I claim physical matching implies total matching, across the set of nomically possible worlds.'

As a final illustration of the power of possible worlds, I turn to determinism. I briefly discussed this at the end of Section 3 above. We saw that all physical theories postulate a space of instantaneous states of the system they describe, so that a possible history of the system (i.e. life-history, comprising both past and future) is represented by a curve through the state-space. Thus I reported the idea of determinism, as follows. A physical theory is deterministic if the state of the system at one time (together with a specification of the influences it is subject to in the past and future) determines its state at all past and future times. In terms of histories as curves through the state-space: through any point in the state-space, there is a unique curve to the future and the past.

It is now clear that, like physicalism, determinism is a supervenience claim---as my word just now, 'determines' signals. For in view of the word 'determine', the definition just given means: any two systems (of the sort that the theory describes) that match exactly in their states at one time, match exactly in their states at all past and future times. This is clearly a statement of supervenience. Namely: the past and future (of the system concerned) supervene on the present---fixing the latter implicitly fixes the former.

Besides, we here see again the contrast, being confident or being cautious, about a concept; (cf. Chapter 1.4). What I just said gave only a cautious construal of determinism as a property of a given theory that applies to a given type of system; and the possible worlds involved were modest ones, viz. instantaneous states. But one might be more confident. Thus suppose we accept the idea, not just of the laws of a given theory, but of a law of nature. Recall the discussion in Chapter 1.4 and Section 6 above.

Then we can think of the conjunction of the laws of nature at a given possible world w as 'the theory of w'. (We might call it 'the theory of everything at w'. But nowadays, the phrase 'theory of everything' is always used for an ambitious and more specific idea, that is like physicalism, as defined above. Namely: the facts described by a single theory of physics might determine all the facts of all the sciences. But 'the theory of w', as just defined, might well not be a physical theory.)

Given this notion of the theory of a given possible world, we can define what is for an entire possible world to be deterministic. It is for the theory of that world to be a deterministic theory, in the previous sense. But here, the space of possibilities will be the "ambitious" space of all possible worlds: not a "modest" state-space of a single theory, such as Newtonian mechanics. So we say that a world w is deterministic if: for any possible world that also makes true the theory of w (i.e. all the laws of nature at w), and whose state at some time matches exactly the state of w at some time---the two worlds match exactly at all times (to both past and future of the assumed matching).

Let me sum up this last discussion. If we confidently accept the idea of the theory of a possible world, then determinism of a world is, again, supervenience. Namely, supervenience of all the past and future states, of an entire possible world, on its present state.

Chapter 3.9: Existential *angst*: what are possible worlds?

So much by way sketching the philosophical benefits of using possible worlds. So much by way of tasting the fruits in the philosophers' paradise. For the rest of this Chapter, I turn to the question which I announced in the preamble to this Chapter: what exactly is a possible world?

As I said there: this question is compulsory, for cautious conceptions of possible worlds as much as confident conceptions; and several possible answers are defended in the philosophical literature. It is also agreed to be a very hard question. Though we can readily agree that our thought and language, everyday and scientific, continually invokes non-actual possibilities (cf. Section 3 above), what exactly they are is an open, and stubbornly difficult, question. Hence this Section's title says: *angst*. Besides, focusing on this question accords with Chapter 1's announcement that the discussion of each of the three multiverse proposals will end by urging an open philosophical problem that the proposal prompts.

So unsurprisingly (and as I admitted in this Chapter's preamble): I cannot honestly urge one answer as correct. I will instead address the question by, first, refuting two tempting suggestions. They are tempting, but definitely false. And they are suggested, I am sorry to say, by proposals from renowned philosophers: Berkeley and Wittgenstein. Then I will discuss a third suggestion that fares better; but is still, I fear, wrong. The upshot (in the next Section) will be that the Chapter ends where it began: by stressing that Lewis' modal realism is a coherent intellectual possibility, even though probably, you---like I---find it incredible.

Chapter 3.9: A: Acts of Imagination?

One natural suggestion is that a non-actual possibility is something we imagine. But here, we have to be careful to distinguish the actual event or state of affairs---a person, say you, imagining that P---from the proposition P being imagined to be true. The distinction applies not just to imagination, but to many mental acts, such as hope, belief, desire, regret. Thus we say that 'John imagines/hopes/believes/desires/regrets that P'. Philosophers' jargon for this is that to imagine, to hope, to believe etc. are propositional attitudes; and that the proposition P "on which the mind is focussed" is the content of the attitude; (i.e. the content of the event or state of affairs of John imagining etc.).

This distinction is clear enough. But it makes trouble for the 'something-we-imagine' suggestion. There is a dilemma: the first horn makes no progress, and the second is definitely wrong.

Suppose, first, that the suggestion is: the non-actual possibility is the content or proposition. That may well be right, but then the question becomes: what exactly is a proposition? I touched on this at the start of Section 3, when I remarked that the content of any false belief such as 'I go to the cinema tonight' (assuming I in fact stay home) represents a non-actual possibility. Nor does the framework of intensional semantics (reviewed in Section 5) help answer the question. For we saw that, although it systematically portrays how propositions, Fregean senses of words and phrases, truth-values ('true' and 'false') and possible worlds all relate to each other, and get expressed by language---it gives no opinion about what a possible world, or more generally a non-actual possibility or proposition, actually is. So we are no further ahead.

(In Section 5, possible worlds and truth-values were basic posits, not further analysed. Thus at the end of that Section, a proposition was taken as a mathematical function from possible worlds to the set of two truth-values. Agreed: one might instead take propositions, or Fregean senses of sub-sentential words and phrases, as basic, and build possible worlds from them, using the language of functions; or more generally, using set-theory. For example, a possible world might be taken as a maximally logically strong ('maximally opinionated') proposition. But the question, of the nature of the basic posits, would remain.)

Suppose, on the other hand, that the suggestion is: the non-actual possibility is the event or state of affairs of imagining. That certainly makes the non-actual possibility unproblematic, and "down to earth" as a short-lived episode (of a mind or brain) within the actual world. But it is definitely wrong. For obviously, there are countless non-actual possibilities that nobody ever actually imagines.

Besides, this suggestion implies that any non-actual possibility has as a necessary concomitant, as an implication, the existence and imaginative activity of a mind. Which is not so. Here we return to the clarifying comment (2) at the end of Section 5 above. There, I stressed that possible worlds provide a semantics for our language as we actually speak it; and that (setting aside some contrived sentences about how we might have spoken), how people in other worlds---if there are any---speak is in general irrelevant to the semantics of our sentences. This also means that a possible world with no people, indeed no animals, or other sentient beings, is entirely coherent. Such a world can make true a proposition of our language, such as 'the world consists entirely of five boulders of granite floating in a Newtonian space, without any living or conscious being'. No sentience—in particular, no visualisation of the boulders---is needed within the world. In short: the idea of possibility as such does not imply the existence and imaginative activity of a mind.

Incidentally, here we also see the flaw in the claims by the eighteenth-century idealist philosopher Berkeley: that (i) we cannot imagine an unperceived object, and therefore (allegedly!) that (ii) any object must be perceived. For there is an equivocation. If (i), i.e. 'we cannot imagine an unperceived object', means 'we cannot imagine a possible world that contains no sentient beings', then (i) is false. (Just think of the boulder world.) But I agree: so understood, (i) implies (ii): if (i) were true, (ii) would also be true. If on the other hand, (i) means 'we cannot imagine ourselves within a possible world that contains no sentient beings', then (i) is of course true. Indeed, it is necessary if 'ourself' implies being sentient. But it by no means implies (ii). (Again: just think of the boulder world.)

This ends my rebuttal of appealing to imagination as the way to understand possibility. I turn to my second tempting, but wrong, suggestion.

<u>Chapter 3.9: B: Combinations?</u>

Here the "culprit" will be Wittgenstein, in his early work, the <u>Tractatus Logico-Philosophicus</u>; (which he later disavowed, partly for the reasons I will present). Indeed, we will see that he is "more guilty" than Berkeley, since he does not just make claims that prompt the suggestion: he explicitly makes the suggestion.

The idea of the suggestion is, at the start, modest. It proposes that we should lower our sights about understanding what a possibility, or possible world, or proposition, really is; and assuming we accept these notions, we should focus instead on the following question---which, admittedly, is vague: How can a proposition be <u>necessary</u>? What explains that?

This echoes the questions we pursued at the end of Chapter 2 and the start of this Chapter (Sections 2.8 and 3.1). Namely: what is pure mathematics really about? what is its subject-matter? and can it be reduced, as logicism claimed, to logic?

It is in the context of those questions that Wittgenstein, in the <u>Tractatus</u>, suggested that for any necessary proposition, whatever its subject-matter, its necessity is like that of what propositional logic calls <u>tautologies</u>. These are special sentences whose necessity can be agreed by all parties to be utterly unproblematic. For they are defined as those sentences built from others using 'and', 'not' and 'or' (which, as discussed in Section 5 above, are <u>truth-functions</u>), with the feature that whatever the truth-values of the component sentences, the compound sentence must be true---just because of the order in which the truth-functions, 'and', 'not' and 'or', have been applied to the components. Examples starting with one component sentence include: 'P or not-P', and 'not-(P and not-P)'. An example starting with two components, P and Q, is: '(P or Q) or (not-P) or (not-Q)'. (Here 'or' is understood, as usual, in our inclusive and-or sense.)

A simple diagram, called a truth-table, makes the idea clear. We give each component sentence a column, and underneath we assign a row to each combination of truth-values that could occur, and then apply the truth-functions to calculate that the entire sentence comes out true in every row. Thus to calculate the truth-table for '(P or Q) or (not-P) or (not-Q)', we need four rows: one for 'P true and Q true', one for 'P true and Q false', one for 'Q true and P false' and one for 'P false and Q false'. Then, by applying the truth-functions 'not' and 'or' appropriately, we calculate that '(P or Q) or (not-P) or (not-Q)' comes out 'true' in every row.

So we think of each row, each combination of truth-values for the component sentences, as a 'way the world could be', as described by those sentences: as a toy-model of a possible world. Then calculating that the truth-value must be 'true' in every row explains why the whole sentence is necessary, in a completely unproblematic way.

The crucial word here is 'combination', as in 'combination of truth-values'. No necessity, nor any other unexplained modal status or mutual logical relation, is attributed to the component sentences. They can be true or false, quite independently of each other: all combination of truth-values are genuinely possible. This is called being logically independent. So the idea is: whatever their combination of truth-values, the placing of 'and', 'not' and 'or' in the whole sentence forces it to be true: i.e. true in that row, that combination. Thus Wittgenstein proposed that all necessity had this lucid combinatorial origin and explanation. He declared that all necessary propositions are really tautologies.

But this is entirely programmatic: a mere declaration. Indeed, there are three large groups of necessary propositions, whose necessity seems to have little if anything to do with the placing of 'and', 'not' and 'or' in any sentences. First: the truths of pure mathematics, like '2+2=4', 'there are infinitely many prime numbers', 'equilateral triangles are equiangular', seem to be necessary because of their special non-empirical subject-matter, such as numbers and geometric figures. Second: there are propositions (with any subject-matter) whose necessity turns upon the placing of those other logical words, 'all', 'any', 'some' and 'none', that are studied in predicate logic. For example, 'If everything is both A and B, then something is A'. Third: there are propositions whose necessity turns upon, not the placing of logical words, but relations between the meanings of other words, such as 'All bachelors are unmarried', and 'A vixen is a female fox'.

To show that all these necessary propositions are really tautologies, one would have to show, somehow or other, that they were built up from components that are logically independent, i.e. for which every combination of their truth-values is genuinely possible. That would be a programme of reduction, in the sense of Section 1 above. In the Tractatus, Wittgenstein was of course influenced by the logicism of Frege and Russell, and so made some suggestions about how to cope with the first and second groups above. For example, maybe 'all' could be reduced to the idea of conjunction, though a possibly infinite one; (and similarly 'some' to a possibly infinite disjunction). But there were few details, especially about the third group; (the problem is just glimpsed at assertion 6.3751 of the Tractatus). Indeed, this shortcoming was one of the main reasons why Wittgenstein, a few years later, abandoned its claims.

Furthermore, no one else has succeeded in this combinatorial approach to explaining necessity. In particular, I note that its prospects do not improve if we adopt a cautious conception of possible worlds, suggested by the state-spaces of physical theories. There are two problems.

First: because the different values of a quantity---whether a physical quantities like position of a point-particle, or a psychological quantity like 'magenta in the top-left of the visual field'---exclude one another, the propositions ascribing such values cannot be logically independent. One might try to address this by interpreting the values as a new kind of truth-value. So for example: a point-particle having position in the x-direction equal to 5 metres is to be understood as the "proposition", 'has position in the x-direction', having as its "truth-value", 5 metres.

But even if one swallows this, there is a second problem. Namely: in the state-spaces of most physical theories, not every combination of values of the physical quantities defining the space is possible. (In the jargon of set-theory: the space is not a Cartesian product of the sets of values of those quantities. For two quantities, such a Cartesian product is naturally pictured as a rectangular array of points, with each point being a pair of values for the two quantities; for three quantities, one pictures a brick-shaped 3-dimensional array of triples of values; and so on.) Nor is there any reason to think that in a putative physical 'theory of everything' (mentioned at the end of Section 8), the state-space would be a Cartesian product.

Chapter 3.9: C: Sentences and sets?

I turn to my third suggestion about what a possible world is. As I announced, it fares better than the first two. But I fear (following Lewis' critique of it) that it too is wrong.

The idea is that a possible world is like a novel: that is, the sentences, rather than the propositions they express. (Saying 'the propositions expressed' would get us no further ahead, as we discussed under the first suggestion.) At first sight, the advantage of this suggestion is that a sentence is an unproblematic object to believe in. For it can surely be taken as the set of all its physical inscriptions in pencil, ink etc., and all the events of its being spoken.

But being more precise brings difficulties. Surely most possible worlds that our thought and talk invokes (cf. Section 3) cannot be represented by a finite sentence, even of a richly expressive language like English. And surely, infinitely long sentences do not exist, i.e. exist in the actual world.

These difficulties prompt one to generalize the idea from sentences to sets of actual objects, and also actual properties and relations. For set theory provides countlessly many sets, many with very intricate structures; (since the operation of making a set out of some given sets---'putting a curly bracket around them'---can be iterated endlessly). Here we return to the discussions in Chapter 2.8 and Section 1 above, about set theory as a lingua franca for expressing all of pure mathematics. The idea now is, in effect, that it is a lingua franca for expressing anything.

Thus the suggestion is that (i) possible worlds are sets of a certain kind built from actual constituents, i.e. actual objects properties and relations; and (ii) the structure of such a set, i.e. the pattern of curly brackets by which it is built up, encodes how it represents a possibility, in a manner similar to that in which the grammatical structure of a sentence encodes how it represents. (Recall Section 5's idea of compositional semantics.) Or in other words: the idea is that the set's structure exactly mirrors the structure of the possible world: and this makes the set the preferred official proposal for being the possible world.

Since nowadays most philosophers accept sets built from actual objects etc. as being themselves legitimate objects, this suggestion seems promising. It reassures us that although the actual world, the cosmos around us, is, presumably, not a set---it is concrete ("material"), not abstract ("immaterial")---any non-actual possible world is a set. Non-actual possible worlds are thereby abstract, in vivid contrast with the concrete cosmos around us. But since they are sets, we can and should accept them as objects: for they are no less acceptable than any other sets.

But I fear that this suggestion does not work. There are several objections; but I shall present just one. (It is urged by Lewis himself, along with others; cf. especially Section 3.2, p. 150 f. of his On the Plurality of Worlds.)

Namely: the suggestion assumes the notion of possibility, it does not analyze or explain it. For if a possible world is a set of sentences, then they must be consistent with each other, i.e.

possibly all true; (equivalently, a possible world taken as a long conjunction must be possibly true). But there is no unproblematic, in particular syntactic, test for consistency. For inconsistency is not just a matter of containing 'P' and 'not-P'. That is only one, very simple, way to be inconsistent. Relations between the meanings of non-logical words provide many other examples. Think of 'Fred is a married bachelor', or 'A male vixen got into the chicken-hutch'. And there is no reason to think, as the early Wittgenstein did, that we can analyze or reduce all our language (e.g. to a set of logically independent propositions), so as to devise a syntactic test for consistency.

Nor does it help to move from sentences to sets. Just as there are sets of sentences, or conjunctions, that are inconsistent without "wearing it on their sleeve", i.e. without a syntactic sign of it: there are countless sets that, once we endeavour to interpret each of them as representing a possibility, in fact represent an impossibility---without the structure of the set encoding any sign of it. So again, the suggestion assumes, but does not explain, the notion of possibility.

Here is a simple example. Consider: 'Butterfield is in Rome in February 2022.' That is false, but possibly true. Following the suggestion, we are to apply set theory to actual objects and properties so as to build, with appropriate representational conventions, a set-theoretic "mock-up" or "replica" of this possibility.

Let us adopt very simple representational conventions, as follows. (They probably work smoothly only for very simple examples, but that will not matter.) We take a period of time, such as February 2022, to be a spacetime region: for example, the Earth during that month. (I set aside the need for further conventions about where and when on Earth, the month begins and ends.) In terms of semantics (Section 5), the referent of 'February 2022' is the spacetime region. And let us for simplicity take a city, such as Rome, during a period of time to be the set of all its contents during that period, or any part of the period. So there is a set we can label by the description, 'Rome-in-February-2022'. This set in the actual world contains e.g. Pope Francis, and the actual Italian Prime Minister; but not Butterfield.

But with our representational conventions, we can still represent very simply the possibility that Butterfield is in Rome in February 2022. For recalling how intensional semantics gives descriptions like 'the capital of Denmark' different referents at different worlds (cf. Section 5), we see that, according to the suggestion: this possibility just is a set-theoretic fact. Namely: it is the fact that Butterfield is a member of (the set that is) the referent of 'Rome-in-February-2022', at various worlds. (Amongst these worlds, those most similar to the actual world, according to our prevailing criteria of similarity, will no doubt "retain" most of Rome's actual contents during February 2022, e.g. Pope Francis. "There is room in town for both of us".)

So far, so good. So far, the suggestion that a possibility is a set has held up. For I have exhibited a set that is appropriately structured to be the possibility that Butterfield is in Rome in February 2022. But the problem for possibilities as sets is parallel to that for possibilities as sentences (or sets of them). Given our representational conventions (about periods of times, about cities as sets of their contents etc.), we are equally committed to countless sets that represent an impossibility, with no sign of why they do---and with no hope of evading the problem, by some change of representational conventions.

For example, I take it to be impossible that I am a fried egg. So 'Butterfield is a fried egg' is necessarily false. Yet there are countless sets that, in an exactly parallel manner to the previous example, put me in the extension (set of instances) of the predicate 'is a fried egg'.

Here I admit: if we assume we have in place a framework of intensional semantics that respects the meanings of our words, so that all assignments of extensions to predicates at the various worlds are genuinely possible, and are not ruled out like a married bachelor, male vixen or human fried egg, then all will be well. That is: ex hypothesi, the sets mentioned by our semantics as representing possibilities, e.g. properties that a man or a fox could have, will succeed in doing so. They will not "lead us astray" by making impossibilities appear possible, masquerading in an appropriately structured nest of curly brackets.

But of course the problem remains, as it did for sentences, rather than sets. Namely: assuming such a framework of meaning-respecting intensional semantics means assuming, not explaining, the notion of possibility.

We can sum up this critique of possible worlds (or possibilities) being sentences or sets, as follows. Saying that the sentence 'Butterfield could be in Rome in February 2022' is made true by the existence of a certain set looks plain wrong. For by parity of reasoning, one would have to also say that 'Butterfield could be a fried egg' is made true by the existence of an equally legitimate set.

Chapter 3.10: Lewis' modal realism

So I end with what began this Chapter: Lewis' modal realism. Lewis believes that:

(i) all the possible worlds are equally real;
(ii) the actual world is in no way special, except from our standpoint within it; and
(iii) although we use 'actual' (and 'real' and similar words) restrictedly, for the actual world, 'actual' is like the word 'here': it is what philosophers call an indexical, i.e. it is a word whose referent depends on the context of utterance---but for 'actual' the relevant aspect of context (with respect to which one asks for the referent) is the world, not the spatial place.

So this is the philosophical multiverse, *par excellence.*

As I said in the preamble to this Chapter: Lewis does not claim to have an irrefutable argument in favour of his view. His extended defence of it (especially in On the Plurality of Worlds) claims only to show that on balance, it is more credible than rival views. He gives several of these rival views a good run for their money. (This includes the last Section's 'sentences and sets' suggestion, which is roughly equivalent to what he calls 'linguistic ersatzism'.) His defence also includes much else. For example:

(a) he specifies in what senses possible worlds are concrete (on his view);

(b) he replies to objections to his view (several of which he himself thought of);

(c) he defends his view about what it is for an object to be in two worlds, as in the last Section's example of Butterfield and Rome; and

(d) he explains, using his persuasive account of causation, why there is no causation between worlds, i.e. why no event in any world is a cause of an event in another; so that his views satisfy our requirement, at the end of Chapter 1 (Chapter 1.6), that advocacy a multiverse should not be undermined by the bewildering idea that most of one's readers or hearers are in another universe.

But let me try to live up to Chapter 1's announced standards of being honest about what one can believe, and self-aware about one's intellectual temperament. I must admit that (like most philosophers) I simply cannot believe Lewis' view.

So for me, concerning the question what a possible world exactly is: the jury is still out. So this Chapter is inconclusive, and perhaps disappointing. But there is some consolation: the next two Chapters will not depend on my having endorsed an answer. And in the next Chapter, the Everettian interpretation of quantum theory will suggest another answer. It may not be persuasive; but it is certainly worth considering.

In any case, this is not the place to further expound or assess Lewis' views. For this book's purposes, it suffices to have shown in this Chapter that: (i) we are, in our thought and language, up to our neck in modality; (ii) logicians and philosophers have developed detailed frameworks for describing and analysing modal concepts; and (iii) nevertheless, the basic question, 'what exactly is a possibility, or a possible world?', still remains stubbornly difficult.

So I shall end with a glimpse of "Lewis in action". He was a very active philosophical correspondent; and in a letter of 15 June 1984 to the cartoonist Roz Chast, in which he asked to use her cartoon 'Parallel Universes' (cf. **Figure 1**) as a frontispiece of his book, he gave a vivid and witty summary of his modal realism. He wrote:

Dear Roz Chast,

I'm writing to explore the possibility of using your 'Parallel Universes' as a frontispiece in a forthcoming book of mine about possible worlds.

I have gained some notoriety among philosophers by claiming that this world we are part of is just one of many possible worlds; in no way is it special, except from the standpoint of us who inhabit it. It turns out that systematic philosophy goes more smoothly if we suppose that there are many worlds, and I take that to be a good reason why we should believe that there are. My views are highly controversial, to put it mildly; I think 'crazy' is how many would prefer to put it. For years, I've bene helping myself to the other worlds when I wrote about one or another philosophical problem. But I never wrote at length about what it mean to believe in them, and why we ought to. Now I have. I've written (well, almost finished writing) a book titled On the Plurality of Worlds. … It is written in prose, not math; but I fear that it still will be a book mostly for specialists, because it presupposes familiarity with a good deal of recent philosophical writing. I'd be glad to send you a copy of the manuscript if you like, but I didn't want to inflict it on you uninvited.

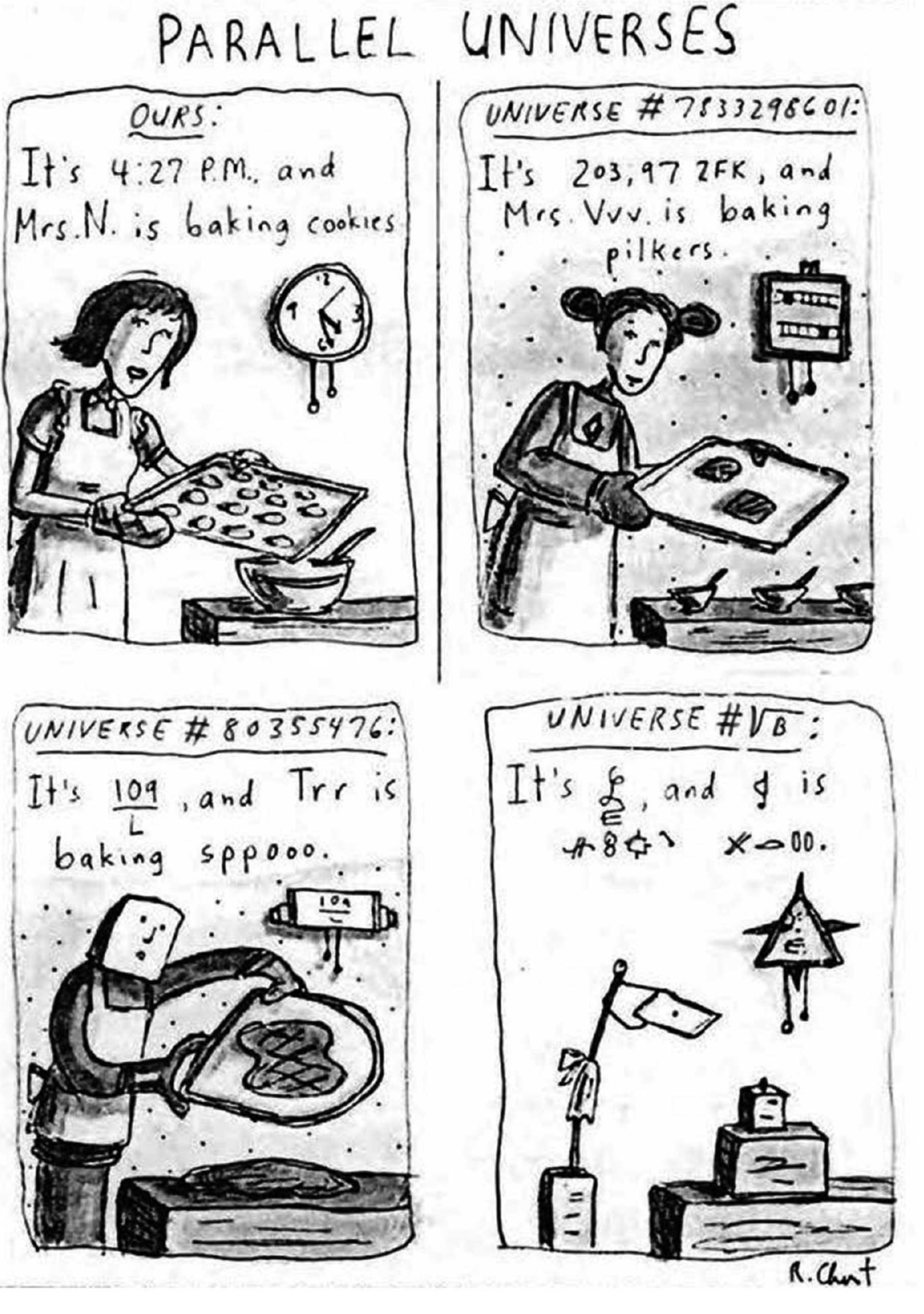
When 'Parallel Universes' appeared, it put many philosophers who saw it in mind of my notorious views. And rightly so: I do claim that there are four such universes. So I thought it would be quote appropriate and fun if your cartoon could appear as a frontispiece in my book. It would please me very much if that could be arranged. …

… It wouldn't do for me to use it if some other author on possible worlds already has. Of course, there are infinitely many other authors who are using it; but I hope all of them are safely off in other worlds, and no thisworldly author has beaten me to it!

Thank you very much for considering my request. And thank you also for the enjoyment that 'Parallel Universes' has given me. …

Sincerely,
David Lewis.

INSERT SCAN OF CARTOON


PARALLEL UNIVERSES
OURS:
It's 4:27 P.M., and Mrs. N. is baking cookies
UNIVERSE #7833298601:
It's 203,97 ZFK, and Mrs. Vvv. is baking pilkers.
UNIVERSE #80355476:
It's 109/L, and Trr is baking sppooo.
UNIVERSE #√B:
It's
, and
is
R. Chast