

THE MULTIVERSE: A PHILOSOPHICAL INTRODUCTION

by Jeremy Butterfield

ISBN 978-1-77385-699-5

THIS BOOK IS AN OPEN ACCESS E-BOOK. It is an electronic version of a book that can be purchased in physical form through any bookseller or on-line retailer, or from our distributors. Please support this open access publication by requesting that your university purchase a print copy of this book, or by purchasing a copy yourself. If you have any questions, please contact us at ucpress@ucalgary.ca

Cover Art: The artwork on the cover of this book is not open access and falls under traditional copyright provisions; it cannot be reproduced in any way without written permission of the artists and their agents. The cover can be displayed as a complete cover image for the purposes of publicizing this work, but the artwork cannot be extracted from the context of the cover of this specific work without breaching the artist's copyright.

COPYRIGHT NOTICE: This open-access work is published under a Creative Commons licence. This means that you are free to copy, distribute, display or perform the work as long as you clearly attribute the work to its authors and publisher, that you do not use this work for any commercial gain in any form, and that you in no way alter, transform, or build on the work outside of its use in normal academic scholarship without our express permission. If you want to reuse or distribute the work, you must inform its new audience of the licence terms of this work. For more information, see details of the Creative Commons licence at: <http://creativecommons.org/licenses/by-nc-nd/4.0/>

UNDER THE CREATIVE COMMONS LICENCE YOU MAY:

- read and store this document free of charge;
- distribute it for personal use free of charge;
- print sections of the work for personal use;
- read or perform parts of the work in a context where no financial transactions take place.

UNDER THE CREATIVE COMMONS LICENCE YOU MAY NOT:

- gain financially from the work in any way;
- sell the work or seek monies in relation to the distribution of the work;
- use the work in any commercial activity of any kind;
- profit a third party indirectly via use or distribution of the work;
- distribute in or through a commercial body (with the exception of academic usage within educational institutions such as schools and universities);
- reproduce, distribute, or store the cover image outside of its function as a cover of this work;
- alter or build on the work outside of normal academic scholarship.



Acknowledgement: We acknowledge the wording around open access used by Australian publisher, **re.press**, and thank them for giving us permission to adapt their wording to our policy <http://www.re-press.org>

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 (sections 1 and 2) 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 (section 3), I will urge that in everyday life, technical science, and philosophy, we are up to our necks in *modality*. This word is philosophical jargon for the topic of necessity, possibility, and impossibility. That is, to state what we believe to be true, whether in everyday life or 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 (sections 4 to 8), 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 (indeed, much 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 towards them. The most confident attitude says: "They are all equally real; the non-actual worlds are merely 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 merely 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 in support of modal realism. 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 conceded that most of the philosophical benefits of using a set of logically possible worlds do not require his modal realism. The benefits 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 using the framework of possible worlds. There are many such concepts and doctrines. But I will restrict my examples to those we will need in later chapters.

Finally, in the fourth stage (sections 9 and 10), 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.

A final preliminary. There is another perspective on the material in the second to fourth stages of this chapter. For the most part, I will not articulate it, since it will be clear what one would say about it. But I will be explicit at the start of the fourth stage (section 9). In short, this other perspective focuses on the idea of a proposition rather than, as I will, on possibility. It will be clear that this difference is largely a matter of jargon, reflecting the fact that “proposition” (and similar words one might use, like “statement”) are really terms of art, to be defined by the logician or philosopher as they

see fit. So, for the most part, when I talk about a possibility or about a possible world, one could instead talk about a proposition or (corresponding to a possible world) about a maximally specific, i.e., logically strongest proposition. But as I said, more details will be provided in the fourth stage.

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 that have persisted. They both involve the idea of *reduction*, and they are the topic of this section. I discuss them in 1.1 and 1.2, respectively.

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

1.1 *Pure Mathematics Reduced to Set Theory*

The 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, 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 subtask, i.e., even if 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 is that a paradox-free set theory adequate for formulating all mathematics can be written in a formal language, with precise vocabulary, and rules of grammar and 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, " $x \epsilon y$ " means that x (which may itself be a set) is an element of the set y .

Besides, the rules of grammar and the rules of inference were also very simple. They were the rules proposed for predicate logic, which 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 for "any" and "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 placement of these words within the argument. Here are two examples, 1 and 2; example 2 also uses some propositional logic.

1. Premise 1: "Some As are Bs."
Premise 2: "All Bs are Cs."
Conclusion: "Some As are Cs."
2. Premise 1: "Some As are Bs."
Premise 2: "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").
Conclusion: "Some As are Cs or Ds" (where "or" is again understood as inclusive).

Thus, the vision has three parts. The first part is about set theory, the second about pure mathematics (apart from set theory), and the third 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, " $\dots \epsilon \dots$," representing set-membership, and whose rules of grammar and rules 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), “ $2/5$,” “ $42/9$,” ..., and the signs $+$ and $\times$ 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.” And so on, for other areas of mathematics.
- 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 explicit 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 reinterpretations 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 reinterpreted, 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 *reduction* 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 *reduced* to set theory. So set theory is called the *reduction basis*.

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, there 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 on 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 pedagogical 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 numerals 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 “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.

1.2 Philosophy Aspires to Provide Reductions

As discussed in chapter 1, much of philosophy has throughout the centuries been about “conceptual housekeeping.” This involves 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* (marketplace) with requests for definitions (analyses) of virtue, courage, etc. And much of philosophy since—focusing on many diverse topics, such as virtue, free will, knowledge, causation, number, and 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 the important accepted truths involving the concept (and of course, kindred concepts) are. 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 as labelled above, by adding to a previously understood and accepted body of doctrine A, some judiciously chosen definitions and 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) he called “logical atomism,” proposed to analyze all of 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. Yet 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, they all gave 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.

2. Logic as a Toolbox of Formal Systems: Modal Logics

I will describe how programmes like Russell’s and Carnap’s led eventually to a modest conception of logic’s role in philosophy, and then I will consider modal logic.

2.1 Logic’s Role in Philosophy

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 on 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 until today. It is 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, and 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 than action 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.”)

2.2 Modal Logic

Aristotle himself initiated modal logic 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 that it is, 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. However, as we saw in chapter 2, philosophy in the modern period, i.e., from the seventeenth century, neglected logic 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 spearheaded in around 1915 by the philosopher Clarence Lewis (usually referred to as “C.I. Lewis,”: no relation to David Lewis, about whom more shortly).

C.I. Lewis was the first person to write down formal logics of modality, called “modal logics.” These 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 analyzes 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 will 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.” (Beware: though I write “N” for “It is necessary that . . .,” it is traditional to write for this, either “L” or a box: $\Box$.)

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)).” (But some expositions do use a separate symbol for “It is possible that . . .” For this, it is traditional to write either “M” or a diamond: $\Diamond$.)

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.” Into this valid form, one inserts “N(P)” in place of “P.”) But what about the converse argument, i.e., the argument “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. Namely, the principles that “It is necessary that P; therefore P” is a valid argument, and that “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 let us 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 in question 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 to come into view, as my mind’s eye goes from the actual world to another possible world.”

So, the interplay between modality and the notion of object is controversial. And the controversies show 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, Saul Kripke, David 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 sections 5 to 7, this semantics also led to detailed proposals about the semantics of natural languages.

Furthermore, some questions central to philosophy turn on examples of principles of the above sort, i.e., principles combining modal operators with the “every,” etc. of predicate logic. In section 8’s discussion of materialism and physicalism, we will see an example of this that uses the principle I have just discussed: i.e., the principle that for a property F, “if everything is necessarily F, then it is necessary that everything is F.” In section 8, the property F in question will be “being material” (or, in the jargon of philosophy, being concrete, as against abstract). As we will see, it is a problematic property.

This completes this chapter’s first stage: a summary of logic’s role in philosophy up to around 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.

3. Up to Our Necks in Modality

In this section—which is the second stage of the chapter—I will argue that 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?

3.1 *What Might Be True: Decision Theory*

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 makeup encoding that it produces chlorophyll, i.e., the green pigment essential for photosynthesis. But this dismissal is unconvincing. For suppose we agreed that grass must be green, and 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 that *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. And 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 use 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, which 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, that is, 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 (as per (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, which 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.

Besides, this involvement with modality holds good for propositions about the past, just as much as for those about the future. For think of memory rather than decision-making. Suppose that yesterday I stayed home, and I learn today that I missed a good film at the cinema, and I regret not going. In such a case, my mind is again focused on a non-actual possibility.

Turning from decision-making and memory to technical science, it is also 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 prepare us for the next chapter.

3.2 Chance and State Spaces in Physics

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, a 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—the atom decayed after an hour, and the atom did not decay.

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. I will 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, momentum (i.e., mass times velocity), or energy. In the jargon of physics, the objects are called

systems, and their properties like position, etc., are called *quantities* (or “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 a particular shade of 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 the system’s values are 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 (i.e., 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 (as I mentioned in chapter 2), mechanics often idealizes the situation. We imagine the sphere is so small and rigid as to be effectively extensionless, i.e., 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 , and writing v_x for the speed in the x -direction and so on, we could write this 6-tuple of all position and momentum values 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 increasingly 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 formulate 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 are they 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, magnetism, and 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*. Of course, for these theories, the quantities involved will in general not be just position and momentum, as in our example of Newtonian mechanics. Quantities such as energy or electric field might be included. (Chapter 4 will give more details about the state space for a quantum system.)

In short, the theories simply cannot be formulated 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.

3.3 Determinism Introduced

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, the fact 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 of determinism 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). Consider again the case of a point particle at some position, with some momentum, at a given time ("now"). And suppose the forces that are exerted on it, not just at the given time ("now"), but throughout the past and the future, are specified. Then, according to Newtonian mechanics, there is a unique history or curve in the state space that passes through the particle's present instantaneous state. (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—for instance, we'll see in the next chapter that the Everettian version of quantum theory is deterministic. (I will return to determinism later in this chapter, in section 8, when I discuss an important philosophical notion: *supervenience*.)

4. A Philosopher's Paradise

So much by way of arguing that our everyday and our scientific beliefs commit us to non-actual possibilities. I turn to this chapter's third stage. In the next five sections (4 to 8), 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, they can be the elements, the states, in 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?” In particular, they can be had without endorsing Lewis’ modal realism, a flexibility that Lewis himself emphasized. (As announced in the chapter’s preamble, I will address this question only in the chapter’s fourth stage, in sections 9 and 10.)

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 in section 1, 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.”

In section 5, I will take as my first example of how useful possible worlds are, *semantics*. More specifically, I will discuss a scheme called *intensional semantics*, which is inspired by ideas of Frege and Carnap about how words get their references in the world. It is clearest to start with Frege, who expresses the ideas without using 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 section 1, 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 our focus, the possible worlds will be the system’s possible histories (curves through state space).

5. Paradise, Part I: Intensional Semantics

I proceed in three stages. In this section, I first present Frege's basic ideas about meaning for simple sentences, without regard to modality. I then extend this to truth-functional compound sentences. Finally, I bring in modality, i.e., possible worlds.

5.1 *Meaning as Referent and Sense*

Frege's first idea is that the meaning of a word or phrase has two main aspects. The idea applies to proper names like "Plato" and "Copenhagen"; to definite descriptions (i.e., "the F" phrases), like "the most famous pupil of Socrates" and "the capital of Denmark"; and to predicates, like "is red," "walks," or "is a horse."

The first and more obvious aspect is the *referent* (*Bedeutung*, in German). This is the object or objects in the world that, as we say, the word or phrase denotes or refers to. For my examples of proper names and definite descriptions, these are, respectively, a human being and a city. By "referent," Frege means the "concrete" object—a human body, a conurbation etc.—located in space and time, and with its countless properties. Such properties can include being tall, for a human; being populous, for a city; being famous, for a human or a city. Thus, the referent is not some feature of the object that is connoted or signalled by the word. Nor is it someone's, e.g., the speaker's, ideas or beliefs about the object. The referent is the object itself. 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," "Plato is the teacher of Aristotle," "Copenhagen is the capital of Denmark," "Copenhagen is the hometown of Kierkegaard," 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, like "the tallest human alive today." Suppose it happens to be a tall German doctor named Gustav Lauben, who lives in Hamburg and is the best-known doctor there. Then clearly, "the tallest human alive today" presents Lauben to us in a different way than "the best-known doctor in Hamburg."

Frege's term for these modes of presentation is *sense* (or *Sinn*, in German). 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 each of us refers by saying “Plato,” he must hold that we each associate an appropriate sense with “Socrates” and “Aristotle,” respectively. In fact, the name “Gustav Lauben” is Frege’s example, which he uses 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. The set of instances is called the predicate’s *extension*. Frege says that the extension, i.e., the set of instances, is the referent of the predicate. (Here, I have simplified the history. In fact, Frege said that a predicate’s referent is a special notion of his own, which he called a *concept* (*Begriff*, in German). Then he added that a concept has an extension. But according to Frege, two predicates refer to the same concept if and only if they have the same extension. In philosophical jargon, we say that Frege’s concepts are *individuated* by their extensions. As a result, most systems of semantics in Frege’s spirit simplify as I have done: they say the referent of a predicate is its extension.)

But the same set of 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, i.e., extension or 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” and “Dr. Lauben walks,” so also their senses combine to make a proposition. And this proposition is true if and only 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.

5.2 Truth Values and Truth Functions

Frege then extends the ideas of referent and sense to propositions and to sentences. Thus, he proposes that the *truth value*, True or False, i.e., the property or status of being true or false, of any proposition is the referent of the sentence expressing it. Thus, the referent of any true sentence is the truth-value True, and the referent of any false sentence is the truth-value False.

Frege’s rationale for this proposal 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, *and* may be treated as a two-place sentence operator. Given two sentences, P and Q, it maps them to the compound sentence “P and Q.” Similarly, *or* is a two-place sentence operator mapping P and Q to “P or Q.”

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).

Thus, Frege can propose that the referent of “and” is this function, from pairs of referents of sentences as inputs/arguments, to referents of sentences, i.e., True or False, as outputs/values. 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, according to Frege, the referent of “and” is a truth function. This function can be exhibited in a *truth table* in which each row shows, for a pair that is an argument of the function, what the corresponding value is. Writing “T” for True and “F” for False, we have the truth table below in Table 5.1.

Table 5.1. Truth table for *and*

P	Q	P & Q
T	T	T
T	F	F
F	T	F
F	F	F

The first two columns display the four possible combinations of truth values for the sentences P and Q, from both True to both False. The third column displays the corresponding truth value of the conjunction, as determined by the truth function *and*.

And similarly, for “or.” In our inclusive and/or sense of “or” the truth table displaying the truth function, i.e., the Fregean referent of “or,” is seen below in Table 5.2.

Table 5.2. Truth table for *or*

P	Q	P or Q
T	T	T
T	F	T
F	T	T
F	F	F

Again, the first two columns display the four possible combinations of truth values for the sentences P and Q. The third column displays the corresponding truth value of the disjunction, as determined by the truth function *or*.

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*.

5.3 *Compositional Semantics with Possible Worlds*

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.

Consider, for example, that 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, for example, for “has a heart.”

All this can be neatly formulated in terms of functions. Since at a possible world “the capital of Denmark” denotes 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 name such as “Denmark,” or a definite description such as “the capital of 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, 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 definite description, 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. I turn to the second, philosophically important, clarification.

(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 that 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. Namely, this semantics—though it invokes other worlds, in some of which people use our words but with different senses—is a semantics for *our* language. That is, it is a semantics for *our* language as we actually speak it, with our senses.

Though this clarification is straightforward, it is important. For as we will see later in this chapter (section 9), the erroneous temptation goes along with a wrong answer to our main philosophical question: what exactly are the worlds?

So much by way of the two clarifications. To finish this exposition, note that just as Frege put to work his assignments to words, of senses and thereby of referents, 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 the Fregean referent of "and" (recall subsection 5.2 above), 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" can be stated in a less cumbersome way: it is precisely the operation of intersection on sets of worlds.)

Finally, a note about jargon. You may 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, this usage is adopted even by those who are wary of 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.

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 the ellipsis . . . So, these are sentence operators: they make a sentence as “output” from a sentence P as “input.”

But note that, unlike “and,” “or,” and “not” (discussed in section 5), they are not *truth-functional*. For suppose P is true but contingent, i.e., could have been false, e.g., “I stay at home tonight,” while Q is true and necessary, e.g., “ $2+2=4$.” Then “Necessarily, P ” is false, while “Necessarily, Q ” is true. So, the truth value of the output of the sentence operator “Necessarily, . . .” does not depend solely on the truth value of the input.

To give a semantics of these operators, the main idea will, of course, be the intuitive (and Leibnizian) one that we already discussed. The idea will be that “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 at that world of the argument/input proposition P) for “ $L(P)$ ” to be true. So here, “ L (. . .)” is not short for “Necessarily, . . .” or “Possibly, . . .”: it is just my notation for some other operator that we will interpret in terms of a subset of W , not the whole of W . (So, it would not be appropriate to call such an operator “Necessarily, . . .,” “Possibly, . . .,” etc. But philosophers still use the word “modality.” That is, they say such an operator represents a *restricted* notion of modality.)

One philosophically important example of such a restriction is the idea of a *law of nature*. (In chapter 1, section 4, 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 eight 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 eight 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—e.g., supplies of rocket fuel, funding, etc.”

Such examples motivate 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) that we can hardly imagine being overturned by future physics. Nevertheless, strictly speaking, “known” implies “being true.” And we must accept that all laws, as presently formulated, even the laws of relativity theory, are fallible.

Thus, such examples suggest that we have a notion of a *law of nature*. Roughly speaking, this is 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 popular suggestion is by David Lewis, building on the ideas of John Stuart Mill and Frank Ramsey. In short, it is that “deeply informative” means being both logically strong and simple. But we do not need the details of this suggestion or of other rival suggestions. We just need the idea that the laws of nature form an elite minority of the countlessly many true propositions about the cosmos, and that the conjunction, I say, of all these laws is thus an elite proposition that is deeply informative about the way the cosmos works.

The Humean tradition (cf. chapter 1, section 4 and chapter 2, section 5) then suggests that although this conjunction L is true, it is *contingent*. It is 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 aficionados: To make this more precise and more convincing, let me keep matters simple by imagining that there is no massive or charged matter, for I agree that 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 to 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 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 eight 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 eight 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 eight minutes.

A final note about 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: or nomological modality).

7. Paradise, Part III: Counterfactual Conditionals

My next example of the philosophers' paradise is *counterfactual conditionals*. These are 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 (it's

“contrary to fact,” hence the name). To say it in terms of possible worlds: P is false at the actual world.

Our discussion will have two stages, in 7.1 and 7.2 below. The first stage is uncontroversial: I will report these propositions’ curious logical behaviour, which was noticed by philosophers and logicians in the 1950s and 1960s. The second stage is more controversial: I will report the proposal, made by Lewis and Stalnaker in about 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 section 6’s idea of restricted modality. For the proposal will be that to formulate what makes a counterfactual true—in our 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, where the collection gets defined in terms of similarity between worlds.

7.1 The Curious Logical Behaviour

A conditional connective “if . . . , then . . .” is a sentence operator. Like “and,” it accepts two sentences as “inputs,” and “outputs” a third sentence. The intuitive idea of a conditional of “if . . . , then . . .” suggests several logical principles that one naturally expects the connective to obey.

One example is *transitivity*. This is the principle that, writing the conditional connective as $\rightarrow$, the following inference is valid, for any sentences P, Q, and R: “ $P \rightarrow Q$; $Q \rightarrow R$. So, $P \rightarrow R$.” 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 (usually making it logically stronger) yields another true conditional. That is, one expects the following inference to be valid, for any propositions P, Q, and R: “ $P \rightarrow Q$. So, $(P \text{ and } R) \rightarrow Q$.”

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

Here is one example showing that counterfactuals do *not* obey transitivity, an example taken from the Cold War in 1950s USA. So, we need to

recall that J. Edgar Hoover was then (in the actual world!) head of the FBI and an ardent anti-Communist. Then the first two statements below are true. Or at least we can take them to be true, 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 that Hoover's ambitious but conformist temperament is retained under the supposition that he grew up in Russia.)
2. If J. Edgar Hoover were a Communist, he would be a traitor.
(The idea here is that 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 least 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 the principle of 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 contexts.)

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 were wet, it would ignite.

7.2 Explained by Degrees of Similarity

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. 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 move away 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, more different from actuality, will carry us to worlds more dissimilar to the actual world, which are represented by dots in a bigger circle. And these explanations are readily visualized.

Here is how this goes, with diagrams, for our two examples. For simplicity, I will adopt Stalnaker’s proposal that for any world *w*, and any proposition *P* not true at *w*, there is a unique world that is most similar to *w* while making *P* true. (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 outlandish, or more different from actuality, to imagine Hoover growing up in Russia than his being a Communist within the USA. So, a world where he is Russian and Communist (and so not a traitor) is more dissimilar from the actual world than a world where he is American and Communist (and so a traitor). To make a diagram to display this, with worlds as dots on the page, we put the dots closer together the more similar the worlds are. It is also usual to write @ for the actual world and *w*₁, *w*₂, etc. for other worlds, and to write beside each world the propositions that are true at it. The

counterfactual conditional is usually symbolized with a box-arrow $\Box \rightarrow$. Hence, with “R” for “Hoover is Russian,” “C” for “Hoover is Communist,” and “T” for “Hoover is a traitor,” we have the following diagram. The distances between the dots, i.e., the varying amounts of similarity between the worlds, make the two counterfactuals “ $R \Box \rightarrow C$ ” and “ $C \Box \rightarrow T$ ” true at the actual world @; and make “ $R \Box \rightarrow T$ ” false at the actual world @. This is seen below in Table 7.1.

Table 7.1. Counterfactuals violate transitivity: the Hoover example

@	• w ₁	• w ₂
$\neg R, \neg C, \neg T$	$\neg R, C, T$	$R, C, \neg T$
$R \Box \rightarrow C, C \Box \rightarrow T,$ $\neg(R \Box \rightarrow T)$		

Transitivity fails because it is a larger departure from actuality to imagine Hoover being Russian and a Communist than his being a Communist within the USA. In the Table, the columns represent the worlds, with a pair of worlds that are adjacent being more similar than a non-adjacent pair. In each column, the first row displays which basic propositions are true at the world in question. The second row shows which counterfactuals are true (though only at the actual world).

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). So, we adopt an obvious notation: “S” for “I strike the match,” “I” for “the match ignites,” and “W” for “the matchbox is wet.” Then the obvious diagram takes all these three propositions to be actually false, gives us the following Table 7.2:

Table 7.2. Counterfactuals violate strengthening of the antecedent: the match example

@	. w ₁	. w ₂
¬S, ¬I, ¬W	¬W, S, I	S, W, ¬I
S □ → I, ¬([S & W] □ → I)		

Strengthening of the antecedent fails because (since the matchbox is actually dry) it is a larger departure from actuality to imagine that I strike the match from a wet matchbox than that I strike the match from a dry one. As in Table 7.1, the columns represent the worlds, with a pair of worlds that are adjacent being more similar than a non-adjacent pair. In each column, the first row displays which basic propositions are true at the world in question. The second row shows which counterfactuals are true (though only at the actual world)

8. Paradise, Part IV: Supervenience: Materialism, Physicalism, and Determinism

My last example of “the philosophers’ paradise,” i.e., the uses of possible worlds, is the notion of *supervenience* (also known as *determination*). This notion is important in many philosophical discussions. But for our purposes, I need to describe how it is useful for formulating (and so also assessing) just three ideas: materialism, physicalism, and determinism. As we will see, for both supervenience in general and for its application to these three ideas, one can be either confident or cautious in the sense of section 4 of chapter 1. In 8.1, I introduce supervenience. Then in 8.2, 8.3, and 8.4, I discuss materialism and physicalism. In 8.5, I discuss determinism.

8.1 Supervenience Introduced

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 concern some single subject matter or topic. Because of this kinship between the properties, it is common to call the sets “families.” So, philosophers speak of supervenience as a relation between families of properties (or attributes). Thus, the objects in the subject matter can be described by the properties they have, and the set of properties amounts to a taxonomy or classification scheme for the objects.

For example, think of botanical taxonomy 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, botanical taxonomy 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-taxonomic facts is just the classification of all appropriate objects (especially the plants) with respect to this family, i.e., 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 (or better, biological taxonomy). 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 jargon: a relation between subject matters, taxonomies, or classification schemes. What relation? The answer: the classification of any of the objects using one set of properties implies how it is classified by the other set. This is as the word “determination” (the alternative jargon to “supervenience”) suggests: the classification of an object using one set of properties *determines* (also “fixes”: in the sense of “makes unique,” 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, F_1 , say, are fixed by all the facts about another family, F_2 . In other words: specifying all the facts about F_2 involves, *ipso facto*, specifying all the facts about F_1 . We say: F_1 *supervenies on* F_2 . We also say: F_2 *subvenes* F_1 . (Thus, “subvenes” is a synonym for “determines.”)

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, e.g., being an oil painting, having magenta in the top left square centimetre, etc. So, the idea is that 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. (Agreed, the example is not wholly uncontroversial. For example, one might claim that being an original is an aesthetic property of a picture, not a “merely historical” property, and if so, aesthetic properties certainly do not supervene on pictorial properties.)

8.2 Materialism

There are various major topics in philosophy where the question of 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? (Here, I use “natural science” to include only physics, chemistry, and biology, i.e., to exclude psychology and other sciences.) 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 I agree that 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.

8.3 Physicalism and Modal Range

Thus arises the doctrine of *physicalism*. This is a strengthening of materialism that gives physics a pre-eminent role in comparison to the other sciences. So, as a claim of supervenience, physicalism amounts to the doctrine that 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 out 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 section 4 of chapter 1. 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 into the details of these debates. Here, I only want to describe one main way that possible worlds help us to 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 only about actual objects, and we set aside possible worlds. That is, suppose we say that two actual objects that match exactly as regards all the properties in a set (family) F_2 also match as regards all the properties in another family F_1 .

Then there is likely to be a problem. For the families of properties F_2 , 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 F_2 has to be rich to have a chance of fixing all of F_1 . And there is the rub. For any reasonably rich taxonomy (family of properties), the actual objects are likely to be a very varied set. I agree that two actual objects often match for some property; both have it, or both lack it. But it is very likely that no two actual objects match for *every* property in F_2 . And if so, the supervenience claim restricted to actual objects—“if they match in this way, then they also match for all of F_1 ”—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 to this problem lies in recognizing that the basic idea of supervenience is modally involved. It is not an assertion only about a case of two actual objects matching for all of F_2 . For as we have just seen, for most of philosophy’s interesting supervenience claims, there are no such cases. Rather, it is about the *trans-world matching* of objects. 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 show how supervenience is about trans-world matching of objects, let us take as an example physicalism, and what it says about, say, the mental life of a cat supervening on its physical state, i.e., on all its physical properties. For illustration, I again take just one mental property (i.e., in the family, F_1): viz. seeing yellow in the top-left of the visual field. Then physicalism says, in particular, that 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 on the answer. What matters for us are three points, all of which echo some previous themes.

8.4 Materialism and Physicalism as Contingent Supervenience Claims

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 whatnot. 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 toward 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 6 of this chapter, one might well be cautious, rather than confident, about the very idea of a law of nature.) So, for example, it is natural for a physicalist to say, “I accept the idea of a law of nature, and I believe they are all contingent. And I claim that physical matching of any two objects implies their *total* matching, across the set of nomically possible worlds.”

Third, my last point returns to the theme at the end of section 2 about the principle about some property F that if all objects are necessarily F, then necessarily all objects are F. This principle's connection with materialism becomes clear when we take as the property F in question being material, that is, being made of matter. So, for this property, the principle says that if all objects are necessarily material, then necessarily all objects are material. This conditional statement obviously relates to the formulation of materialism, and, in particular, to its modal range. Thus, one can imagine a materialist who believes the antecedent of the conditional. They believe that all

actually existing objects are material, and that for each such object, it could not have been immaterial. As they might put it, this actual rock, plant and animal are each of them material; and though each might have had other features (this rock might have been heavier, this plant taller, etc.), none of them could have been immaterial—on pain of not being that very object. But if this materialist takes materialism to be a contingent supervenience thesis (my first point above), they will probably deny the consequent of the conditional statement. That is, they will deny that necessarily all objects are material. For once they consider possible worlds beyond the range of their supervenience thesis—as it might be: worlds that lack the actual world’s laws of nature (my second point above)—they may well allow that some such worlds contain immaterial objects.

Again, in this book, we do not need to pursue the intricacies that these three points reveal about the formulation and assessment of materialism. For us, these points have two relevant morals. First, they show the value of the possible worlds framework for articulating the various philosophical issues and relating them to each other.

Second, these points (especially the third) bring out a feature that will be important for section 9’s question, about what exactly a possible world is. Namely, the feature that the property I called “being material” or “being made of matter” is vague. If the materialist uses it (or some similar phrase) to formulate their materialism, should they take it to require matter of the known kinds—whatever they might choose to mean by “known”? For example, does it include the dark matter that cosmologists nowadays believe in, though they do not know what it is composed of, nor what laws it obeys? In this book, I will not need to resolve this vagueness (not least because materialism is not our main topic). But in sections 9.3 and 10, we will see that this vagueness causes trouble for an otherwise attractive account of what exactly a possible world is.

8.5 Determinism Again

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 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. (More precisely, this uniqueness holds good, once we specify the external influences that the system is subject to during the past and future. Recall the need, at the end of section 3, to know the forces on the point particle.)

It is now clear that, like physicalism, determinism is a supervenience claim—as my word just now, “determines,” rightly 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 (and in the external influences they are subject to) and match exactly in their states at all past and future times. This is clearly a statement of supervenience, namely, the past and future states of the system concerned supervene on its present state—fixing the latter implicitly fixes the former.

Besides, we here see the contrast again, being confident or being cautious, about a concept (cf. section 4 of chapter 1). 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 cautious ones, viz. instantaneous states of the given type of system.

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, section 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, an idea that is like physicalism, as defined above. This idea is that 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 it 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 confident (ambitious) space of all possible worlds, not a cautious (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, i.e., to both past and

future of the assumed matching. (Note: incidentally, a benefit of accepting the idea of the theory of an entire possible world. Since by definition it cannot be subject to external influences—in physics terminology: it is a closed system, not an open one—the definition of determinism as supervenience of past and future on the present does not need to include the qualification about specifying such influences.)

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

9. Existential Angst: What Are Possible Worlds?

So much by way of sketching the philosophical benefits of using possible worlds. And so much by way of tasting the fruits in the philosophers' paradise. I turn, in this section and the next, to the fourth and final stage of this chapter. That is, I turn to the question which I announced in the chapter's preamble: What exactly is a possible world?

As I said there, this question is compulsory, for cautious conceptions of possible worlds as much as for confident conceptions; and several possible answers are defended in 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 invoke non-actual possibilities (cf. section 3 above), what exactly they are is an open and stubbornly difficult question. Hence, this section's title includes *angst*. Besides, focusing on this question illustrates chapter 1's announcement that my discussion of each of the three multiverse proposals will end up 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 (in 9.1 and 9.2). They are tempting, but definitely false. And they are suggested, I am sorry to say, by proposals from renowned philosophers: Berkeley and Wittgenstein. Then, in 9.3, I will discuss a third suggestion that fares better. But it 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, in the literal sense.

9.1 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 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, and regret. Thus, we say that “John imagines/hopes/believes/desires/regrets that *P*.” Philosophers have a term for this distinction. They say that to imagine, to hope, to believe, etc., are propositional attitudes; and that the proposition *P* “on which the mind is focused” 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. (But there will be some good news as a consolation: each horn will teach a philosophical lesson.)

Suppose, first, that the suggestion is that the non-actual possibility is the content or proposition. This suggestion may well be right. For, as I said at the end of this chapter’s preamble, the question of the nature or possibility is tantamount to the question of the nature of a proposition. (I touched on this again 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.) So, our question becomes: What exactly is a proposition?

Now we see that we are no further ahead. For (as I also said) “proposition” and similar words like “statement” are terms of art, with no agreed precise meaning: it is up to each philosopher or logician to say what they mean. Nor does the framework of intensional semantics (reviewed in section 5) help answer the question. There, we saw how it systematically portrays how propositions (taken as the Fregean senses of sentences), the Fregean senses of words and phrases, truth values (True and False), and possible worlds all relate to each other in a compositional semantics. We also saw there how the various senses get expressed by language. But that survey gave no opinion on what a possible world, or more generally, a non-actual possibility or proposition, actually *is*. So, we are no further ahead.

(To further justify a little this verdict of “no progress,” let me sketch the kind of trade-off between taking as basic possible worlds or propositions. Thus, in section 5, possible worlds and truth values were basic posits, not further analyzed. 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. But I agree that 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 that 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 subsection 5.3 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 beings.” No sentience—in particular, no visualization 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 claims made by the eighteenth-century idealist philosopher Berkeley, in his *Treatise Concerning the Principles of Human Knowledge* (1710). Berkeley claims that (i) we cannot imagine an unperceived object, and therefore (allegedly!), that (ii) any object must be perceived. (He sums this up in a famous slogan, that for an object, to be *is* to be perceived, or in Latin, *esse est percipi*.)

The flaw lies in an equivocation. If (i), i.e., “we cannot imagine an unperceived object,” means “we cannot imagine a possible world containing an unperceived object,” then (i) is false. (Just think of the boulder world, above.) But I am willing to concede: so, understood, (i) implies (ii). That is,

if (i) were true, (ii) would also be true. If, on the other hand, (i) means “we cannot imagine ourselves in a possible world without perception,” then I can concede that (i) is true. Indeed, it is necessary if “ourselves” implies being able to perceive. 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 now turn to my second tempting, but wrong, suggestion.

9.2 Combinations?

Here, the “culprit” will be Wittgenstein, in his early work, the *Tractatus Logico-Philosophicus* (1921), a work which he later disavowed, partly for the reasons I will present. Indeed, we will see that for our purposes, he is “more guilty” than Berkeley. For he does not just make claims that prompt the false 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 *necessary*? What explains that?

This echoes the questions we pursued at the end of chapter 2 (section 8) and the start of this chapter (section 1). These questions are: 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 *Tractatus*, suggested that for any necessary proposition, whatever its subject matter, its necessity is exactly like that of what propositional logic calls *tautologies*. 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 *truth functions*), 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, such as “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).” (Here “or” is understood, as usual, in our inclusive and/or sense.)

The simple diagrams called *truth tables*, introduced in section 5, make 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 we apply the truth functions to calculate, in each row, what the truth value is of the compound sentence. We say that the compound sentence is a *tautology* if it comes out True in every row.

Thus, to calculate the truth table for “(P or Q) or (not-P),” we need four rows: one for “P is true and Q is true,” one for “P is true and Q is false,” one for “Q is true and P is false,” and one for “P is false and Q is false.” Then, by applying the truth functions “not” and “or” appropriately, we calculate that “(P or Q) or (not-P)” comes out True in every row. This is seen below in Table 9.1.

Table 9.1. Truth table for a tautology

P	$\neg P$	Q	P or Q	(P or Q) or $\neg P$
T	F	T	T	T
T	F	F	T	T
F	T	T	T	T
F	T	F	F	T

For some compound sentences, such as (P or Q) or $\neg P$, the placing of connectives, such as “or,” “and”, and “not,” makes the sentence come out true for any combination of truth values of the constituent sentences.

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. In short, it is 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 combinations of truth values are genuinely possible. This is called their being *logically independent*. So, the idea is that 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 or that combination.

Thus, Wittgenstein proposed that all necessity had this lucid *combinatorial* origin and explanation: that all necessary propositions are really—could be analysed into—tautologies.

So far, this is entirely programmatic: a mere declaration. Indeed, there are three large kinds 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,” and “equilateral triangles are equiangular,” seem to be necessary. But this necessity seems to be very different from—and much more problematic than—the placing of “and,” “not,” and “or” in any sentences. Recall from chapter 2, sections 7 and 8, the struggles since Kant to explain the necessity of mathematics, and in particular, the growing separation of pure and applied mathematics, with the former accorded a special *non*-empirical subject matter, such as numbers and geometric figures. And though the logicism of Frege and Russell tried to reduce pure mathematics to logic, they really succeeded in reducing it to set theory. Agreed, that was a great achievement, as extolled in section 1 above. But set theory is not logic: it has a special non-empirical subject matter, viz. sets.

Second, there are propositions (about any subject matter) whose necessity turns upon the placing of logical words *other than* “and,” “not,” and “or,” especially those other logical words, “all,” “any,” “some,” and “none,” that—as also explained in section 1 above—are studied in predicate logic. For example, consider the necessary proposition (for any predicates A and B) “Either it is not true that everything is both A and B, or something is A.” Its necessity is a matter of the placing of, not just “and,” “not,” and “or,” but also “everything” and “something.” So, even if we set aside pure mathematics, i.e., the first kind of necessary proposition above (whether or not it is really set theory), here is another kind of necessary proposition that is not a matter of tautologies. Besides, we saw in section 2 that the logical words on which propositional and predicate logic focus—“and” and its brethren, “all” and its brethren—are not all the logical words, i.e., words on the placing of which the necessity of a proposition (or the validity of an argument) can depend. We saw, in particular, that C.I. Lewis developed modal logic, about the logical words “It is necessary that . . .” and “It is possible that . . .”

(An incidental comment. Here, the propositional form “Either not-P or Q” is logic’s much-used “weaker cousin” of “If P then Q.” Thus, one readily agrees that the proposition expressed by the sentence “If everything is both A and B, then something is A” is necessary; whereas my example above is a tongue-twister, whose necessity is hard to see. But I use the tongue-twister form, “Either not-P or Q,” to avoid subtleties about the meaning of the

English “If . . . , then . . . ,” subtleties that are not needed in this book, but that are rather like those we saw for the counterfactual conditional in section 7.)

Third, there are propositions whose necessity depends not upon the placing of any logical words in the sentence expressing them, but upon relations between the meanings of other words. For example, consider “All bachelors are unmarried” and “A vixen is a female fox.” (Though philosophers use “analytic” in different senses, the common thread in their usages is undoubtedly the idea of being true in virtue of the meanings of words. So, these sentences would certainly count as analytic.)

To sum up, in order to show that all the necessary propositions of these three kinds are really tautologies, one would have to show, somehow or other, that they are built up from—can be analysed into—component propositions that are logically independent, i.e., component propositions for which every combination of their truth values is genuinely possible. That would be a programme of *reduction*, in roughly the sense of section 1 above.

Wittgenstein, in the *Tractatus*, was committed to such a programme. He 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 kinds of necessary propositions above. For example, maybe “all” could be reduced to the idea of conjunction, though a possibly infinite one (and similarly, “some” might be reduced to a possibly infinite disjunction). But few details were given. The main lacuna is his silence about the third kind—the difficulties about analyzing them as tautologies—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. The problem is simply stated. The different values of a quantity—whether a physical quantity like the position of a point particle, or a psychological quantity like “magenta in the top-left of the visual field”—obviously exclude one another. And this means that the propositions ascribing such values cannot be logically independent. Far from it: That values exclude each other means precisely that each proposition ascribing a specific value implies the negation of each of the other propositions ascribing specific values.

9.3 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 invoke (cf. section 3) never get represented in even one inscription or utterance of a sentence. And setting aside whether there is an actual inscription or utterance, surely most cannot be represented in all their myriad details 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 of an existing language, such as English, to sets of actual objects, and also sets of 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, section 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 5.3 above'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.

To many philosophers, this suggestion has seemed promising. One main reason they find it attractive is that it seems to secure for us a deeply felt contrast between the actual world and all other worlds, viz. that the actual world is “concrete” while non-actual worlds are “abstract.” As the scare-quotes show, these two words are philosophical jargon, and vague: “concrete” does not mean “cement”! But many philosophers, though they accept sets built from actual objects, etc., as being themselves legitimate objects, think of sets as abstract objects; and they also think of ordinary actual objects as not abstract, which they dub “concrete.” So, to these philosophers, the suggestion that any non-actual possible world *is* a set, while the actual world is not a set, seems to classify worlds rightly, as regards concrete vs. abstract.

But I fear that this suggestion does not work. There are several objections, but I shall present just two. They are urged by Lewis himself, along with others. My presentation summarizes some passages in his *On the Plurality of Worlds*, viz. in section 1.7 and section 3.2, 150 f.

9.3.1 TROUBLES WITH THE CONCRETE/ABSTRACT DISTINCTION

The first objection is that the advantage just mentioned, of classifying worlds rightly as regards concrete vs. abstract, is spurious, for three reasons, of which I think the third is more important.

1. Since there will be a set-theoretic representation, a mock-up or replica, of the concrete actual world, just as there are of non-actual worlds, intensional semantics will presumably specify the actual referents of our words as ingredients (barely visible, deep in a forest of curly brackets) of the actual world’s mock-up. But that apparently conflicts with the idea that we refer to concrete objects.
2. Having the actual world be concrete, and all the other worlds abstract, is an “absolute fact,” holding across all the worlds, rather than a proposition that is true at individual worlds. This engenders a conflict with the idea that much of what is actually true is contingently so, i.e., that it is contingent which world is actual. We should no doubt hold fast to that idea, since it is at the root of our commitment to other possibilities, at the root of our being up to our necks in modality (section

- 3). But if the actual world is concrete, and all the others abstract, the idea that “things could have gone differently” requires that an abstract item could have been concrete—which most philosophers who endorse a concrete/abstract distinction would resist.
3. The words “concrete” and “abstract” are not just vague, in the sort of way that being red or being bald are vague, namely, made precise by specifying a position somewhere in one, or perhaps a few, reasonably well-defined spectra, like hue, brightness, and number of hairs on the head. Concrete and abstract are not so much vague as ambiguous, in the sense that making them precise is not a matter of position in one or a few spectra. Thus, different philosophers make “concrete” precise in very different ways. Some only give examples. The paradigm, much-used example, is tables and chairs (wittily dubbed “medium-sized dry goods” by the philosopher J.L. Austin), but one surely should add much smaller and much bigger “ordinary objects,” such as bacteria and stars. On the other side, the paradigm, much-used example, of “abstract” is sets and other objects of pure mathematics, such as numbers and geometrical figures. But listing examples gives no indication of where the boundary lies. And the general notions often invoked cut across one another. Thus, should “concrete” be taken to mean being material? As I mentioned in section 8, this is itself ambiguous: Is it matter of a familiar sort? Having mass? Having energy?). Or should “concrete” be taken to mean being located in space and time? Or as being either a cause or an effect (or both)? Besides, a good case can be made that those paradigm abstract objects, sets, can be concrete according to some of these proposals. For example, why not say that a set of ordinary objects, each of which is located in space and time, is multiply located at the places where the objects are? And some philosophical accounts of causation take causes and effects (i.e., events, the relata of causation) to be sets. To sum up this discussion, the concrete/abstract distinction is very unclear. Of course, this is not the place to try and “clean it up.” But all I need here is the point

that, because it is unclear, the suggestion above, that non-actual worlds are sets, cannot claim any substantive merit in classifying such worlds as abstract.

9.3.2 ASSUMING THE NOTION OF POSSIBILITY

The second objection is that 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, no syntactic, test for consistency. For inconsistency is not just a matter of the set containing “P” and “not-P,” for some P. (Equivalently, it’s not just a matter of the long conjunction containing “P” and “not-P” as conjuncts.) That is only one very simple way to be inconsistent. Relations between the meanings of non-logical words provide many other examples. Here we return to the third kind of proposition that beset the early Wittgenstein’s combinatorial account of necessity (cf. section 9.2 above). Think of “Fred is a married bachelor” or “A male vixen got into the chicken hutch.” And there is no reason to think that we can somehow analyze all our language, so as to devise a syntactic test for consistency. (In particular, there is no reason to think as the early Wittgenstein did that any proposition is a truth function of a set of logically independent propositions, so that consistency can be tested by truth tables, i.e., ascertained by finding at least one row with a “T.”)

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, so also 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 August 2024.” 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 the following very simple representational conventions. (They probably work smoothly only for very simple examples, but that will not matter.) We take a period of time, such as August 2024, 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 “August 2024” 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-August-2024.” 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 August 2024. 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. It is the fact that Butterfield is a member of (the set that is) the referent of “Rome-in-August-2024,” at various worlds. (Among 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 August 2024, 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 well. For I have exhibited a set that is appropriately structured to *be* the possibility that Butterfield is in Rome in August 2024.

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 time, 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*, indeed, 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. The sets will not “lead us astray” by

making impossibilities appear possible, masquerading in an appropriately structured nest of curly brackets.

But, of course, even if we make this assumption, the basic 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 objection to possible worlds (or possibilities) being sentences or sets as follows. Saying that the sentence “Butterfield could be in Rome in August 2024” 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.

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. Here, I just briefly report three main aspects, out of many.

1. He replies in great detail to various objections to his view (several of which he himself thought of). In particular, he “takes the sting” out of the objection that, according to him, the possible worlds are each concrete, just like the actual world is concrete—and that surely non-actual worlds should be classified as abstract. Namely, as I reviewed in section 9.3 above, the concrete/abstract distinction is so unclear that this “surely” claim stumbles.
2. He defends his view about what it is for an object to be in two worlds, as in the previous section’s closing example of Butterfield and Rome. In short, he denies that the selfsame object can be in two worlds. Instead, given an object in one world, another world may contain a suitably similar object, which Lewis calls a *counterpart* of the given one.
3. He also 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, his views satisfy our requirement, at the end of chapter 1 (section 6), that advocacy advocating a multiverse should not be undermined by the bewildering idea that most of one’s readers or hearers are in another universe.

I shall say more about 1 to 3 in section 2 of chapter 6.

But to conclude, 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 of 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 to this question. And the topic of the next chapter, the Everettian interpretation of quantum theory, will suggest another answer—another conception of what a possible world exactly is. I myself do not find it persuasive, but it is certainly worth considering, and I will do so in chapter 6.

In any case, this is not the place to further expound or assess Lewis' views. For our purposes, it suffices to have established in this chapter the following three main points. (i) We are, in our thought and language, up to our necks in modality; (ii) logicians and philosophers have developed detailed frameworks for describing and analyzing 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 from June 15, 1984, to the cartoonist Roz Chast, he asked to use her witty cartoon "Parallel Universes" from *The New Yorker* as the frontispiece of his book, *On the Plurality of Worlds*. The cartoon represents "our universe" as a realistic scene of a woman baking cookies, and three successively more dissimilar universes by weirder and weirder analogues of the first scene. See Figure 10.1. (Unfortunately, the cartoon was not used in the book.)

In this letter, Lewis 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 been helping myself to the other worlds when I wrote about one or another philosophical problem. But I never wrote at length about what it means 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 quite 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. . . .

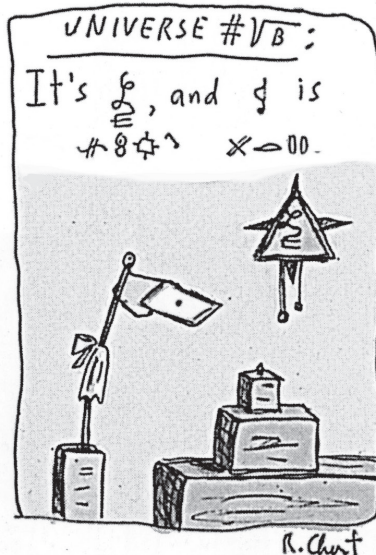
It wouldn't do for me to use it if some other author on possible worlds already has. Of course, there *are* infinitely many other authors who are using it; but I hope all of them are safely off in other worlds, and no this worldly author has beaten me to it!

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

Sincerely,

David Lewis.

PARALLEL UNIVERSES



R. Chast

© R. Chast, from the New Yorker, 1984. Used with permission. The cartoonist has serendipitously illustrated Lewis' doctrine that possible worlds can be more or less similar to the actual world: in this sequence, ever less similar...

11. Notes and Further Reading

11.1 Primary Sources

Since this chapter's topic, logic, has been centre-stage in philosophy for a century, there is an enormous amount of literature on it. But, like for chapter 2, my main suggestion for reading is some of the original masterpieces. As I said for chapter 2, though daunting, one should at least dip into them. For such masterpieces, I will emphasize Frege and Lewis, who—as is clear from the chapter—are my heroes.

FREGE

Foundations of Arithmetic (1884) and *The Frege Reader*, ed. M. Beaney (Wiley-Blackwell, 1997) are essential. For material in section 5 of this chapter, I recommend two of his essays (both included in *The Frege Reader*):

1. Gottlob Frege. "On Sense and Reference" (reprinted in *The Philosophical Review*, 1948); available at: JSTOR: <https://www.jstor.org/stable/2181485>
2. Gottlob Frege. "The Thought: A Logical Inquiry" (reprinted in *Mind*, 1956); available at: <https://www.jstor.org/stable/2251513>

Note: In the title of 2, *Thought* (*Gedanke* in German) is Frege's technical term for what analytic philosophers usually call a proposition, i.e., the content or meaning of a sentence. Frege's term *gedanke* is unfortunate since he intended his notion to abstract away from the psychological aspects of meaning.

LEWIS

On the Plurality of Worlds (Blackwell, 1986) is the pre-eminent reference for this chapter. As I said in section 4, the phrase "a philosopher's paradise" is his. It is the title of that book's chapter 1, which develops the themes of my sections 5 to 8. The book's other chapters (chapters 2 to 4 respectively) (i) answer objections, (ii) rebut accounts of possible worlds that are rivals to his modal realism, and (iii) develop his counterpart theory account of what it is for an object to be in two worlds (as mentioned in my subsection 2 of section 10).

Further details are in some other masterpiece papers by Lewis. His main exposition of intensional semantics (my section 5) is in “General Semantics” in the journal *Synthese* (1970), and is available in the JStor archive at: <http://www.jstor.org/stable/20114749>. A companion paper synthesizing this semantics with his account of the pragmatic and social aspects of language (about which I have said nothing) is “Languages and Language,” in *Minnesota Studies in the Philosophy of Science*, vol. 7, edited by K. Gunderson (University of Minnesota Press, 1975). Both these papers are reprinted in Lewis’ first collection of selected papers, called *Philosophical Papers*, vol. I (Oxford University Press, 1983), which is available at: <https://academic.oup.com/book/36015>

Furthermore, these papers, together with all (so far as I know!) of Lewis’ papers, can be downloaded from a website built by Andrew Bailey (which also contains details of his books and book reviews). It is available at: <https://andrewmbailey.com/dkl/>

Of his books other than *On the Plurality of Worlds*, the one most relevant to this chapter is his *Counterfactuals* (Blackwell, 1973). It expounds his version of the analysis I summarized in section 7, invoking similarity between possible worlds. It also gives (in its section 3.3) Lewis’ first statement of his account of laws of nature, which is Humean and kindred to ideas in Mill and Ramsey (my section 6 above), which has been very influential in philosophy of science.

Finally, Lewis’ philosophical correspondence has been published by Oxford University Press in two volumes, edited by H. Beebe and A. Fisher. Volume 1 (about causation, modality and ontology, and containing the letter to Roz Chast, which I quoted in section 10) is available at: <https://global.oup.com/academic/product/philosophical-letters-of-david-k-lewis-9780198855453?lang=en&cc=gb>. And Volume 2 (about mind, language and epistemology) is available at: <https://global.oup.com/academic/product/philosophical-letters-of-david-k-lewis-9780198855842?cc=gb&lang=en&q=Industrial%20Policy%20and%20Development:%20The%20Political%20Economy%20of%20Capabilities%20Accumulation&tab=overview>

BOOKS BY BERKELEY AND WITTGENSTEIN

- **Berkeley:** *A Treatise Concerning the Principles of Human Knowledge* (1710); available at: <https://www.cambridge.org/core/books/>

berkeleys-a-treatise-concerning-the-principles-of-human-knowledge/DAB1D1CB81E7D0659900B4CDF270E3C2

- **Wittgenstein:** *Tractatus Logico-Philosophicus* (1921); available at: <https://archive.org/details/tractatuslogicop1971witt/page/n5/mode/2up>

11.2 Secondary Sources

Secondary readings provide context and interpretation.

Internet resources

- *Stanford Encyclopedia of Philosophy*, entry on “Possible Worlds”; Lewis’ modal realism (section 2.1) and combinatorialism as in my section 9.2 (section 2.3); available at: <https://plato.stanford.edu/entries/possible-worlds/>

BOOKS

Here, by way of example, are two books. The first is about Leibniz’s views, the second about determinism (cf. my section 8.5). Most of Earman’s writings, throughout the philosophy of science, can be downloaded from the site listed.

- Mates, Benson. *The Philosophy of Leibniz: Metaphysics and Language*. Oxford University Press, 1989; Oxford Scholarship Online; available at Internet Archive: <https://archive.org/details/benson-mates-the-philosophy-of-leibniz-metaphysics-and-language>
- Earman, John. *A Primer on Determinism*. Kluwer, 1986; available at: https://sites.pitt.edu/~jearman/Earman_1986PrimerOnDeterminism.pdf (accessed January 13, 2026)

11.3 Reflections on Two Themes

So much by way of masterpieces and secondary reading in the philosophy of logic and language. Going beyond this, I will here give a bit more detail and some more references about just two of this chapter’s themes. First, I will discuss the subject matter of pure mathematics; this will develop themes

in my sections 1 and 9.3.1. Next, I will further discuss the endeavour of reduction (cf. section 1).

My rationale for these two choices is that discussion of the first theme will lead to a brief discussion of another “Pythagorean” multiverse proposal, while discussion of the second will connect reduction with other philosophical themes such as supervenience (cf. section 8) and emergence, which will figure in the next chapter.

PURE MATHEMATICS

This chapter touched on the philosophical question of what the subject matter of mathematics is, at two places. Section 1 reviewed the achievement of the early twentieth century in casting all of pure mathematics as part of set theory; section 9.3 criticized the concrete/abstract distinction as being very unclear, so that, in particular, saying that sets are abstract might not prevent them from being located in space and time, or from being causes or effects.

The point now is that those two discussions pull in opposite directions—and that seeing the tension between them can prompt a “Pythagorean” view of the nature of mathematics. For the first discussion consolidates the late nineteenth-century distinction between pure and applied mathematics (cf. sections 7 and 8 of chapter 2), a distinction that seems, when one hears philosophers say “concrete” and “abstract,” to be precisely a distinction between studying abstract objects (as in pure mathematics) and studying concrete objects (as in applied mathematics). So, since the second discussion criticized the concrete/abstract distinction as unclear, what should we conclude about the validity of the distinction between pure and applied mathematics?

This is a live question in the philosophy of mathematics. We philosophers do *not* have an agreed uncontroversial view of what the objects of mathematics—numbers, geometrical figures etc.—*really are*, notwithstanding the twentieth century’s achievement in showing that by adopting appropriate definitions of them as sets, one can recover, i.e., derive, the sentences taken as true in mathematics: sentences like “ $2+2=4$ ” and “all equilateral triangles are equiangular”—albeit now understood as about certain sets, not about numbers and triangles as *sui generis* entities.

Of course, this is not a book about the philosophy of mathematics. Fortunately for me. So, I do not need to justify an answer to the above question, or to related ones, like (i) how best to repair (i.e., make precise) the concrete/abstract distinction, or indeed the basic question (ii) what exactly is a number, or a triangle?

But in philosophy, perhaps more than any other discipline, one question leads to another. And indeed, if one rejects the concrete/abstract distinction, one may be tempted by what I called a “Pythagorean” view of the nature of mathematics. This is the view that the world, i.e., the actual world of tables and chairs (“medium-sized dry goods”) and bacteria and stars (cf. section 9.3) *is* mathematical. That is, the world is not just accurately described by mathematics but is *made of* mathematical objects. Numbers and triangles are literally *in* the actual (allegedly concrete!) world.

And for us, with our focus on multiverse proposals, this Pythagorean view is relevant in two ways. First, if true, it would alter (but not necessarily solve) the problem that in section 9.3, we saw confronting the suggestion that non-actual possible worlds are sets (“abstract”) while the actual world is not a set (“concrete”). The problem was that set theory is too “profligate” in its ability to construct sets. That is, under whatever representational conventions we adopt, there are bound to be sets that represent impossibilities (like Butterfield being a fried egg) in just the same way that there are sets representing possibilities. So, on the Pythagorean view, the status of this problem—Is it solved? Is it still recalcitrant?—will depend on what the Pythagorean says about the constructive, or generative, power of its in-the-actual-world mathematical objects. Maybe it can somehow avoid being profligate and representing impossibilities. But the jury is out.

Second, I should report that the popular book advocating a multiverse (in several senses), which I recommended in the final section of chapter 1, also advocates this Pythagorean view of mathematics. The book is Max Tegmark’s *The Mathematical Universe* (2014). In a review of it, I criticized Tegmark’s proposals, especially his Pythagorean view, at some length. I will not here repeat the details of my criticisms (for the review is also cited in that section and is on the internet). But to help orient the reader, I will just summarize his claims and the “core” of my critique.

First, Tegmark advocates the cosmological multiverse and the Everettian multiverse, which I will treat in the next two chapters. He labels these multiverses as “Levels,” and he distinguishes within the cosmological multiverse whether the laws of physics vary across the different universes that are contained in the multiverse. (I offer more details about this idea in chapter 5.) So, Tegmark labels the cosmological multiverse as comprising “Level I” and “Level II.” And he labels the Everettian multiverse as “Level III” (more details in my chapter 3). But the relevant point here is that he then goes on to argue not just that his cosmological-cum-Everettian multiverse

is described by mathematics, i.e., instantiates a mathematical structure, but that it *is* mathematics. This is, of course, the Pythagorean view discussed above.

He also says that all mathematical structures exist, including all the structures large and intricate enough to encode or represent, as we would naturally say—though Tegmark would say: be—various possible cosmological-cum-Everettian multiverses. (Tegmark does not say how “various possible” should be understood, but never mind that here.) So, the upshot is that the cosmological-cum-Everettian multiverse that Tegmark first advocated, and labelled Levels I to III, is just one of countless many mathematical structures. They are all equally real, just as real as the multiverse at Levels I to III, which he first advocated. Thus, Tegmark is claiming that all of reality, comprising the physical and the mathematical, is a mathematical multiverse, which he labels “Level IV.”

So much by way of summary. The “core” of my critique lies in the fact that even if one is a Pythagorean like Tegmark, the distinction between pure and applied mathematics remains. One sees the distinction in play, in the idea of a *physical quantity*.

All agree that when physics describes the world using mathematics, it does not just attribute a pure (“raw”) number (or similar quantitative measure or magnitude) to “bits of reality.” The attribution is always of some number of units of a physical quantity: 5 units of energy, 7 units of angular momentum, 9 units of electric charge, etc. Without mention of the quantity concerned, the description is so incomplete as to be meaningless, e.g., “This object has number 5.” Thus, even if numbers and the other objects considered part of the subject matter of pure mathematics are in the world, as the Pythagorean claims, nevertheless, there is undoubtedly more to the physical world than these numbers, etc. Namely, there is the pattern of occurrence of the quantities (where “pattern” includes their relations to one another and the relations of their values, as stated in the laws of a physical theory).

On the other hand, no such quantity gets mentioned in a work of pure mathematics. That is, the enterprise of pure mathematics, as it is conceived today, wholly disregards which quantities physics needs (energy, charge) and which it does without (such as erstwhile contenders like caloric). (If like me, you endorse the Humean doctrine from chapter 2 that physics is contingent, you may well see (as I do) this disregard as part and parcel of pure

mathematics' enterprise being to formulate, and justify by proof, necessary propositions.)

In this way, the distinction between pure and applied mathematics is, I think, mandatory in the light of the development of logic and mathematics in the last 150 years, as we reviewed at the end of chapter 2 and in this chapter. Indeed, it is a defect of Tegmark's book that he does not connect his multiverse proposals to this development, nor to its philosophical ramifications, such as this chapter's philosophical, i.e., modal multiverse. In any case, the upshot for Tegmark is that we can accept his "Level IV" claim that all mathematical structures exist . . . but if this means *pure* mathematical structures in the modern sense, i.e., regardless of physical quantities, then his claim has little bearing on the debates about the physical multiverses of his Levels I to III. And accordingly, the rest of this book will set his Level IV aside.

Finally, a note about using the label "Pythagorean" for the view that the empirical world is made of numbers. The historical Pythagoras (ca. 572–497 BC) is lost in the mists of time; see, for example, the entries "Pythagoras" and "Pythagoreanism" in *The Stanford Encyclopedia of Philosophy*. So, for our purposes, the label is admittedly anachronistic: it is just rooted in the fact that Pythagoras seems to have led a sect of "number-mystics."

The broader theme here is, of course, the fact that philosophers have for centuries—from Plato to Russell—been preoccupied by the nature of mathematics, especially as a realm of certain knowledge apparently not derived from experience. Cf. again, the end of chapter 2, and from its last section, the three essays by J. Bennett, M. Burnyeat, and I. Hacking in *Mathematics and Necessity: Essays in the History of Philosophy* (edited by T. Smiley).

REDUCTION

In section 1, I reviewed the idea of the *reduction* of one theory to another, understood as adding to the latter ("reducing") theory some judiciously chosen definitions of the terminology of the former ("to-be-reduced") theory that enable a deduction of the claims of the latter (those claims now reinterpreted through the definitions). In the first half of the twentieth century, the paradigm example was the reduction of pure mathematics (arithmetic, algebra, geometry, etc.) to set theory, which, as I said, was very influential in the philosophy of science, as a template for how a pair of scientific, in particular, physical, theories might be related.

Of course, this doesn't apply to all pairs of scientific theories. Reduction might fail: one theory could be irreducible to another. And there seem to be other important intertheoretic relations apart from reduction and its denial. In particular, a theory might *supervene* on, or be *determined* by, another theory (in senses of "supervenience" and "determination" like those in section 8 above), and a theory might be *emergent* from another (a topic which will figure in the next chapter).

Hence, there is nowadays a large literature in the philosophy of science about these various intertheoretic relations. Since I have written three articles that each try to both survey the situation and to argue for some topical, perhaps contentious, claims, I recommend them here.

The first two (from 2011) form a pair, focused on the relations between reduction, supervenience, and emergence. The third relates reduction to a doctrine, *functionalism*, which was first introduced as a label for a position in the philosophy of mind, but was recently brought into the philosophy of physics. (The position is vague, but normally associated with a failure of reduction, though compatible with supervenience. I aim to show that the work of Lewis (the same one!) makes functionalism more precise and shows it to be compatible with reduction.)

Here are my three articles:

- Butterfield, Jeremy. "Emergence, Reduction and Supervenience: A Varied Landscape." *Foundations of Physics* 41 (2011): 920–960. Available at: <https://arxiv.org/abs/1106.0704>; <http://philsci-archive.pitt.edu/5549/>
- Butterfield, Jeremy. "Less is Different: Emergence and Reduction Reconciled." *Foundations of Physics* 41 (2011): 1065–1135. Available at: <https://arxiv.org/abs/1106.0702>; <http://philsci-archive.pitt.edu/8355/>
- Butterfield, Jeremy, and H. Gomes. "Functionalism as a Species of Reduction." In *Current Debates in Philosophy of Science*, edited by C. Soto, 123–200. Springer, Synthese Library 477, 2023. Available at: <https://arxiv.org/abs/2008.13366>; <http://philsci-archive.pitt.edu/18043>