4. Foundations

4.1 Possible Worlds

It is time to face the fact that my analysis rests on suspect foundations. Doubly so: possible worlds are widely regarded with suspicion, and so is similarity even among entities not themselves suspect. If the common suspicion of possible worlds and of similarity were justified, then my analysis could have little interest: only the interest of connecting mysteries to other mysteries. I shall argue, however, that the suspicions are not well justified.

I believe that there are possible worlds other than the one we happen to inhabit. If an argument is wanted, it is this. It is uncontroversially true that things might be otherwise than they are. I believe, and so do you, that things could have been different in countless ways. But what does this mean? Ordinary language permits the paraphrase: there are many ways things could have been besides the way they actually are. On the face of it, this sentence is an existential quantification. It says that there exist many entities of a certain description, to wit 'ways things could have been'. I believe that things could have been different in countless ways; I believe permissible paraphrases of what I believe; taking the paraphrase at its face value, I therefore believe in the existence of entities that might be called 'ways things could have been'. I prefer to call them 'possible worlds'.

I do not make it an inviolable principle to take seeming existential quantifications in ordinary language at their face value. But I do recognize a presumption in favor of taking sentences at their face value, unless (1) taking them at face value is known to lead to trouble, and (2) taking them some other way is known not to. In this case, neither condition is met. I do not know any successful argument that my realism about possible worlds leads to trouble, unless you beg the question by saying that it already *is* trouble. (I shall shortly consider some unsuccessful arguments.) All the alternatives I know, on the other hand, do lead to trouble.

If our modal idioms are not quantifiers over possible worlds, then what else are they? (1) We might take them as unanalyzed primitives; this is not an alternative theory at all, but an abstinence from theorizing. (2) We might take them as metalinguistic predicates analyzable in terms of consistency: '*Possibly* ϕ' means that ϕ is a consistent sentence. But what is consistency? If a consistent sentence is one that could be true, or one that is not necessarily false, then the theory is circular; of course, one can be more artful than I have been in hiding the circularity. If a consistent sentence is one whose denial is not a theorem of some specified deductive system, then the theory is incorrect rather than circular: no falsehood of arithmetic is possibly true, but for any deductive system you care to specify either there are falsehoods among its theorems or there is some falsehood of arithmetic whose denial is not among its theorems. If a consistent sentence is one that comes out true under some assignment of extensions to the non-logical vocabulary, then the theory is incorrect: some assignments of extensions are impossible, for instance one that assigns overlapping extensions to the English terms 'pig' and 'sheep'. If a consistent sentence is one that comes out true under some possible assignment of extensions, then the theory is again circular. (3) We might take them as quantifiers over so-called 'possible worlds' that are really some sort of respectable linguistic entities: say, maximal consistent sets of sentences of some language. (Or maximal consistent sets of atomic sentences, that is *state-descriptions*; or maximal consistent sets of atomic sentences in the language as enriched by the addition of names for all the things there are, that is *diagrammed models*.) We might call these things 'possible worlds', but hasten to reassure anyone who was worried that secretly we were talking about something else that he likes better. But again the theory would be either circular or incorrect, according as we explain consistency in modal terms or in deductive (or purely model-theoretic) terms.

I emphatically do not identify possible worlds in any way with respectable linguistic entities; I take them to be respectable entities in their own right. When I profess realism about possible worlds, I mean to be taken literally. Possible worlds are what they are, and not some other thing. If asked what sort of thing they are, I cannot give the kind of reply my questioner probably expects: that is, a proposal to reduce possible worlds to something else.

I can only ask him to admit that he knows what sort of thing our actual world is, and then explain that other worlds are more things of *that* sort, differing not in kind but only in what goes on at them. Our actual world is only one world among others. We call it alone actual not because it differs in kind from all the rest but because it is the world we inhabit. The inhabitants of other worlds may truly call their own

worlds actual, if they mean by 'actual' what we do; for the meaning we give to 'actual' is such that it refers at any world *i* to that world *i* itself. 'Actual' is indexical, like 'I' or 'here', or 'now': it depends for its reference on the circumstances of utterance, to wit the world where the utterance is located.*

My indexical theory of actuality exactly mirrors a less controversial doctrine about time. Our present time is only one time among others. We call it alone present not because it differs in kind from all the rest, but because it is the time we inhabit. The inhabitants of other times may truly call their own times 'present', if they mean by 'present' what we do; for the meaning we give to 'present' is such that it is indexical, and refers at any time *t* to that time *t* itself.

I have already said that it would gain us nothing to identify possible worlds with sets of sentences (or the like), since we would need the notion of possibility otherwise understood to specify correctly which sets of sentences were to be identified with worlds. Not only would it gain nothing: given that the actual world does not differ in kind from the rest, it would lead to the conclusion that our actual world is a set of sentences. Since I cannot believe that I and all my surroundings are a set of sentences (though I have no argument that they are not), I cannot believe that other worlds are sets of sentences either.

What arguments can be given against realism about possible worlds? I have met with few arguments—incredulous stares are more common. But I shall try to answer those that I have heard.

It is said that realism about possible worlds is false because only our own world, and its contents, actually exist. But of course unactualized possible worlds and their unactualized inhabitants do not *actually* exist. To actually exist is to exist and to be located here at our actual world—at this world that we inhabit. Other worlds than ours are not our world, or inhabitants thereof. It does not follow that realism about possible worlds is false. Realism about unactualized possibles is exactly the thesis that there are more things than actually exist. Either the argument tacitly assumes what it purports to prove, that realism about possibles is false, or it proceeds by equivocation. Our idioms of existential quantification may be used to range over everything without exception, or they may be tacitly restricted in various ways. In particular, they may be restricted to our own world and things in it. Taking them as thus restricted, we can truly say that there exist nothing but our own world and its inhabitants; by removing the restriction we pass illegitimately from that truth to the conclusion that realism about possibles is false. It would be convenient if there were one idiom of

* For more on this theme, see my 'Anselm and Actuality', *Noûs* 4 (1970): 175–188.

quantification, say '*there are* . . .', that was firmly reserved for unrestricted use and another, say '*there actually exist* . . .', that was firmly reserved for the restricted use. Unfortunately, even these two idioms of quantification can be used either way; and thus one can pass indecisively from equivocating on one to equivocating on another. All the same, there are the two uses (unless realism about possibles is false, as has yet to be shown) and we need only keep track of them to see that the argument is fallacious.

Realism about possible worlds might be thought implausible on grounds of parsimony, though this could not be a decisive argument against it. Distinguish two kinds of parsimony, however: qualitative and quantitative. A doctrine is qualitatively parsimonious if it keeps down the number of fundamentally different *kinds* of entity: if it posits sets alone rather than sets and unreduced numbers, or particles alone rather than particles and fields, or bodies alone or spirits alone rather than both bodies and spirits. A doctrine is quantitatively parsimonious if it keeps down the number of instances of the kinds it posits; if it posits 10^{29} electrons rather than 10^{37}, or spirits only for people rather than spirits for all animals. I subscribe to the general view that qualitative parsimony is good in a philosophical or empirical hypothesis; but I recognize no presumption whatever in favor of quantitative parsimony. My realism about possible worlds is merely quantitatively, not qualitatively, unparsimonious. You believe in our actual world already. I ask you to believe in more things of that kind, not in things of some new kind.

Quine has complained that unactualized possibles are disorderly elements, well-nigh incorrigibly involved in mysteries of individuation.* That well may be true of any unactualized possibles who lead double lives, lounging in the doorways of two worlds at once. But I do not believe in any of those. The unactualized possibles I do believe in, confined each to his own world and united only by ties of resemblance to their counterparts elsewhere (see Section 1.9) do not pose any special problems of individuation. At least, they pose only such problems of individuation as might arise within a single world.

Perhaps some who dislike the use of possible worlds in philosophical analysis are bothered not because they think they have reason to doubt the existence of other worlds, but only because they wish to be told more about these supposed entities before they know what to think. How many are there? In what respects do they vary, and what is common to them all? Do they obey a non-trivial law of identity of indiscernibles? Here I am at a disadvantage compared to someone who

* Willard V. Quine, 'On What There Is', in *From a Logical Point of View* (Harvard University Press: Cambridge, Mass., 1953): 4.

pretends as a figure of speech to believe in possible worlds, but really does not. If worlds were creatures of my imagination, I could imagine them to be any way I liked, and I could tell you all you wish to hear simply by carrying on my imaginative creation. But as I believe that there really are other worlds, I am entitled to confess that there is much about them that I do not know, and that I do not know how to find out.

One comes to philosophy already endowed with a stock of opinions. It is not the business of philosophy either to undermine or to justify these preexisting opinions, to any great extent, but only to try to discover ways of expanding them into an orderly system. A metaphysician's analysis of mind is an attempt at systematizing our opinions about mind. It succeeds to the extent that (1) it is systematic, and (2) it respects those of our pre-philosophical opinions to which we are firmly attached. Insofar as it does both better than any alternative we have thought of, we give it credence. There is some give-and-take, but not too much: some of us sometimes change our minds on some points of common opinion, if they conflict irremediably with a doctrine that commands our belief by its systematic beauty and its agreement with more important common opinions.

So it is throughout metaphysics; and so it is with my doctrine of realism about possible worlds. Among my common opinions that philosophy must respect (if it is to deserve credence) are not only my naive belief in tables and chairs, but also my naive belief that these tables and chairs might have been otherwise arranged. Realism about possible worlds is an attempt, the only successful attempt I know of, to systematize these preexisting modal opinions. To the extent that I am modally opinionated, independently of my philosophizing, I can distinguish between alternative versions of realism about possible worlds that conform to my opinions and versions that do not. Because I believe my opinions, I believe that the true version is one of the former. For instance, I believe that there are worlds where physics is different from the physics of our world, but none where logic and arithmetic are different from the logic and arithmetic of our world. This is nothing but the systematic expression of my naive, pre-philosophical opinion that physics could be different, but not logic or arithmetic. I do not know of any non-circular argument that I could give in favor of that opinion; but so long as that *is* my firm opinion nevertheless, I must make a place for it when I do metaphysics. I have no more use for a philosophical doctrine that denies my firm, unjustified modal opinions than I have for one that denies my firm, unjustified belief in chairs and tables.

Unfortunately, though, I am not opinionated enough. There are too many versions of realism about worlds that would serve equally well to systematize my modal opinions. I do not know which to believe; unless

I become more opinionated, or find unsuspected connections between my opinions I may never have any way to choose. But why should I think that I ought to be able to make up my mind on every question about possible worlds, when it seems clear that I may have no way whatever of finding out the answers to other questions about noncontingent matters—for instance, about the infinite cardinals?

Quine has suggested one way to seek fixation of belief about possible worlds by proposing that worlds might be put into correspondence with certain mathematical structures representing the distribution of matter in space and time.* Suppose, for simplicity, that we are concerned with worlds where space-time is Euclidean and four-dimensional, and where there is only one kind of matter and no fields. (Quine calls these *Democritean* worlds.) We can represent any such world by a mapping from all quadruples $\langle x, y, z, t \rangle$ of real numbers to the numbers 0 and 1. We are to think of the quadruples as coordinates, in some coordinate system, of space-time points; and we are to think of the quadruples mapped onto 0 as coordinates of points unoccupied by matter, and of quadruples mapped onto 1 as coordinates of points occupied by matter. Thus the entire mapping represents a possible distribution of uniform matter over Euclidean space-time. Since there are many different coordinate systems—differing in the location of the $\langle 0, 0, 0, 0 \rangle$ point, the length of the units of spatial and of temporal distance, and the directions of the spatial axes—there are many different mappings (differing by a transformation of coordinates) that we regard as representing the same distribution of matter. To overcome this dependence of the mapping on an arbitrary choice of coordinates, we take not the mappings themselves, but equivalence classes of mappings under transformations of coordinates. We get a perfectly well-defined, well-understood set of mathematical entities, exactly one for every different possible distribution of matter.

Of course, this is a simplified example. The construction must be generalized in several ways to cover possibilities so far overlooked. Space-time might be non-Euclidean; there might be scalar, vector, or tensor fields independent of the distribution of matter; there might be more than one kind of matter, or more or less density of matter, even in the small. We would have to go on generalizing as long as we could think of possibilities not yet taken into account. But generalizing Quine's simplified example is easy mathematical work. We can hope that soon we will reach the end of the generalizations required and permitted by our opinions about what is possible, and then we will have a well-defined set of mathematical entities of a familiar and

* Willard V. Quine, 'Propositional Objects', in *Ontological Relativity* (Columbia University Press: New York, 1969): 147–155.

well-understood sort, corresponding one-to-one in a specified way with the possible worlds.

I do not, of course, claim that these complicated mathematical entities *are* the possible worlds. I cannot believe (though I do not know why not) that our own world is a purely mathematical entity. Since I do not believe that other worlds are different in kind from ours, I do not believe that they are either. What is interesting is not the reduction of worlds to mathematical entities, but rather the claim that the possible worlds stand in a certain one-to-one correspondence with certain mathematical entities. Call these *ersatz possible worlds*. Any credible correspondence claim would give us an excellent grip on the real possible worlds by their ersatz handles. It would answer most of our questions about what the possible worlds are like.*

We already have a good grip, in this way, on at least *some* of the possible worlds: those that correspond to mathematical ersatz worlds constructed at the highest level of generality that our modal opinions clearly require and permit. It is only because there may be higher levels of generality that we have failed to think of, and because our modal opinions are indecisive about whether there really are possibilities corresponding to some of the levels of generality we have thought of (what about letting the number of spatial dimensions vary? what about letting there be entities that are temporally but not spatially located? what about letting the distinction between space and time be local rather than global, like the distinction between up and down?), that we fail to have a good grip on all the worlds.

The mathematical construction of ersatz worlds may seem to depend too much on our current knowledge of physics. We know that we must generalize enough to include non-Euclidean worlds, for instance, just because the physicists have found reason to believe that we live in one. But physics is contingent. If we look to physics to tell us what is possible, will we get all possible worlds? Or only the physically possible worlds, according to current physics?

More, at least, than the latter. We will certainly construct ersatz worlds that disobey currently accepted physical laws; for instance,

* Even the indefinite correspondence claim that *some* generalization of Quine's simplified example is right is enough to answer one important question about the possible worlds. How many are there? Answer: at least $\beth_2$, the infinite cardinal of the set of all subsets of the real numbers. It can easily be shown that this is the number of ersatz worlds in Quine's original construction. Indeed, it is the number of ersatz worlds at any level of generality that seems to me clearly called for. Here is another reason why possible worlds are not sets of sentences of a language. If we take 'language' at all literally, so that sentences are finite strings over a finite alphabet, there are not enough sets of sentences to go around. There are at most $\beth_1$, the infinite cardinal of the set of all real numbers.

ersatz worlds where mass-energy is not conserved. Still, we cannot be sure of getting all possible worlds, since we cannot be sure that we have constructed our ersatz worlds at a high enough level of generality. If we knew only the physics of 1871, we would fail to cover some of the possibilities that we recognize today. Perhaps we fail today to cover possibilities that will be recognized in 2071. Our modal opinions do change, and physicists do a lot to change them. But this is *not* to say that we can argue from the contingent results of empirical investigation to conclusions about what possibilities there are. It is only to say that when we find it hard to locate our actual world among the possibilities that we recognize, we may reasonably be stimulated to reconsider our modal opinions. We may try to think of credible possibilities hitherto overlooked, and we may consider whether we are still as sure as we were about those of our modal opinions that have turned out to be restrictive. It is this reconsideration of modal opinions that may influence our construction of ersatz worlds, not the results of empirical investigation itself. We are concerned not with physics proper, but with the preliminary metaphysics done by physicists.

4.2 Similarity

It may be said that even if possible worlds are tolerable, still the notion of comparative overall similarity of worlds is hopelessly unclear, and so no fit foundation for the clarification of counterfactuals or anything else. I think the objection is wrong. 'Unclear' is unclear: does it mean 'ill-understood' or does it mean 'vague'? Ill-understood notions are bad primitives because an analysis by means of them will be an ill-understood analysis. (It may yet be better than no analysis at all.) But comparative similarity is not ill-understood. It is vague—very vague—in a well-understood way. Therefore it is just the sort of primitive that we must use to give a correct analysis of something that is itself undeniably vague.

Overall similarity consists of innumerable similarities and differences in innumerable respects of comparison, balanced against each other according to the relative importances we attach to those respects of comparison. Insofar as these relative importances differ from one person to another, or differ from one occasion to another, or are indeterminate even for a single person on a single occasion, so far is comparative similarity indeterminate. As Goodman says,* 'Importance is a highly

* Nelson Goodman, 'Seven Strictures on Similarity', in L. Foster and J. W. Swanson, *Experience and Theory* (University of Massachusetts Press: 1970): 27.