

## *Possible Worlds*

ROBERT C. STALNAKER

CORNELL UNIVERSITY

According to Leibniz, the universe—the actual world—is one of an infinite number of possible worlds existing in the mind of God. God created the universe by actualizing one of these possible worlds—the best one. It is a striking image, this picture of an infinite swarm of total universes, each by its natural inclination for existence striving for a position that can be occupied by only one, with God, in his infinite wisdom and benevolence, settling the competition by selecting the most worthy candidate. But in these enlightened times, we find it difficult to take this metaphysical myth any more seriously than the other less abstract creation stories told by our primitive ancestors. Even the more recent expurgated versions of the story, leaving out God and the notoriously chauvinistic thesis that our world is better than all the rest, are generally regarded, at best, as fanciful metaphors for a more sober reality. J. L. Mackie, for example, writes "... talk of possible worlds ... cries out for further analysis. There *are* no possible worlds except the actual one; so what are we up to when we talk about them?" ([3]: 90). Lawrence Powers puts the point more bluntly: "The whole idea of possible worlds (perhaps laid out in space like raisins in a pudding) seems ludicrous" ([4]).

These expressions of skepticism and calls for further analysis are of course not directed at Leibniz but at recent uses of parts of his metaphysical myth to motivate and give content to formal semantics for modal logics. In both formal and philosophical discussions of modality, the concept of a possible world has shown itself to have considerable heuristic power. But, critics have argued, a heuristic device should not be

confused with an explanation. If analyses of modal concepts (or of the concept of a proposition) in terms of possible worlds are to be more than heuristic aids in mapping the relationships among the formulae of a modal logic, the concept of a possible world itself must be explained and justified.

Although it is commonly taken to be an obvious truth that there really are no such things as possible worlds—that the myth, whether illuminating or misleading, explanatory or obfuscating, is nevertheless a myth—this common opinion can be challenged. That is, one might respond to the possible worlds skeptic not by explaining the metaphor but by taking the story to be the literal truth. David Lewis responds in this way, and he cites common opinion and ordinary language on his side:

I believe 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 have been 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 that 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 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 which might be called 'ways things could have been'. I prefer to call them 'possible worlds'. ([2]: 84.)

Lewis does not intend this as a knockdown argument. It is only a presumption that the sentences of ordinary language be taken at face value, and the presumption can be defeated if the naive reading of the sentences leads to problems which can be avoided by an alternative analysis. The aim of the argument is to shift the burden to the skeptic who, if he is to defeat the argument, must point to the problems which commitment to possible worlds creates, and the alternative analysis which avoids those problems. Lewis does not think the skeptic can do either.

The rhetorical force of Lewis's argument is in the suggestion that possible worlds are really not such alien entities as the metaphysical flavor of this name seems to imply. The argument suggests not that ordinary language and our common beliefs commit us to a weighty metaphysical theory, but rather that what appears to be a weighty metaphysical theory is really just some ordinary beliefs by another name. Believing in

possible worlds is like speaking prose. We have been doing it all our lives.

But for this to be convincing, the shift from “ways things might have been” to “possible worlds” must be an innocent terminological substitution, and I do not believe that, as Lewis develops the concept of a possible world, it is. To argue this point I will state four theses about possible worlds, all defended by Lewis. Together they constitute a doctrine which I will call extreme realism about possible worlds. It is this doctrine against which the skeptic is reacting, and against which, I shall argue, he is justified in reacting. I believe the doctrine is false, but I also believe that one need not accept or reject the theses as a package. The main burden of my argument will be to show the independence of the more plausible parts of the package, and so to defend the coherence of a more moderate form of realism about possible worlds—one that might be justified by our common modal opinions and defended as a foundation for a theory about the activities of rational agents.

Here are Lewis’s four theses:

- (1) *Possible worlds exist.* Other possible worlds are just as real as the actual world. They may not *actually* exist, since to actually exist is to exist in the actual world, but they do, nevertheless, exist.
- (2) *Other possible worlds are things of the same sort as the actual world—“I and all my surroundings”* ([2]: 86). They differ “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” ([2]: 85).
- (3) *The indexical analysis of the adjective ‘actual’ is the correct analysis.* “The inhabitants of other worlds may truly call their own world 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” ([2]: 85-6).
- (4) *Possible worlds cannot be reduced to something more basic.* “Possible worlds are what they are and not

another thing.” It would be a mistake to identify them with some allegedly more respectable entity, for example a set of sentences of some language. Possible worlds are “respectable entities in their own right” ([2]: 85).

The first thesis, by itself, is compatible with Lewis’s soothing claim that believing in possible worlds is doing no more than believing that things might have been different in various ways. What is claimed to exist are things which ordinary language calls “ways things might have been”, things that truth is defined relative to, things that our modal idioms may be understood as quantifiers over. But the first thesis says nothing about the nature of the entities that play these roles. It is the second thesis which gives realism about possible worlds its metaphysical bite, since it implies that possible worlds are not shadowy ways things could be, but concrete particulars, or at least entities which are made up of concrete particulars and events. The actual world is “I and my surroundings”. Other possible worlds are more things like that. Even a philosopher who had no qualms about abstract objects like numbers, properties, states and kinds might balk at this proliferation of fullblooded universes which seem less real to us than our own only because we have never been there.

The argument Lewis gives for thesis one, identifying possible worlds with ways things might have been, seems even to be incompatible with his explanation of possible worlds as more things of the same kind as I and all my surroundings. If possible worlds are ways things might have been, then the actual world ought to be *the way things are* rather than *I and all my surroundings*. *The way things are* is a property or a state of the world, not the world itself. The statement that the world is the way it is is true in a sense, but not when read as an identity statement (Compare: “the way the world is is the world”). This is important, since if properties can exist uninstantiated, then *the way the world is* could exist even if a world that is that way did not. One could accept thesis one—that there really are many ways that things could have been—while denying that there exists anything else that is like the actual world.

Does the force of thesis two rest, then, on a simple equivocation between “the actual world”, in the sense that is roughly captured in the paraphrase “I and all my surroundings”,

and the sense in which it is equivalent to "the way things are"? In part, I think, but it also has a deeper motivation. One might argue from thesis three—the indexical analysis of actuality—to the conclusion that the essential difference between our world and the others is that we are here, and not there.

Thesis three seems to imply that the actuality of the actual world—the attribute in virtue of which it is actual—is a world-relative attribute. It is an attribute which our world has relative to itself, but which all the other worlds have relative to themselves too; so the *concept* of actuality does not distinguish, from an absolute standpoint, the actual world from the others. But if there is no absolute property of actuality, does this not mean that, looking at things from an objective point of view, merely possible people and their surroundings are just as real as we and ours?

The mistake in this reasoning, I think, is in the assumption that the absolute standpoint is a neutral one, distinct from the view from within any possible world. The problem is avoided when one recognizes that the standpoint of the actual world is the absolute standpoint, and that it is part of the concept of actuality that this should be so. We can grant that fictional characters are as right, from their point of view, to affirm their fullblooded reality as we are to affirm ours. But their point of view is fictional, and so what is right from it makes no difference as far as reality is concerned.

My point is that the *semantical* thesis that the indexical analysis of "actual" is correct can be separated from the metaphysical thesis that the actuality of the actual world is nothing more than a relation between it and things existing in it. Just as one could accept the indexical analysis of personal pronouns and be a solipsist, and accept the indexical analysis of tenses and believe that the past exists only as memory and the future only as anticipation, one can accept the indexical analysis of actuality while excluding from one's ontology any universes that *are* the way things might have been.

In fact, I want to argue, one must exclude those analogues of our universe from one's ontology. The thesis that the actual world alone is real is superficially analogous to solipsism—the thesis that I alone am real, but solipsism has content, and can be coherently denied, because it says something substantive about what alone is real. In effect, solipsism says that the actual world is a person, or a mind. But the thesis that the actual world alone

is real has content only if “the actual world” means something other than the totality of everything there is, and I do not believe that it does. The thesis that there is no room in reality for other things than the actual world is not, like solipsism, based on a restrictive theory of what there is room for in reality, but rather on the metaphysically neutral belief that “the actual world” is just another name for reality.

So the moderate realism whose coherence I am trying to defend accepts theses one and three, and rejects thesis two. What about thesis four? If we identify possible worlds with ways things might have been, can we still hold that they are “respectable entities in their own right”, irreducible to anything more fundamental? Robert Adams has argued that to avoid extreme realism we must find an eliminative reduction of possible worlds. “If there are any true statements in which there are said to be non-actual possible worlds,” he argues, “they must be reducible to statements in which the only things there are said to be are things which are in the actual world, and which are not identical with non-actual possibles” ([1]: 224). Unless the reminder that by “possible world” we mean nothing more than “way things might have been” counts as such a reduction, I do not see why this should be necessary. Why cannot *ways things might have been* be elements of the actual world, as they are?

Two problems need to be separated: the first is the general worry that the notion of a possible world is a very obscure notion. How can explanations in terms of possible worlds help us to understand anything unless we are told what possible worlds are, and told in terms which are independent of the notions which possible worlds are intended to explain? The second problem is the specific problem that believing in possible worlds and in the indexical analysis of actuality seems to commit one to extreme realism, which (many believe) is obviously false. Now to point to the difference between a way our world might have been and a world which *is* the way our world might have been, and to make clear that the possible worlds whose existence the theory is committed to are the former kind of thing and not the latter, is to do nothing to solve the first problem; in fact it makes it more acute since it uses a modal operator to say what a possible world is. But this simple distinction does, I think, dissolve the second problem which was the motivation for Adams’s demand for an analysis.

Not only is an eliminative reduction of possible worlds not necessary to solve the second problem, it also may not be sufficient to solve the first. I shall argue that the particular reduction that Adams proposes—a reduction of possible worlds to propositions—by itself says nothing that answers the critic who finds the concept of a possible world obscure. His reduction says no more, and in fact says less, about propositions and possible worlds than the reverse analysis that I would defend—the analysis of propositions in terms of possible worlds.

Adam's analysis is this: "Let us say that a *world-story* is a maximal consistent set of propositions. That is, it is a set which has as its members one member of every pair of mutually contradictory propositions, and which is such that it is possible for all of its members to be true together. The notion of a possible world can be given a contextual analysis in terms of world-stories" ([1]: 225). For a proposition to be true in some or all possible worlds is for it to be a member of some or all world-stories. Other statements that seem to be about possible worlds are to be replaced in a similar way by statements about world-stories.

There are three undefined notions used in Adams's reduction of possible worlds: *proposition*, *possibility*, and *contradictory*. What are propositions? Adams leaves this question open for further discussion; he suggests that it might be answered in various ways. Little is said about them except that they are to be thought of as language independent abstract objects, presumably the potential objects of speech acts and propositional attitudes.

What is possibility? The notion used in the definition of world-story is a property of *sets* of propositions. Intuitively, a set of propositions is possible if all its members can be true together. This notion cannot, of course, be defined in terms of possible worlds, or world-stories, without circularity, but it should be a consequence of the theory that a set of propositions is possible if and only if its members are simultaneously true in some possible world (are all members of some world-story). Presumably, an explicit formulation of the world-story theory would contain postulates sufficient to ensure this.

What is a contradictory? This relation between propositions might be defined in terms of possibility as follows: *A* and *B* are contradictories if and only if  $\{A, B\}$  is not possible, and for every possible set of propositions  $\Gamma$  either  $\Gamma \cup \{A\}$  or  $\Gamma \cup \{B\}$  is

possible. The theory tacitly assumes that every proposition has a contradictory; in an explicit formulation, this would be an additional postulate.

These definitions and postulates yield a minimal world-story theory. It is minimal in that it imposes no structure on the basic elements of the theory except what is required to justify what Adams calls the "intuitively very plausible thesis that possibility is holistic rather than atomistic, in the sense that what is possible is possible only as part of a possible completely determinate world" ([1]: 225). But the theory justifies this thesis only by postulating it.

It will be useful to compare this reduction of possible worlds to propositions with the competing reduction of propositions to possible worlds. What is at stake in choosing which of these two notions to define in terms of the other? Adams refers to the "not unfamiliar trade-off between non-actual possibles and intensions (such as propositions); given either, we may be able to construct the other, or do the work that was supposed to be done by talking about the other" ([1]: 228). But the two proposals are not equivalent. Part of what distinguishes them is an elusive question of conceptual priority, but there are also more substantive differences, both in the structure imposed on propositions and possible worlds and in the questions left to be answered by further developments of the respective theories.

If we set aside questions of conceptual priority—of which concepts and principles should be primitive and which defined or derived—what is the difference between the two analyses? The world-story theory is weaker, leaving open questions which are settled by the possible worlds analysis of propositions. The following two theses are consequences of the possible worlds analysis, but not of the world-story theory; the first concerns identity conditions; the second is a closure condition:

- (I) Necessarily equivalent propositions are identical.
- (C) For every set of propositions, there is a proposition which, necessarily, is true if and only if every member of the set is true.

Are these consequences of the possible worlds analysis welcome or not? I believe that thesis (I) can be defended independently of the possible worlds analysis of propositions, but that is a long



story for another occasion. The thesis does have some notoriously problematic consequences, but I believe, first, that it is implied by a widely held and plausible assumption about the nature of propositional attitudes—the assumption that attitudes like belief and desire are dispositions of agents displayed in their rational behavior—and second, that the apparently paradoxical consequences of the thesis can be explained away. But for now let me just point out that the possible worlds analysis has this substantive consequence, and leave the part of my argument which depends on it conditional on the assumption that this consequence is welcome. The thesis is not implied by the minimal world-story theory, but it is compatible with it, so the world-story theorist who agrees with me about thesis (I) can add it to his theory as an additional postulate.

Thesis (C) seems reasonable on almost any theory of propositions and propositional attitudes. Whatever propositions are, if there are propositions at all then there are sets of them, and for any set of propositions, it is something determinately true or false that all the members of the set are true. If one is willing to talk of propositions at all, one will surely conclude that that something is a proposition. It may not be possible to express all such propositions since it may not be possible, in any actual language, to refer to all such sets; it may not be humanly possible to believe or disbelieve some such propositions, since it may not be humanly possible to grasp them. But if this is so, it is surely a contingent human limitation which should not restrict the range of potential objects of propositional attitudes. So I will assume that the world-story theorist will want to add thesis (C) to his theory.

If (I) and (C) are added as postulates to the minimal world-story theory, then it becomes equivalent to the possible worlds analysis with respect to the structure it imposes on the set of propositions, and on the relation between propositions and possible worlds. The sole difference that remains between the two theories is that one takes as primitive what the other defines. And even this difference will be eliminated if we make one more change in response to a question about the further development of the world-story theory.

The next question for the world-story theorist is this: can he say more about his fundamental concept, the concept of a proposition? In particular, are there some *basic* propositions out of which all the rest can be constructed? The usual way to

answer this question is to model basic propositions on the atomic sentences of a first order language; propositions are constructed out of individuals and primitive properties and relations in the same way that sentences are constructed out of names and predicates. But this strategy requires building further structure into the theory. There is another way to answer the question which needs no further assumption. We can deduce from what has already been built into the world-story theory that there is a set of propositions of which all propositions are truth-functions: this is the set of strongest contingent propositions—those propositions which are members of just one world-story. It is thus a harmless change, a matter of giving the theory a more economical formulation, to take these to be the basic propositions. (This change does not foreclose a further reduction of what are here called basic propositions. Any alternative reduction could be expressed as a further reduction; this is why the move is harmless.) We can then define propositions generally as sets of basic propositions (or, for a neater formulation, call the basic elements *propositional elements* and let their unit sets be the basic propositions.) A non-basic proposition will be true just in case one of its members is true. This reduction has the added advantage that it allows us to define the previously primitive property of possibility, and to derive all of the postulates. With these primitive notions and assumptions eliminated, the world-story theory looks as good as the theory that takes possible worlds as primitive and defines propositions. This is, of course, because it is exactly the same theory.

I have gone through this exercise of changing the world-story theory into the possible worlds analysis of propositions in order to make the following point: first, the minimal world-story theory with which I began is indeed a minimal theory of propositions, a theory that assumes nothing about them except that they have truth values and are related to each other by the standard propositional relations (entailment, compatibility, and so forth). But second, every step in the metamorphosis of this minimal theory into the possible worlds analysis is motivated by independently plausible assumptions about propositions or by theory-neutral considerations of economy of formulation. If this is right, then the possible worlds analysis is not just one theory which makes the assumptions about propositions that I have made. More than this, it is the whole content of that

analysis to impose the minimal structure on propositions which is appropriate to a theory which understands propositions in this way. Anyone who believes that there are objects of propositional attitudes, and who accepts the assumptions about the formal properties of the set of these objects, must accept that there are things which have all the properties that the possible worlds theory attributes to possible worlds, and that propositions can be reduced to these things.

Is the form of realism about possible worlds that I want to defend really realism? It is in the sense that it claims that the concept of a possible world is a basic concept in a true account of the way we represent the world in our propositional acts and attitudes. A full defense of this kind of realism would require a development and defense of such an account. All I have tried to do here is show that there is a coherent thesis about possible worlds which rejects extreme realism, but which takes possible worlds seriously as irreducible entities, a thesis that treats possible worlds as more than a convenient myth or a notational shortcut, but less than universes that resemble our own.<sup>1</sup>

#### REFERENCES

- [1] Robert Merrihew Adams, "Theories of Actuality" *NOÛS* 8(1974): 211-31.
- [2] David Lewis, *Counterfactuals* (Cambridge, Mass.: Harvard University Press, 1973).
- [3] J. L. Mackie, *Truth, Probability and Paradox* (Oxford: Clarendon Press, 1973).
- [4] Lawrence Powers, COMMENTS on R. Stalnaker, "Propositions" (forthcoming in the proceedings of the 1971 Oberlin Philosophy Colloquium).

#### NOTE

<sup>1</sup>I am indebted to the John Simon Guggenheim Memorial Foundation for support during the time when this paper was written.