INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI®
800-521-0600

IMPOSSIBLE WORLDS

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

David A. Vander Laan, B.A.

Alvin Plantinga, Director

Department of Philosophy
Notre Dame, Indiana

December 1999

UMI Number: 9950668

UMI Microform 9950668

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

IMPOSSIBLE WORLDS

Abstract

by

David A. Vander Laan

The theory of possible worlds has permeated analytic philosophy in recent decades, and its best versions have a consequence which has gone largely unnoticed: in addition to the panoply of possible worlds, there are a great many impossible worlds. A uniform ontological method alone should bring the friends of possible worlds to adopt impossible worlds, I argue, but the theory's applications also provide strong incentives. In particular, the theory facilitates an account of counterfactuals which avoids several of the implausible results of David Lewis's account, and it paves the way for the analogues of Kripkean semantics for epistemic and relevant logics.

On the theories of possible worlds as abstract objects, worlds bear a strong resemblance to propositions. I contend that if there are distinct necessarily false propositions, then there are likewise distinct impossible worlds. However, one who regards possible worlds as concrete objects must not recognize impossible worlds, in part because concrete worlds cannot misrepresent certain features of reality (the plurality of worlds, for example), as some impossible worlds must. Accordingly, I defend and develop a theory of impossible worlds as (abstract) maximal impossible states of affairs.

Impossible worlds perform admirably in the analysis of counterfactuals with impossible antecedents. I argue that, contrary to standard accounts, not all counterpossibles are trivially true, and I develop a Lewis-style semantics

which allows this result. The point is crucial, since many views presuppose that some counterpossibles are substantive philosophical truths.

Finally, I show that impossible worlds hold great promise for doxastic and relevant logics. Epistemic logic needs a domain of propositions which is not closed under strict implication to avoid the problem of logical omniscience, and relevant logic needs such a domain to avoid the famous paradoxes of implication.

In sum, impossible world theory promises natural, elegant solutions to philosophical problems in numerous areas where possible worlds alone flounder. These solutions come to most possible world theorists at no cost, since the existence of impossible worlds is entailed by theses they already hold.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I owe a great deal of thanks to

Al Plantinga, for patient guidance, for ushering my attention toward the big picture, and for unflagging cheer;

Peter van Inwagen, for philosophical inspiration and valuable words of encouragement;

Michael Kremer, for extremely helpful comments which resulted in significant improvement and expansion of the final chapter;

Dean Zimmerman, for eagerly serving as a reader, and for tireless help with the job search;

Tom Flint, for encouraging and helpful comments in the early stages;

Paddy Blanchette, for her characteristically thorough comments on the material of chapter four;

Michael Loux, for planting the idea that my term paper on impossible worlds might grow into a dissertation;

Mic Detlefsen, for alerting me to the Notre Dame Journal of Formal Logic's special issue on impossible worlds, and for kindly inviting a submission to the issue;

Graham Priest, for his interest in the piece I submitted;

David K. Lewis, to whom my philosophical debt will be obvious;

Tom Crisp, Matt Davidson, and especially Mike Thrush, for extensive and penetrating discussion of issues raised in several chapters; and

Anthony Everett, Brian Leftow, and Del Ratzsch whose helpful discussions are noted in the text.

INTRODUCTION

It scarcely need be remarked that the contemporary theory of possible worlds has been tremendously influential and fruitful in philosophy in recent decades. This dissertation proposes an extension of that important theory. The extension--the theory of impossible worlds--is, I think, a useful one and one which flows quite naturally from the most sensible views of possible worlds. In fact, the existence of impossible worlds is an overlooked consequence of those views (given some reasonable premises about the nature of states of affairs).

Lest anyone be unduly suspicious of the topic, let me point out that the theory of impossible worlds to be presented is not Meinongian: it does not say that there are objects which do not or could not exist. Nor does my theory earn the incredulous stares that David Lewis's concretist theory of possible worlds tends to attract. Nor does it assert that there are true contradictions. Rather, nearly anyone who is inclined to accept the prevalent, abstractionist approach to possible worlds should find impossible worlds perfectly congenial.

The aim of the first chapter is to show that the sorts of reasons that we have for thinking that there are possible worlds provide us with equally good reasons for thinking that there are impossible worlds. Along the way we will examine the notion of content as it relates to states of affairs.

Objections will be dealt with in the second chapter. Most of these can be defused without great difficulty. The only one which poses any serious threat to impossible worlds, I conclude, is equally problematic for possible worlds.

The third chapter explores the nature of impossible worlds. It attempts to say what can be said about what impossible worlds are like and how they

are to be individuated. The chapter features a whirlwind tour of "illogical space," highlighting some impossible worlds with fascinating formal properties (worlds, for example, such that no two propositions true in them are inconsistent with each other).

My principal reason for arguing that there are impossible worlds is that it seems to me there really are such things, and it seems to me that those who hold the dominant view of possible worlds should think that there are such things. My motivation has less to do with potential philosophical application of impossible worlds than with facts about what follows from a certain widely-held theory. I want to make it clear, however, that I do think that impossible worlds can function as useful philosophical tools in a number of different contexts, some of which I discuss in the last two chapters.

An important application of impossible worlds--perhaps the most important application--is in the semantics for counterfactuals. The fourth chapter presents a modified version of David Lewis's semantics which gives much more intuitive truth conditions to counterfactuals with impossible antecedents and which preserves, without exception, the intuitive interdefinability of 'would' and 'might' counterfactuals.

The final chapter sketches some applications of impossible worlds in the semantics of epistemic and relevant logic. The main point is that impossible worlds are ideally suited to handle the "problem of logical omniscience" (in epistemic logic) and the "paradoxes of implication" (in relevant logic). These are both areas where possible world analyses tend to founder, and where impossible worlds step in in a very natural way.

In the *Notre Dame Journal of Formal Logic*'s recent special issue on impossible worlds, guest editor Graham Priest comments, "the notion of an impossible world is coming to play a role in the theorization and unification of a number of issues in philosophical logic similar to that which the notion of a

possible world itself did some twenty-five years ago. ... My prediction, for what it is worth, is that the debate concerning them will go the same way and for exactly the same reasons...."[1] My hope is that Priest is right, and that this dissertation may help bring impossible worlds the philosophical attention they deserve.

Some readers will have a higher tolerance for unapplied metaphysics than others. For those that have a relatively high tolerance, I recommend that you read the chapters in the order in which they are presented. For those who need to see the cash value of impossible worlds before they are willing or able to listen to metaphysical arguments, I recommend that you read chapters 4 and 5 first and afterward read the earlier chapters. Chapters 4 and 5 do at times presuppose some of the arguments and nomenclature of the earlier chapters, particularly of chapter 3, but most of the material will be independently intelligible and readers will easily grasp the main points of those chapters without the help of the earlier ones.

[1] p. 487.

CHAPTER 1: THE EXISTENCE OF IMPOSSIBLE WORLDS

I want to argue for the existence of impossible worlds, though I suspect that many or most of those who believe in possible worlds already believe in impossible worlds, whether or not they have thought to call them by that name. On the view of possible worlds that I prefer, there are states of affairs[2]. Some of these, like *an apple's being colored if red* and *California's being populous*, are actual; these states of affairs obtain. Others, like *the Axis powers' having won World War II*, are merely possible; they do not obtain, but it is possible that they obtain. Still others, like *Paul's having squared the circle* or *the number 9's being a Caesar salad*, are impossible; they do not possibly obtain. Impossible worlds constitute a subclass of this last class of states of affairs, namely, the impossible states of affairs which are maximal (in a sense we shall discuss later).

Some philosophers, Roderick Chisholm, for example, prefer to identify states of affairs with propositions. If these philosophers are correct, then possible worlds are propositions which are possibly true and maximal. In this case impossible worlds are the necessarily false propositions which are maximal in the requisite sense.[3]

For many philosophers, then, very little needs to be said to establish the existence of impossible worlds; it is simply a matter of pointing out which objects are deserving of the title. Nevertheless, it may be worthwhile to review

[2] I think these are the same as the situations referred to, e.g., by John Perry in "From Worlds to Situations."

[3] I shall assume that states of affairs are not propositions. However the position that states of affairs are propositions is not necessarily inimical to the central claims of this dissertation and, indeed, would simplify much of the argumentation that follows.

the principal arguments that have been offered for the existence of possible worlds, for in each case there are remarkably similar parallel arguments for impossible worlds. Both of the arguments that I will consider have been put forward by David Lewis. I will call them the argument from ways and the argument from utility.

The Argument from Ways

I believe that, besides the wide variety of possible worlds, there are impossible worlds as well. If an argument is wanted, it is this. It is uncontroversially true that certain things could not have been otherwise than they are. I believe, and so do you, that things could not have been different in countless ways. But what does this mean? Ordinary language permits the paraphrase: there are many ways things could not have been. On the face of it, this is an existential quantification. It says that there exist many entities of a certain description, to wit 'ways things could not have been'. Taking this statement at face value, there exist entities that might be called 'ways things could not have been'. In keeping with the terminology used for their possible analogs, I prefer to call them 'impossible worlds'.

The above argument is an adapted quotation of David Lewis's argument for possible worlds in Counterfactuals,[4,5] and I think the argument is a good one. (Strictly, though, both Lewis's argument and my adaptation are arguments for the existence of certain states of affairs, not merely for the worlds, which must meet a maximality requirement of some sort.[6]) We *do* speak of ways things couldn't be; told that some object is black and white and red all over, for

[4] p.84. As we shall see, Lewis himself does not endorse impossible worlds.
[5] Since writing this section I have learned that Margery Bedford Naylor gives the same argument in "A Note on David Lewis's Realism About Possible Worlds."
[6] Since according to Lewis possible worlds are concrete rather than abstract objects, he takes the argument from ways to establish something quite different than what I think it establishes. Despite Lewis's position, I will treat the argument from ways as an argument for states of affairs. I will treat the argument from utility as an argument for possible worlds, postponing the further question whether possible worlds are abstract or concrete.

example, we might say, "It couldn't be *that* way!" One could object that ordinary language is ambivalent about this manner of speaking, since we might just as well respond, "There is *no* way things could be like *that*!" But the ambivalence of ordinary language is not a reason to reject the argument. We make the same sort of remark with respect to logically possible situations: "There is no way I can make it to the church on time" when the only obstacle is my being ten miles away two minutes before the deadline. In both cases, the denial of the existential quantification is implicitly qualified. There is no way I can make it to the church on time; that is, no way compatible with all of my circumstances, and the laws of physics, metaphysics, and logic that I can make it to the church on time. There is no way any object can be black and white and red all over; that is, no way compatible with certain necessary truths about color and the laws of logic, no *possible* way. But there is a way things could *not* be, such as a situation in which white is a texture and red is a flavor. Ordinary language permits this way of speaking, just as it permits talk of ways things could be. And if language permits it, we have the *prima facie* existential quantification we need.

Is there any reason to take the *prima facie* existential quantification at face value in the argument for possible states of affairs but not in the argument for impossible states of affairs? As Lewis says,

> I do not make it an inviolable principle to take seeming existential quantifications in ordinary language at 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 in some other way is known not to (84).

So far as I can tell, the thesis that there are impossible worlds does not lead to trouble. The thesis is surprising (or at any rate people are sometimes surprised by it), world-talk having traded exclusively in possible worlds for so long. But none of the objections that one might initially be inclined to array against the

thesis strike me as being very powerful upon consideration. Certainly no obvious objection is so clear and compelling as to justify the dismissive attitude toward impossible worlds one occasionally finds. The objections themselves will be taken up in the next chapter, after the thesis has been put forth in greater detail. For now, let us keep in mind that we have a presumption in favor of impossible worlds that objections will need to be strong enough to overcome if they are to be successful.

Are there other ways of taking our ordinary discourse that are known not to lead to trouble? Lewis thinks so; earlier in Counterfactuals[7] he says (without argument) that this part of our modal discourse may be founded on a confused fantasy. But suppose that talk about ways things could be is the literal truth of the matter.[8] Then, I think, rejection of impossible worlds does lead to trouble. Why should one think that talk about ways things couldn't be is founded on a confused fantasy? There would be something rather arbitrary about excluding from one's ontology those worlds which are metaphysically impossible. Why not also exclude the nomologically impossible worlds? Why not exclude the unlikely worlds, or better yet, the nonactual worlds? The fact that impossible worlds cannot possibly obtain provides no ground whatsoever for supposing that *this* part of our modal discourse is less truthful than the rest.

There is a strong analogy between states of affairs and propositions. Both states of affairs and propositions are, in some sense, representational. Both states of affairs and propositions "describe" things as being in a certain way. Both may be accurate or inaccurate. (It is true that some--David Armstrong, for example--say states of affairs are not representational, but the objects I call "states of affairs" are not the objects Armstrong calls "states of

[7] p. 24.

[8] Of course this is a disputable supposition. Many--e.g. nominalists--will insist that possible worlds talk itself is founded on a confused fantasy. I disagree, but I do not provide arguments here. My principal audience consists of those who are already willing to countenance possible worlds, and my aim is to point out some consequences of that position.

affairs." The issue is whether *his* states of affairs are representational is not germane at present.)

Now should we exclude representations of the impossible from our ontologies? Necessarily false propositions--impossible propositions, we might say--are not usually regarded as ontologically suspect (at any rate, no more suspect than other propositions). It is no strike against them that they cannot possibly be true. Why should it count against a state of affairs that it cannot possibly obtain? What could it be about this kind of representation what weighs against its existence? How is the metaphysical possibility of obtaining even a relevant consideration?

I have claimed that Lewis's argument from ways has an equally good parallel argument for impossible worlds. But is Lewis's argument a good one? Let's look at it a bit more closely.

One feature of the argument from ways may be highlighted by an objection to its parallel. Mark Sharlow says one who accepts

(N2) Things could not have been different in countless ways

need not also accept

> (N3) There are many ways things could not have been besides the way they are.[9]

The move is invalid, he says, since the former is simply the claim that there are countless necessary propositions, and this truth can be analyzed without impossible worlds. What Sharlow denies, in effect, is that latter claim (taken with all "ontological seriousness") is really a paraphrase of the former, as Lewis says.

There is a kind of point here, namely, that the believer in possible worlds is not compelled to accept the intended reading of (N3) simply in virtue of the truth conditions of (N2). However this objection seems to manifest a

[9] "Lewis's Modal Realism: A Reply to Naylor". The labels are Sharlow's.

misunderstanding of both Lewis's argument from ways and its parallel. Lewis does not assert that

(L2) Things could have been different in countless ways

entails

(L3) There are many ways things could have been besides the way they actually are

taken as a genuine existence claim. He only says that the sentence (L3) is a "permissible paraphrase" of (L2). Then, since even a *prima facie* existential quantification like (L3) might be taken as something other than an existential quantification, Lewis commends to us the face value of (L3). This interpretation is not inevitable. Lewis specifically points out that he does not always take such sentences at face value. The objection that the move from (L2) to (L3) (or from (N2) to (N3)) is invalid is thus beside the point. The sentence (L3) *is* a permissible paraphrase of the sentence (L2); the question is then how (L3) ought to be understood.

I suggest that what Lewis is really doing is calling our attention to a certain plausible view. One natural way to understand our claims about the possible is to regard them as claims about *objects*, objects we might call possibilities. It is not so much that our assertions frequently have the external form of existential quantifications as that the existence of possibilities has a plausibility reflected by our manner of speech and writing.

So understood, Lewis's argument from ways and its parallel avoid another charge, the charge that such arguments are really just bits of ordinary language philosophy and are therefore to be eschewed. It is not our goal--or even one of our strong preferences--that our theory of things conform to our ordinary ways of speaking, the objection goes. If the relation between the two has any importance at all, our language should change to suit our theories, not the other way around. Otherwise we would find ourselves burdened with the

likes of astronomical theories in which the sun really does rise each day, and we would be disconcerted by the apparent non-existence of the average American family. But if Lewis is really inviting us to consider a certain plausible understanding of the possible, his argument does not really have much to do with ordinary language. Our language may stem from that plausible understanding, but the language itself is not required to serve as evidence for that understanding.

Lewis hints that the existence of states of affairs may be the sort of thing for which we might not even require an argument, and perhaps this is right: we seem to have a familiarity with states of affairs that makes it tolerably clear that there are such things. The argument from ways may be viewed as a description of our familiarity and the manner in which our language reflects it. This same familiarity also makes it tolerably clear that among the states of affairs are *9's being even*, *motherhood's being (necessarily) transitive*, *something's coming to be from nothing at all*, *something's being identical to something with different properties*. These, of course, are impossible states of affairs. There are also the likes of *there being a private language*, *someone's discovering a unicorn*, *an iron ball's having all the same non-relational properties as a distinct ball*, *Peter's freely refraining from an action he strongly desires to do and has no countervailing desires not to do*. These are states of affairs which at least some philosophers have thought to be impossible, though they are less clearly impossible than the foregoing examples. But the question whether these can be instantiated has no bearing here. The latter examples are quite obviously examples of states of affairs. We do not need to stop and ask whether it is possible that there be a private language before we can say whether or not there is such a state of affairs as *there being a private language*.

The above considerations indicate that there should be at least a presumption in favor of impossible worlds. Whether or not impossible worlds

are to be accepted as entities depends on whether the strength of contrary arguments is sufficient to overcome this presumption. When we come to objections, I will argue that each is faulty in one way or another, and hence that no objection gives us any strong reason for rejecting impossible worlds.

The Argument from Utility

Lewis puts the second argument in Quinean terms: Improvements in unity and economy of ideology are sometimes worth controversial ontology. The economy and power of set theory gives mathematicians (and the rest of us) good reason to believe in sets. The cost of believing in sets is well worth the benefits for one's total theory. Says Lewis, so it is with possible worlds. The benefits for our understanding of necessity and possibility and for analyses of numerous objects of philosophical inquiry in nearly all subfields of philosophy make the cost of believing in possible worlds worth paying. Weighing the costs of an ontology against its benefits for ideology is a matter of judgment, but in this case, Lewis says, the price is right, even if less obviously so than in the case of sets.[10]

I do not claim that impossible worlds by themselves will prove as useful as possible worlds have. My theory is merely an addendum to an already established theory. It counts in favor of the addendum that simply by positing entities of the same kind (states of affairs) as the entities of the main theory, the main theory and addendum together have substantially greater analytic power than the main theory alone. (For example, the accurate world-based semantics for counterfactuals of chapter 4 cannot be given with possible worlds alone.)

How useful do impossible worlds have to be in order to be acceptable as entities? Of course it is difficult to quantify utility, but even if we had a precise

[10] See On the Plurality of Worlds, pp.3-5.

measure of utility I think it would be difficult to say what degree of utility would make a given theory worth adopting. The situation is further complicated by the fact that the theory in question is an addendum, and so the better part of the utility of the main theory plus addendum derives from the main theory. Suppose a mathematician believes in real numbers because she thinks their utility is enough to warrant that belief. How useful must complex numbers be before she can properly add them to her theory? If complex numbers have one twentieth the utility of real numbers, is that enough?

The argument from utility is significantly less impressive than the argument from ways, I think. It is not clear that mathematicians (or anyone else) believe in sets because of the impressive results of set theory and the discovery that all of mathematics can be modeled by set theory plus definitional extensions. Didn't people believe in sets before any of these results, and weren't they right to do so? In both the case of sets and worlds, utility might have relatively little to do with one's reasons for accepting controversial ontology. This is likely to be so in the case of someone who finds the argument from ways persuasive. The argument stands on its own if it gives us good reason to believe in worlds (or rather, states of affairs), whether or not anyone has demonstrated that they may be used in any enlightening philosophical analysis. The argument from ways obviates considerations of cost, since its success must affect our ontology whatever the cost.

For these reasons, I suspect that in questions about the existence of states of affairs or numbers, at least, utility is only one of a variety of a theory's aspects that need to be evaluated, if the theory's utility is relevant at all. But if utility were the principal consideration, or a principal consideration, would impossible worlds be useful enough to be worth adding to one's ontology? My skepticism about utility's relevance makes it hard for me to say. Impossible worlds are useful, and even indispensable to certain world-based

analyses. I hope that those inclined to regard utility as of central importance will agree that impossible worlds make an impressive showing. Whether impossible worlds are useful enough to meet the costs will be a matter of judgment for such individuals, and perhaps it is best for me to leave that judgment in their hands. But for those that believe in states of affairs (whether because of utility or something like the argument from ways), impossible worlds are objects they already believe in. For them, impossible world theory does not posit additional entities, and so there is no ontological cost beyond what has already been paid.

So I have doubts about whether the argument is successful and about the significance of the role it would play in belief if it were successful. However it is not my primary goal here to discredit the argument from utility but to point out that there is again a parallel argument for impossible worlds. Those who are inclined to accept the argument from utility and to think that it provides an important reason for believing in possible worlds will find a very similar argument for impossible worlds, since impossible worlds, like possible worlds, bring the benefits of unity and analytic power to our total theory.

The meat of this claim must wait until chapters 4 and 5. In chapter 4 I will argue that although Lewis's possible world semantics for counterfactuals do an excellent job of capturing some central semantic intuitions about counterfactuals, the semantics fail to give an accurate account of the truth conditions of counterfactuals with necessarily false antecedents and fail to properly characterize the relationship between 'would' and 'might' counterfactuals with necessarily false antecedents. These failures can be corrected by a straightforward addition of impossible worlds to Lewis's semantics. In chapter 5, we will see how impossible worlds do work in semantics at precisely the point where possible world semantics seem bound to fail. In particular, impossible worlds are needed in the semantics for certain

propositional attitudes, since, e.g., one can believe inconsistent things, and the subtly impossible can be true for all one knows. In the semantics of relevant logic, impossible worlds allow us to think of relevant implication as a "necessitated" version of material implication without requiring that necessary truths are relevantly implied by any proposition, or that necessary falsehoods relevantly imply everything.

And I am confident that there are other applications. Both possible and impossible worlds are liable to come into play whenever modality is relevant, and modality is ubiquitous.

Truth in a State of Affairs

In everyday discourse, it is common to speak of what is *true in* a given state of affairs or situation ("situation" and various other terms often being synonymous with "state of affairs"). We may say, "In this case we have three options" -- awkwardly paraphrased, "That we have three options is true in this state of affairs." We may say, more naturally, "It is true in most situations that shouting at people will only make them angry." Frequently the mention of truth is suppressed: "In his situation there is no escape." Sometimes the state of affairs is treated grammatically as a place. Thus Kripke: "What do we mean when we say 'In some other possible world I would not have given this lecture today?' We just imagine the situation *where* I didn't decide to give this lecture or decided to give it on some other day" (Naming and Necessity, 44, emphasis mine).

This mode of speech has been incorporated into our more technical possible worlds talk. In his famous "Semantical Considerations on Modal Logic," Kripke appears to assume as a matter of course that certain propositions are true in (and possible in) the various possible worlds, and accordingly he defines a model as a binary function $\phi(P, \mathbf{H})$ where 'P' is a

accordingly he defines a model as a binary function $\phi(P, \mathbf{H})$ where 'P' is a propositional variable and '**H**' varies over the elements of a model structure which are to be thought intuitively as possible worlds. ϕ assigns a truth value **T** or **F** to each proposition-world pair, and so each model represents each proposition as being true in or false in each possible world.

Others have taken some steps toward explicating this notion of truth in a possible world. Plantinga, for example, says it is part of both what he calls the "Canonical Conception" and the actualist conception of possible worlds that propositions are true in possible worlds, and he offers this analysis of truth in a state of affairs: "A proposition p is true in a state of affairs S if it is not possible that S be actual and p be false...."[11] For Plantinga, a proposition is true in a state of affairs just in case the proposition is entailed by that state of affairs, in a naturally extended sense of the term 'entailed'. (We will see a difficulty with this analysis below.) Lewis, in contrast, remarks that the phrase 'at W' restricts the domains of quantifiers in its scope, and so behaves much like the modifier 'in Australia'. For Lewis, truth at a world is simply a species of truth, a species whose subject matter concerns only the contents of the world in question (in most cases).

These items in the philosophical literature are, I think, attempts to specify or make use of a pretheoretical notion. The pretheory and the language suggest that propositions may stand in a certain relation to states of affairs, a relation of being "true in".

How seriously should we take this suggestion? Here the dialectical position is reminiscent of the position in which the argument from ways is given. On the face of it, there is a relation which holds between propositions and states of affairs and which our "true in" language describes. Our lanugage is fallible; we needn't always judge that our usual mode of speech reflects the

[11] "Actualism and Possible Worlds," p. 259 in Loux's The Possible and the Actual.

metaphysical truth of the matter. Nonetheless, there is a presumption in favor of that judgment.

I think that our "true in" language springs from an intuitive grasp of states of affairs, propositions, and the relations between them. Two of the things this grasp tells us are these:

> (1) There are *many* impossible state of affairs; in particular, there is not only one,

and

> (2) States of affairs are to be individuated according to what is true in them, by what we might call *propositional content*.

Thesis (1) is as evident as the thesis that there are many necessarily false propositions. It is quite clear that *Socrates' being taller than himself* is not the same state of affairs as *addition's being non-commutative*. Like the propositions *Socrates is taller than himself* and *addition is non-commutative*, the states of affairs are not about the same things. There is an intentional difference between the two.

Note that we may regard (1) as a consequence of the adapted argument from ways. There are many ways things couldn't be, and thus many impossible states of affairs.

Regarding thesis (2): Worlds and other states of affairs, we have been taught, are stipulated, not discovered with powerful telescopes.[12] When we do attempt to specify a state of affairs, we try to characterize its content. It is difficult to say just what content is without recourse to metaphors. Content has to do with what a state of affairs contains, what it's about, what it involves. This is a rough characterization, but it is clear enough that whatever individuates states of affairs must be something in this conceptual neighborhood. Could it be that two different states of affairs have exactly the

[12] Kripke, Naming and Necessity, p. 44.

same content? If two states of affairs share their content, what is left that might distinguish one from the other? (An obvious alternative theory individuates states of affairs according to their logical extension; if, necessarily, state of affairs A obtains iff state of affairs B obtains, then A is identical to B. But this theory has the consequence that there is only one impossible state of affairs, and so is unacceptable. A bit more promising is the theory that identifies A with B whenever A and B mutually imply each other on a relevantist's notion of implication--though this theory, too, seems to conflate states of affairs with distinct content.)

The principal device we have for specifying content is our "true in" language. Indeed, "true in" locutions seem to be designed for that very purpose. Kripke, speaking of how we stipulate possible worlds, says that "A possible world is *given by the descriptive conditions we associate with it*" (44, emphasis his). Here he is speaking of the same thing, the propositions true in a given world. It seems fitting, then, to call the content of a state of affairs a propositional content. It is a content given by a certain class of propositions, those true in it. If we call the class[13] of propositions true in a state of affairs S the *book* on S (writing 'B_S'), we may say that states of affairs are to be individuated by their books.

[13] I use the word 'class' here since there does not appear to be a set of all propositions true in the actual world. Consider this adaptation of a Cantorian argument from Patrick Grim, "Logic and Limits of Knowledge and Truth." Suppose there were a set T of propositions true in the actual world (that is simply a set of true propositions). Then there would be at least one proposition in T for every member of $\mathcal{P}$(T), the power set of T. For example there would be the true proposition that says whether a particular proposition P is or is not a member of that member of $\mathcal{P}$(T). But Cantor's power set theorem says that the cardinality of any set is less than the cardinality of its power set; thus there could not be a different proposition in T for each member of the power set of T. So we must reject our assumption that there is a set of propositions true in the actual world. Similar arguments will hold against sets of propositions true in other worlds.

Admittedly, it is not entirely clear that using the word 'class' enables us to avoid the difficulty. We will consider some related worries in the next chapter.

Issues surrounding the above argument are discussed further by Alvin Plantinga and Patrick Grim in "Truth, Omniscience, and Cantorian Arguments: An Exchange."

Individuation by propositional content coincides with individuation by extension if it is assumed that the books on states of affairs are closed under entailment. Given both (1) and (2), however, we may conclude that not all books are closed under entailment. For suppose they were, and take any impossible state of affairs--say, *9's being a Caesar salad*. A necessary falsehood entails everything, so *every* proposition is true in this state of affairs. In fact, each impossible state of affairs is such that either a necessary falsehood is true in it or some collection of propositions true in it is inconsistent, so every proposition is entailed by the propositions true in each impossible state of affairs. So via (2), on the present assumption there is only one impossible state of affairs, contrary to (1).

As will be apparent later, this result is disastrous for a variety of uses to which we might wish to put the impossible world(s). More importantly, though, the result is highly counter-intuitive. Is *Socrates' being taller than himself* the same state of affairs as *some bachelor's being married*? These seem like different situations. In particular, it seems like *Socrates' being taller than himself* is about Socrates and represents him as having a certain property, whereas *some bachelor's being married* is not and does not. Like the propositions with which they are so closely connected, states of affairs are to be individuated intensionally, i.e., not by their extensions in logical space, as well as intentionally, i.e., by what they are about. Sometimes states of affairs which obtain in precisely the same possible worlds have intensional differences, just as necessarily equivalent propositions are sometimes distinct. *Socrates' being taller than himself* does not obtain in any possible world; neither does *some bachelor's being married*; but the former has a property--being about Socrates--which the latter lacks. And so by the Indiscernibility of Identicals, *Socrates' being taller than himself* and *some bachelor's being married* are distinct. For this reason, I think the conclusion that there is only one

impossible state of affairs (and hence at most one impossible world) is absurd. We ought to reject the assumption that leads us to this error and hold instead that truth-in-a-state-of-affairs is not in general closed under entailment.

Here, then, we find a swift reply to those who object to the notion of impossible worlds--or to the usefulness of the notion--saying that because a contradiction entails everything, there is at most one impossible world. We may grant that each proposition is entailed by a contradiction. But we deny that all books are closed under entailment. Some books *are* closed under entailment, e.g. the books of possible worlds. These books have closure because they are consistent and maximal. Other books, however, may lack either consistency, as in the case of impossible states of affairs, or maximality, as in the case of states of affairs not complete enough to be worlds.

Another consequence of theses (1) and (2) is that the Plantingean analysis of truth in a state of affairs fails. If a proposition is true in a state of affairs just in case it is, in the relevant sense, entailed by that state of affairs, then every proposition is true in every impossible state of affairs. Then, by (2), there is only one impossible state of affairs, contrary to (1). (The condition of a state of affairs S's entailing a proposition P is not a necessary and sufficient condition of P's being true in S; however, the condition is a necessary one. Whenever P is true in S, S cannot obtain unless P is true.)

Why not use Plantinga's other definition of 'true in', the one that says a proposition is true in a world just in case, if that world had been actual, the proposition would be true?[14] Again, (adapting the definition for states of affairs in general) we have a necessary condition; if P is true in S, then it's certainly true that if S had obtained, P would be true. However, it seems the condition is not sufficient. Consider the state of affairs *George Bush's liking broccoli*. Neither the proposition *the population of China is large* nor its negation *the*

[14] Nature of Necessity, p. 46.

population of China is not large is true in it. It may be that if *George Bush's liking broccoli* had obtained then *the population of China is large* would (still) have been true; nonetheless, that proposition is not a member of the book on *George Bush's liking broccoli*. If this is right, then for states of affairs in general the propositions true in a state of affairs and the propositions that would be true if that state of affairs were actual are not the same. (An argument that there are states of affairs like *George Bush's liking broccoli* is implicit in the argument from ways. The next section will provide an argument that there are states of affairs such that for some proposition P, neither P nor its negation is true in that state of affairs.)

If the Plantingean analysis fails, how shall we characterize the "true in" relation? I do not propose to give necessary and sufficient conditions of P's being true in S. Of course, it is nice to have such necessary and sufficient conditions to attach to our philosophical notions when we can, but we must not let clarity trump accuracy. If the only analyses available are faulty, then it is better to make do with an unanalyzed notion until a correct analysis is found than to endorse a false account as the truth. (Often in philosophical and, especially, scientific inquiry we quite properly make false simplifying assumptions in order to aid our investigation. Such cases are unlike the present case, where the flaw in the proposed analysis is directly relevant to the questions at hand and would yield faulty results rather than harmlessly simplifying our inquiry.) Fortunately, there are several features of the "true in" relation which enable us to locate the notion nicely. Two of these features are described above. And as we saw above, we may give a *partial* analysis of it. We may also note that the modal status of a state of affairs is correlated to the status of its book. A state of affairs is possible iff the conjunction of propositions true in it is possibly true, actual iff that conjunction is true, and necessary iff the conjunction is necessarily true.

Maximality and Propositional Content

Impossible worlds are maximal impossible states of affairs; but in what sense is the word 'maximal' being used here? Let us consider a few terms that are sometimes used in discussions of possible worlds.[15] The following are defined for states of affairs S and S^*.

> INC: S *includes* S^* iff it is not possible that S obtain and S^* fail to obtain.
>
> PREC: S *precludes* S^* iff it is not possible that both S and S^* obtain.
>
> MAX: S is *maximal* iff for every state of affairs S^* , either S includes S^* or S precludes S^*.
>
> PW: S is a *possible world* iff S is a maximal possible state of affairs.

Let us add our candidate definition of 'impossible world':

> IW: S is an *impossible world* iff S is a maximal impossible state of affairs.

And finally:

> WORLD: S is a *world* iff it is a possible world or an impossible world, i.e., iff it is a maximal state of affairs.[16]

Given these definitions, every impossible state of affairs is an impossible world. Suppose that R is an impossible states of affairs. It is not possible that R obtain; *a fortiori* it is not possible that R obtain and Q fail to obtain, for arbitrary state of affairs Q. Thus by INC R includes Q. Neither is it possible that R and Q both obtain, so by PREC, R precludes Q. So R includes and precludes every state of affairs, and thus doubly satisfies the definition MAX. Because R is impossible, it is an impossible world.

It seems rather odd, however, that each impossible state of affairs should be a world. *Kermit's being green and uncolored* is an impossible state of

[15] See, e.g., Plantinga's "Actualism and Possible Worlds".

[16] Quite often in the possible worlds literature, 'world' abbreviates 'possible world'. Naturally such an abbreviation would spawn confusion in the present context, so I will use 'world' by itself only in the generic sense of WORLD.

affairs (it's not that easy being green and uncolored), but it is not about nearly so many things as we would expect a world to be. It closely resembles *Kermit's being green*, one of many possible states of affairs which is not a world. *Kermit's being green and uncolored* says, so to speak, only slightly more than *Kermit's being green*, attributing to Kermit one additional property. And it says slightly less than *Kermit's being green and uncolored and a better jumper than himself*, though both are impossible states of affairs. Though *Kermit's being green and uncolored* does meet the maximality condition given in MAX, there is another sense in which it is not maximal; it does not say something about everything; it's not about everything. There is another conception of maximality lurking nearby.

If *Kermit's being green and uncolored* and an impossible world differ in what they say, or what they're about, then they differ in content (as we loosely characterized that notion in the previous section). If content of a state of affairs is indeed given by the propositions true in it, i.e. by its book, what we need is a notion of maximality given in terms of the propositions true in a state of affairs.

The desired content-based definition of maximality is given by

MAX*: A state of affairs S is *maximal* iff, for every proposition P, the book on S contains either P or ~P (and perhaps both).

It is a straightforward matter to see that states of affairs like *Kermit's being green and uncolored* are not maximal impossible states of affairs. We may plausibly suppose that the state of affairs in question is one of the many whose book is relatively small. Perhaps neither the proposition *Gonzo is a human being* nor its negation is true in it.[17] Then, as one would expect, according to MAX* *Kermit's being green and uncolored* is not an impossible world. With

[17] If my memory serves, Gonzo, according to Muppet lore, is technically a *weirdo*. Apparently being a weirdo is an alternative to being human, and this suggests that *human* and *weirdo* are mutually exclusive natural kinds.

MAX* in place, IW satisfactorily defines 'impossible world' as 'a maximal impossible state of affairs'.

The Theory of Unrestricted Books

If the above is correct, then the books on many states of affairs contain propositions which contradict each other. May any collection of propositions whatsoever be the book on a state of affairs, or are there restrictions governing which collections may serve as books? Perhaps we should insist that books are non-empty, since some proposition or other is true in every state of affairs. Even the books of such "null state of affairs" candidates as *nothing's existing* and *nothing's being true* contain (at least) the propositions *nothing exists* and *nothing is true*, respectively. Are books restricted in other ways as well?

The two ways of answering this question lead to two types of impossible world theory. A negative answer yields a theory on which every non-empty collection of propositions is the book on some state of affairs or another, even those collections which are nothing more than a haphazard assortment of propositions with no unifying principle at all. A positive answer cannot require that books be consistent collections of propositions, but it may posit closure under some version of relevant implication, or closure under conjunction, or under modus ponens. A theory of this type might also require, for example, that every necessary truth be true in every state of affairs.

I think that the negative answer is to be preferred, for three reasons. First, we have seen that not everything is true in each impossible state of affairs, and so states of affairs may fail to address certain issues (as Perry puts it), even when the answers to those issues are entailed by what *is* true in that state of affairs. If states of affairs may omit these entailed propositions from their books, books are not governed by the rules of logic in the way that we might have expected. Then it is very difficult to see why we should think

them governed by a rule of closure under conjunction, or by any similar rule. Some may find it intuitively obvious that if propositions P and Q are true in state of affairs S, then so is P&Q, but I suspect that this is generally a disguised form of the thought that all states of affairs are closed under entailment. One might hold that states of affairs are closed under conjunction but not entailment, I suppose, but it is hard to see what might motivate the thought that there is closure under conjunction besides the thought that there is closure under entailment. The obviousness of the entailment from P and Q to P&Q is indeed striking, but it does not follow that the latter is true in every state of affairs in which the former are true.

A second reason for supposing that they are not is that there is significant advantage in that supposition. Better than others, a theory of unrestricted books accommodates conflicting intuitions about what we refer to when we speak of a state of affairs of one description or another. One may think that "Clinton's winning the election" refers to a sparse state of affairs, such as the one whose book is the single-membered {*Clinton wins the election*}. Or one might think that this same phrase refers to a rich state of affairs--if not a world, then at least one whose book includes such items as *Dole lost the election*, *most eligible Americans voted*, and *Gore became vice president*. One might think that this rich state of affairs is not closed under strong inductive inference (perhaps *Clinton is an American* is true in it, but *Clinton probably voted* is not), though others may insist that the phrase refers to a state of affairs whose book is closed under strong inductive inference. The proposed theory both accommodates and explains these views; there are states of affairs of each sort mentioned, and the language we use to refer to them is almost always ambiguous. We refer to a rich state of affairs on one occasion, a sparse one on another, and something in between on a third. We need not choose between them, for each of them exists.

The third and most important reason for rejecting restrictions on books is that a theory with restrictions would miss some of the impossibilities. It is impossible that both of propositions P and Q be true while their conjunction P&Q is not true. Hence one of the impossibilities is that P be true, Q be true, and P&Q not be true. But a theory which says that all books are closed under conjunction does not contain this impossibility, and so there are more impossibilities than it says there are. Since a similar argument will be available in the case of any other restriction, I advocate the theory of unrestricted books (TUB). According to TUB, the content of each state of affairs is given by some (nonempty) class of propositions, and each (nonempty) class of propositions is the content of a state of affairs.

Some Consequences of TUB

If TUB is true, some states of affairs are what John Perry calls 'partial ways.'[18] In Perry's terminology, possible worlds are total ways, ways which provide yes or no answers to all of a collection of basic issues. Partial ways provide answers only to some issues, leaving others unaddressed. Many states of affairs are partial ways. These do not address certain issues; that is, the propositions true in them do not decide certain issues one way or the other. *George Bush's liking broccoli* is one such state of affairs. In the terms of the present theory, each issue is a proposition, and an issue is decided by a state of affairs if either that proposition or its negation is true in the state of affairs.

There are some surprising corollaries that we should note before moving on. Consider the relatively small state of affairs J which corresponds to the class of propositions whose only members are *Jack is nimble* and *Jack is quick*. If, as we have said, the propositions true in a state of affairs are exactly the

[18] See his "From Worlds to Situations". Perry introduces a manner of talking about propositions, states of affairs, and possible worlds "that is, so to speak, outside of the theory" (85) and it may be helpful for present purposes to have this alternate way of speaking available.

propositions contained in the corresponding class of propositions, then the propositions *Jack is nimble* and *Jack is quick* are true in J, and no other propositions are true in J. J is silent on the issue of whether Jack is nimble and Jack is quick (that is, the issue of whether the proposition *Jack is nimble and Jack is quick* is true[19]). Of course J does speak to the closely related issues of whether Jack is nimble and of whether Jack is quick, giving an affirmative answer to each.

As noted above, any complete or comprehensive possible state of affairs (any possible state of affairs which gives an answer to every issue) in which two propositions are true is one in which their conjunction is true. But incomplete states of affairs, or partial ways, need not address the issue of whether such a conjunctive proposition is true. There is no possible world in which the conjuncts are true and the conjunction is not, but not all states of affairs are worlds. The more modest states of affairs do not take a stand one way or the other.

Perhaps some will allege that I have misunderstood the referent of a phrase like 'the state of affairs in which Jack is nimble and in which Jack is quick'. They will say, "When we use such a phrase, we clearly mean to bring to our attention a situation in which the conjunctive proposition *Jack is nimble and Jack is quick* is true, and in which the propositions *Jack is nimble* and *Jack is quick* are true. One would never raise for consideration a state of affairs in which the latter two propositions are true and remain agnostic as to whether the conjunctive proposition would be true in that case, for clearly it would." Similar arguments could be given for claims that books are subject to some other restriction, as those mentioned above.

[19] Notice that there is an ambiguity here. 'The issue of whether P is true' might refer either to a proposition that is about P (e.g., *P is true*) or to the proposition P itself. Phrases like 'the state of affairs *P's being true*' are similarly ambiguous. We will see many examples of this sort of ambiguity in the pages that follow. Hopefully context will be sufficient to make most of these cases clear. (In the above case, I mean to refer to an issue that is about Jack, not an issue that is about a proposition about Jack.)

What we have here, at best, is an argument for the conclusion that the states of affairs *that we normally bring to mind* are closed under conjunction. And perhaps this much is true. It is questionable whether, in the normal course of events, we ever refer to or even consider such a sparse state of affairs as J. More often, we consider states of affairs which come much closer to being complete. We think of states of affairs which are agnostic regarding, say, the price of rice-wine in China, but which do make true the conjunctions of other propositions true in them. But of course the fact, if it is a fact, that one rarely or never considers states of affairs like J does not give us any reason at all to suppose that there are no such states of affairs. There may yet be states of affairs that are incomplete in such a way as to be agnostic even about conjunctions of its book's members. Such states of affairs are like the author who believes each claim in her book, but, thinking it likely that she has erred at some point or other, does not believe the conjunction of those claims.

Compare the above view of states of affairs to theories which say that propositions are sets of possible worlds: most sets of possible worlds will be rather motley collections, very much unlike any proposition that arises in normal conversation. (However, it is not too hard to see that for every set of possible worlds some proposition is true in exactly those possible worlds. If our set X is $\{W_1, W_2, W_3, \ldots\}$, one proposition true in those possible worlds is *some member of X is actual*.) Likewise we should expect that the states of affairs that normally come to mind comprise only a tiny subclass of the states of affairs, and that they may differ significantly from other states of affairs, such as the emaciated J.

So as far as these considerations go, there may yet be states of affairs which leave certain issues unaddressed, even issues whose answers are entailed by issues that the state of affairs does address. The fact, if it is a fact, that one rarely considers states of affairs like J which have only two

propositions true in them does not give us reason to think that there is no such state of affairs as J. In any case, it may be that our everyday reference to situations is ambiguous in this way: in addition to J, there are a host of other states of affairs whose books contain the propositions in {*Jack is nimble*, *Jack is quick*}, including the states of affairs whose books are:

-- {*Jack is nimble*, *Jack is quick*, *Jack is nimble and Jack is quick*}

-- {*Jack is nimble*, *Jack is quick*, *Jack is nimble or Jack is quick*}

-- the closure of these under entailment

In addition, there are the states of affairs whose books are:

-- {*Jack is nimble and Jack is quick*}

--{*Jack is nimble and Jack is quick*, *Jack jumped over a candlestick*, *Jack jumped over a candlestick nimbly and quickly, Jack jumped over a candlestick as a demonstration of his nimbleness and quickness, the candlestick was sitting on the ground when Jack jumped over it*}

-- the closure of the previous set under entailment

When we entertain a situation in which Jack is nimble and quick, which of these states of affairs--if any--are we considering? It may be far from clear. In fact, it may make little or no difference, depending on our reason for bringing the situation to mind.

Suppose someone says, "Imagine what Bill Clinton would look like with Dennis Rodman's hair." The speaker brings a state of affairs to our attention because he or she finds it amusing or startling or otherwise worth imagining. But which state of affairs is being presented for consideration here? Is it the state of affairs whose book is {*Bill Clinton has hair like Dennis Rodman's*}? Or are we meant to imagine some possible *world* in which *Bill Clinton has hair like Dennis Rodman's* is true (perhaps one of the nearest such possible worlds)? Maybe we are meant to imagine one of the multitude of states of affairs which gives more detail than the one with the single-membered book but less detail than any world does. Our thoughts, no doubt, will contain a rather motley

collection of images, facts, and other imaginings, much of which may be irrelevant to the proposed mental exercise. If there is any state of affairs that we entertain in so doing, it will be extraordinarily difficult to say which one.

If the speaker's intent is simply to amuse us, it may make very little difference exactly what is true in the situation we consider; nearly any of those we have mentioned will do the job as well as any other. Likewise if the speaker intends to shock us, or instruct us, or confuse us. The point is that our reference to states of affairs is often ambiguous. Maybe you are inclined to think that every state of affairs we consider in the course of our imaginings is closed under entailment (though this is dubious). Even if it were so, it would be no indication that there are not other states of affairs unlike any that you do entertain. In particular, there may be some which, like J, are not closed under entailment.

One moral of this section is that truth-in-a-state-of-affairs in general need not be regarded as closed under entailment. (Nor is it closed under relevant entailment or obvious entailment or conjunction or any such thing.) In this respect, states of affairs resemble other entities which are closely related to propositions. Sentences, for example, do not express every proposition entailed by a proposition they do express. The proposition *Jack is nimble* entails the necessary truth *the external direct product of Abelian groups is an Abelian group*, but the sentence 'Jack is nimble' hardly expresses the latter. Thus the class of propositions expressed by a particular sentence is not closed under entailment. To make use of an important comparison, states of affairs are something like stories, and, like stories, they may remain silent about all sorts of matters--even about necessary truths. Even inconsistent stories are silent about some things.[20] It may be true according to the story that there is a

[20] See Priest's "Sylvan's Box" on this point.

computer more powerful than itself, but not true according to the story that there are no computers.[21]

This similarity is another moral: states of affairs have content in much the way that stories and sentences and thoughts *et cetera* have content. There will be disanalogies, of course. Stories have the content they do by virtue of linguistic conventions, whereas states of affairs presumably have their contents essentially. But the contents of both stories and states of affairs will be given by collections of propositions which are not necessarily closed under entailment. This similarity is what makes impossible worlds perfect candidates for tools in the analysis of propositional attitudes and in other areas.

Some Consequences of MAX*

It is easily confirmed that possible states of affairs that are maximal in the sense of MAX* (MAX*imal) are also maximal in the sense of MAX (MAXimal). Suppose that W is MAX*imal and is a possible state of affairs. Let S be a state of affairs. Either every proposition in B_S is a member of B_W or not. Suppose that $B_S \subseteq B_W$. A state of affairs obtains iff every member of its book is true. If W obtains each member of B_W is true, and so each member of B_S is true and S obtains. It is not possible that W obtain and S fail to obtain, so W includes S. Suppose on the other hand that $B_S \not\subseteq B_W$. Some member P of B_S is not a member of B_W. Because W is MAX*imal, the negation of P is a member of B_W. Since it is not possible for both P and ~P to be true, it is not possible for both W and S to obtain; W precludes S. Whether or not $B_S \subseteq B_W$, W either includes or precludes S. Since S is any state of affairs whatever, W includes or precludes every state of affairs and is MAXimal.

The converse is not so easily confirmed; in fact, there are counterexamples, possible states of affairs which are MAXimal (and hence are

[21] See Douglas Adams's Dirk Gently's Holistic Detective Agency, p. 200.

possible worlds in the resulting sense of PW) but which are not MAX*imal. One such counterexample is the state of affairs whose book is the class of all contingent truths. Call this state of affairs 'Twin α' for its close resemblance to the actual world, α. First we show that Twin α includes all actual states of affairs. Because all contingent truths are true in $B_{Twin\ \alpha}$, it is not possible that Twin α obtain and some contingent proposition that is actually true be false. In addition, it is not possible that any necessary proposition be false; *a fortiori* it is not possible that Twin α obtain and a necessary truth be false. So for any collection of true propositions, it is not possible that Twin α obtain and some member of that collection be false; it is not possible that Twin α obtain and the state of affairs whose book is that collection fail to obtain. Thus Twin α includes all states of affairs whose books contain only true propositions. These are the actual states of affairs.

Next we show that Twin α precludes all non-actual states of affairs. If S is a non-actual state of affairs, then some member P of its book is false. If P is a necessary falsehood, then clearly it is not possible that S obtain, so Twin α precludes S. If P is a contingent falsehood, then since $B_{Twin\ \alpha}$ contains all contingent truths, $B_{Twin\ \alpha}$ contains ~P, so it is not possible that both Twin α and S obtain. Twin α precludes S. Thus every state of affairs is either included or precluded by Twin α, and Twin α is MAXimal. But Twin α is not MAX*imal ($B_{Twin\ \alpha}$ contains no necessary truths) so a state of affairs can be MAXimal (and hence a world according to definitions MAX and PW) without being MAX*imal (and hence not a world according to the definition of 'maximal' that I propose as correct).

Why the discrepancy between the two versions of maximality? Aren't possible worlds normally thought to be maximal in the sense of MAX* as well as that of MAX? Yes, they are. Recall that 'P is true in W' is sometimes defined

to be true whenever W entails P. If truth in a state of affairs is understood in this way, then any state of affairs that is MAXimal is also MAX*imal. For suppose S is MAXimal. Then for any proposition P, there will be a state of affairs which obtains iff P is true--*P's being true*, say. *P's being true* is either included or precluded by S. If it is included, it's not possible for S to obtain unless P is true, so S entails P. If it is precluded, it is not possible for S to obtain and P to be true, so S entails ~P. So for any P, S either entails P or ~P, and if to be true in a state of affairs is to be entailed by it, S is MAX*imal.

On this understanding of the 'true in' relation there is no such state of affairs as Twin α, that is, no state of affairs whose book contains exactly the contingent truths. On this understanding, every book contains all the necessary truths. It therefore makes a difference--even in the realm of the possible--whether our notion of maximality is based on entailment relations or on propositional content.

Extensionalism, the view that states of affairs are to be individuated by their entailment relations, recognizes only one impossible state of affairs. As we noted earlier, this is an unpalatable result. What happens, then, if one denies extensionalism but still understands the 'true in' relations as an extended form of entailment? In this case the same propositions might be true in distinct states of affairs. For example, all propositions would be true in both *Kermit's being green and uncolored* and *Plato's being a married bachelor*, even if these two were distinct. What would distinguish such states of affairs? The only answer I can imagine here is that there is a difference of content (or *something* in that conceptual neighborhood). So this sort of view seems committed to the idea that a state of affairs's content is not exhausted by the propositions true in it.

This thought puzzles me; our 'true in' locutions do seem designed to specify (some of) the contents of states of affairs. Certainly if it is admitted

that content is not exhausted by entailment relations, it seems clear to me that 'true in' locutions aim to capture the former. Further, the resulting view falls into the difficulty Paul McNamara points out[22]: it strongly suggests that there is more than one actual world. If a world is any state of affairs that is MAXimal or MAX*imal (it makes no difference which definition we use if we understand 'P is true in S' as 'P is entailed by S'), then there seem to be distinct but equivalent possible worlds. McNamara argues that *there being no contingent objects* and *there being no contingent objects and there being no number larger than itself* are distinct possible states of affairs, since one can consider the first without considering the second. He then says that each of these is in fact a world; every contingent state of affairs is either entailed by or inconsistent with a complete lack of contingent objects. So on this view some possible worlds are equivalent to others. He goes on to argue to that "the" actual world can hardly be an exception in this respect. There are a variety of obtaining, maximal states of affairs--actual worlds.

We might disagree about whether *there being no contingent objects* is a world. If it should turn out that certain quantum level events are indeterministic and that there is a truth about whether one of these events E would occur in certain circumstances C, then there is a contingent proposition (*if there were contingent objects and indeterministic laws of nature as per C, then E would occur*) neither entailed by nor inconsistent with *there being no inconsistent objects*. (And whether or not this is in fact the case, it does seem possible.) But the success of McNamara's example is not crucial. If we are already prepared to believe that there are distinct states of affairs with the same entailment relations, then distinct equivalent *maximal* states of affairs are just what we should expect.

[22] In his "Does the Actual World Actually Exist?"

The defender of truth-in-S as entailment-by-S might even be willing to accept this result, and perhaps to soothe our discomfort with the idea of multiple actual worlds. I find this view more awkward than the alternative. It is interesting in any case that we should have thought our notion of worlds and truth-in-a-world gave us a unique actual world. I suspect that--both before anyone attempted to define 'true in' and after the entailment definition was offered--we implicitly linked maximality and individuation and truth-in-a-world with *propositional content* rather than extension. And I think our best theory of states of affairs makes this implicit commitment explicit instead of replacing it with another.

We have the theory. Now on to objections.

CHAPTER 2: OBJECTIONS AND MISCONCEPTIONS

The "Only One Impossible World" Objection

There is one objection to impossible worlds that most readily comes to mind and has already been touched on above. The initial plausibility of this objection may even be one of the primary reasons why relatively little has been made of the scattered mentions of impossible worlds in the metaphysics of modality literature. The objection is this: since a necessary falsehood entails everything, must it not be the case that every proposition is true in every impossible state of affairs? And if we distinguish propositions by what is true in them, doesn't it follow that there is only one impossible state of affairs?

My answer, as I indicated in the first chapter, is not to deny that a necessary falsehood entails everything.[23] Nor do I deny that states of affairs are to be individuated according to their content, i.e., by their books. Instead, I deny that truth-in-a-state-of-affairs is closed under entailment. Truth-in-a-possible-world is closed under entailment, but closure does not hold in the cases of non-maximal states of affairs (partial ways) and inconsistent or impossible states of affairs. The objection has an initially plausible but false premise.

---

[23] I have assumed that entailment is strict implication. If I am right about this, then the proper logic of entailment is not a relevant logic. However, a view of states of affairs and impossible worlds much like my own is nonetheless open to relevantists. Such a view may be particularly attractive to those who have stood in defense of intensions and meanings, as relevantists have.

It does not seem to me that the logic of entailment is a relevant logic, but there is some controversy about this question, and so in chapter 5 I suggest that relevantists may find the theory of impossible worlds useful in the semantics of relevant logic, as Richard Routley, e.g., has proposed. Even if the relevantists should turn out to be wrong about the nature of entailment, there will certainly be use for relevance in the logic of conversation and in other logics, and impossible worlds will play a similar role in these settings. In short, the theory of impossible worlds may be regarded as true and useful regardless of one's position on the nature of entailment.

There is another formulation of this objection which makes no explicit mention of truth in an impossible world, but which instead makes use of the notions of inclusion and preclusion introduced earlier. Any impossible state of affairs, this formulation says, includes and precludes every state of affairs, so there is only one impossible state of affairs. The suppressed premise, of course, is something to the effect that states of affairs are to be individuated according to the modal relations specified by the notions of inclusion and preclusion.[24]

Again, my response to the objection will already be clear from the position outlined above. States of affairs are to be individuated by content, which is to be understood in terms of books, which may include any non-empty collection of propositions. Two states of affairs (such as α and Twin α) may share their inclusion and preclusion relations and yet be distinct, so the suppressed premise of the objection is false. The objector may well agree with me that states of affairs are to be individuated by content; in this case, she mistakenly assumes that the content of a state of affairs is exhaustively specified by its inclusion and preclusion relations. She may even agree that content is to be understood in terms of books; in this case, she mistakenly assumes that books must be closed under entailment, or, equivalently, that to be true in a state of affairs is to be entailed by that state of affairs.

---

[24] A slightly stronger assumption of this sort is apparently made, e.g., by Plantinga: "Obviously *at least* one possible world obtains. Equally obviously, *at most* one possible world obtains; for suppose two [possible] worlds *W* and *W** both obtained. *Since W and W* are distinct worlds, there will be some state of affairs S such that W includes S and W* precludes S*. But then if both *W* and *W** are actual, *S* both obtains and does not obtain; and this, as they say, is repugnant to the intellect" (Nature of Necessity, 45, emphasis added).

I would argue for the uniqueness of the actual world in this way. Suppose that W and W* are distinct actual worlds. Because they are actual, their books B_W and B_{W^*} contain only true propositions. Worlds are to be individuated by their books, so the distinctness of W and W* implies that one of their books, B_W, we may assume, contains some proposition P which the other does not contain. P is true, so ~P is false. Then ~P is not a member of B_{W^*}. But if neither P nor ~P is a member of B_{W^*}, then W* is not maximal and thus is not a world, contrary to supposition.

So although the objection that there could only be one impossible world is initially plausible, both formulations depend upon false assumptions and pose no obstacle to the theory of impossible worlds.

A Modal Misconception

If there are impossible worlds, then the floodgates are open: every proposition that exists must be true in some world. One might worry that some proposition or other is such that it could not be true in any world, possible or impossible. Here I will consider an objection to the effect that one particular type of proposition has this feature, and then I will consider a general worry of this sort. The more specific objection is not an especially good one--the error is quite easy to see--but it is an objection that at least had some initial plausibility to me as I was sorting these issues out, and I hope that something may be gained in making the objection explicit.

There are propositions whose truth in a possible world guarantees that certain things are true in other possible worlds. The necessity operator gives us some such propositions: for any proposition P, there is another proposition $\Box$P which is true iff P is true in all possible worlds. $\Box$P is true in a possible world W just in case it is true in W that P is true in all possible worlds, and it is true in W that P is true in all possible worlds exactly when P is true in all possible worlds. So the truth value of $\Box$P depends on which possible worlds P is true in. We expect (the worry goes) an analog applying to worlds in general, a proposition whose truth depends on what is true both in possible and in impossible worlds. Suppose we define another modal operator, call it the 'ultranecessity' operator, as follows.

> ULTRA: The proposition $\blacksquare$P is true iff the proposition P is true in all worlds.

(We can also define an 'infrapossibility' operator: ◆P is true iff ¬■¬P is true.) It looks as if the truth of ■P depends on what propositions are true in worlds in general.

Now for any P it must be the case that ■P is true in some impossible world. But ■P is true in some world just in case P is true in every world, the actual world included. So P, which may be as nasty or false or contradictory a proposition as you like, is true. If this repugnant conclusion follows from the existence of impossible worlds, then impossible worlds must be rejected.[25]

The problem with the argument is that it has a false premise, viz., the premise that ■P is true in some world just in case P is true in every world. ULTRA gives us only that ■P is true iff P is true in every world. It appears as if the objector has assumed that ■P is true in some world iff ■P is true. But the assumption is mistaken: ■P may be true in an impossible world despite the fact that ■P is in fact false.

The mistaken assumption and the false premise that follows it get what little plausibility they have from similar claims about necessity and possible worlds. It is a fact that □P is true in some possible world iff P is true in every possible world, and that □P is true in some possible world iff □P is true, since what is necessary does not vary between possible worlds. However, what is necessary does vary between worlds generally. Unlike possible worlds, impossible worlds may misrepresent matters of necessity. Possible worlds may represent a contingent claim as true though it is in fact false; impossible worlds may represent any false claim as true, even those that are not possibly true. So it is with ultranecessity. For no P is ■P true; nonetheless an impossible world may represent things as being such that ■P is true.

If it is kept firmly in mind that impossible worlds may misrepresent even matters of necessity, then it should be clear that there is no substance to

[25] Thanks to Anthony J. Everett for help in constructing a clear statement of the worry.

the general worry that certain false modal propositions could not be true in an impossible world because of the consequences for actuality (or for other worlds). There would be no untoward consequences.[26] For although □P's being true in a possible world has implications for where P is true (it must be true in all possible worlds), neither □P's nor ■P's nor any other proposition's being true in an impossible world implies (for any P) that P is true or that P is true in any possible world. Any expectation one might have that impossible worlds would have problems in this neighborhood almost certainly comes from acquaintance with possible worlds, which do tell the truth about modal matters. When one remembers that impossible worlds are not as trustworthy as their possible cousins in this regard, the worry evaporates.

Lewis's story metaphor is a useful, non-technical device which makes this point clear. Worlds are like stories. According to some stories there are unicorns and fire-breathing dragons, even though it's not true. Still, there are such stories. Some stories are even more outlandish. They say things that could never be true--sometimes things about other stories. Some stories say that there is no little girl in the story "Little Red Riding Hood", and that it's in fact about a yellow frog named Jeremiah who had great leaping abilities and the magical power of speech. It's all false, of course, but still, there is such a story.

For another example, take the Amazing Story, which begins, "A long, long time ago, before your grandparents' grandparents were born, there were no stories ...." The Amazing Story goes on to tell of Nathaniel's heroic but ultimately unsuccessful quest to create stories in defiance of the Trolls of the Silver Spring. The tale closes: "And so to this day we hope for someone to overcome the Trolls and bring us stories." According to the Amazing Story,

[26] There would be *some* consequences. For example, if ■P is true in impossible world W, it follows that the propositions *■P is true in W* and *some proposition is true in some impossible world* are true in α. But of course these consequences are true and thus unproblematic. It is not this sort of consequence that the objector has in mind.

there are no stories; the Amazing Story *is* a story; but there is no conflict between these facts. The Amazing Story, after all, is fiction. We don't have to believe what it says about itself or the other stories.

Lewis's Objection

In On the Plurality of Worlds David Lewis rejects impossible worlds:

> For comparison, suppose travelers told of a place in this world--a marvelous mountain, far away in the bush--where contradictions are true. Allegedly we have truths of the form 'On the mountain both P and not P'. But if 'on the mountain' is a restricting modifier, which works by limiting domains of implicit and explicit quantification to a certain part of all that is, then it has no effect on the truth-functional connectives. Then the order of modifier and connective makes no difference.... [T]he alleged truth 'On the mountain P and not P' is equivalent to the overt contradiction 'On the mountain P, and not: on the mountain P'.... But there is no subject matter, however marvelous, about which you can tell the truth by contradicting yourself. Therefore there is no mountain where contradictions are true. (7n)

Lewis goes on to say that he thinks that 'at so-and-so world' is indeed a restricting modifier, unlike 'in such-and-such story', since worlds are like the actual world, not like stories.

It is this last point that is of interest here. Lewis's reasons for rejecting impossible worlds stem from his concretism, that is, his view that worlds are concrete objects much like us and our surroundings. Other worlds differ from the actual world (which he thinks is the same thing as us and our surroundings) in a wide variety of facts, but not in kind.

Worlds are ways things either could or could not have been--maximal states of affairs. States of affairs, I think, are not concrete objects but abstract ones. Hence I think that worlds are more like stories than mountains with respect to how the modifier 'at so-and-so world' ought to be taken. We noted earlier that states of affairs, like propositions, are representational. They represent things, accurately or otherwise, as having certain properties and

standing in certain relations. This representing of things is another feature which states of affairs and stories have in common, and it is this feature which makes it appropriate to use the modifier 'at so-and-so world' much as we use 'in such-and-such story': 'In state of affairs S, ~P' is not equivalent to 'Not: in state of affairs S, P'. We might be misled by the special case of possible worlds, since for any possible world W, 'in W, ~P' is true just in case 'Not: in W, P' is true. The equivalence holds in this special case because possible worlds are both maximal and consistent. But some states of affairs are not maximal; one might be silent about both P and ~P, so that the equivalence fails. Or a state of affairs might be inconsistent, such that it represents both P and ~P as true. Modifiers like 'in state of affairs S' do operate like 'in such-and-such story'. And as Lewis himself comments, "If worlds were like stories or story-tellers, there would indeed be room for worlds according to which contradictions are true" (7n).

Since Lewis's objection to impossible worlds is aimed only at concretist theories like his own, it is not an objection which my own abstractionist theory needs to refute.[27] However, suppose that we digress for a few pages and ask whether the Lewis-style concretist really does have sufficient reason for rejecting impossible worlds.[28] According to the concretist, much of what we say comes with implicit restricting modifiers, like the modifier 'on the mountain' in the above quotation. There are no flying donkeys, the concretists will agree. That is, there are no flying donkeys hereabouts, in this world (and many others), though at some worlds the sky is full of them. Gravitational attraction between bodies is always inversely proportional to the square of the distance between them. That is, gravity so operates hereabouts, in this world (and

[27] I do not take up the issues that divide concretists and abstractionists, since I have little to add to what others have said. See, e.g., van Inwagen's "Two Concepts of Possible Worlds" and Plantinga's "Two Concepts of Modality: Modal Realism and Modal Reductionism".

[28] I thank Michael Thrush for suggesting that the concretist position does not exclude impossible worlds as easily as it might appear to.

many others), though at some worlds gravitational attraction is inversely proportional to the cube of the distance between bodies, and in other worlds there is no such force at all. The concretist might continue: no computer is more powerful than itself. That is, no computer is more powerful than itself hereabouts, in this world and many others (including all the possible worlds), though at some worlds certain computers are more powerful than themselves. Speaking with completely unrestricted quantifiers, there is an infinity of flying donkeys, and gravitation attracts (or repels?) at a wide variety of rates, only some of which depend on the masses of the bodies involved[29], and some computers are more powerful than themselves. For the concretist, to say that some necessary falsehood is true may be merely to point out that, unrestrictedly speaking, some things are true that are not true at any of the possible worlds, and so not true hereabouts.

Perhaps the concretist who opposes impossible worlds will respond that the theory of impossible worlds is too costly. Possible worlds, as Lewis says, enable us to achieve a certain degree of unity and economy in our total theory at the price of some disagreement with common sense, and the price is right. But no theoretical advantage that impossible worlds might afford, the objector says, can be worth the price of contradiction. And contradiction is the cost we pay in adopting a concretist theory of impossible worlds. It commits us to saying that *every* proposition is true somewhere or other, if not anywhere nearby.

There are indeed theoretical advantages to the thesis that there are impossible worlds, but I do not claim that they are so extravagantly

[29] Or in any case gravity does not always attract according to the familiar inverse square law. It is relatively noncontroversial that it is metaphysically possible that the laws of nature be different than they are, but I am not sure how much a force can differ from gravity as we know it and yet be gravity. After some degree of difference, the force is something other than gravity. But in a way the issue of how much difference the notion of gravity can tolerate is moot, since even if no force which, say, repels objects is gravity in any possible world, according to the present argument there will be impossible worlds in which gravity does just these things.

advantageous as to outweigh the cost of contradiction. What I want to explore is whether if Lewis successfully ameliorates the cost of certain disagreements with common sense by pointing out an implicit restriction of quantifiers on those claims (call this 'the Restricted Quantifier Move'), then the Restricted Quantifier Move also offsets the cost of disagreement about whether contradictions are true.

If common sense is what the man on the street would say, then there is no complete escape from the cost of disagreement with common sense. The Restricted Quantifier Move cannot change what common sense says. Rather, the Restricted Quantifier Move attempts to point out where common sense fails to make the required distinction between some proposition regarded as plainly true and another that the theory in question denies. Common sense says that there are no million-carat diamonds; the Restricted Quantifier Move grants that there are no million-carat diamonds spatiotemporally related to us, but denies that there are no million-carat diamonds at all, quantifiers completely unrestricted. The woman who accepts the Restricted Quantifier Move must still be willing to disagree with the man on the street, but she judges this disagreement to be less costly than he does because she has an explanation of how he goes wrong. She judges that he conflates two propositions and that the one she denies inherits most of its plausibility from the other. So although the Restricted Quantifier Move cannot eliminate the cost of disagreement with common sense, it can significantly decrease that cost in the judgment of one who accepts it.[30]

Is the cost of believing in round squares higher than the cost of believing in infinities of donkeys? Probably so. The notion that no necessary falsehoods are true is no mere popular wisdom; our modal and logical intuitions, whatever their limits, are strong and clear in such matters. The claims of common sense,

[30] See On the Plurality of Worlds, pp. 133-5.

taken broadly, about at least some matters of necessity are much less negotiable than are its claims about how many donkeys there are, and so I suppose that the cost of disagreeing with these segments of common sense is greater than it is in other cases. (I don't know how to quantify the degree to which these claims are less negotiable, so it is difficult to characterize the difference precisely.)

Nonetheless it seems to me that the Restricted Quantifier Move, if it is successful, significantly offsets even this great cost. I cannot give a proof that the remaining cost is worth paying, but I have argued that the cost is small. The thesis that there is an uncountable infinity of flying donkeys is, after all, in rather spectacular disagreement with common sense. Surely the cost of this disagreement would be fatal to Lewis's theory if the Restricted Quantifier Move did not cover nearly all of it--if, for example, the theory said that there is an uncountable infinity of flying donkeys in actuality. The Restricted Quantifier Move must be equipped to make great costs negligible, so even a small gain in utility would seem likely to make the costs of a concretist theory of impossible worlds worthwhile if the Restricted Quantifier Move works.

Perhaps impossible worlds are not too costly for the concretist, but there are relevant considerations other than considerations of cost. We need to say not only whether the theory's benefits make it worthwhile to disagree with common sense, but also whether the theory is consistent. Indeed, Lewis's objection to a concretist theory of impossible worlds is that it is not consistent. I have argued that (if concretism about possible worlds is true) the concretist has resources for arguing that the theory is consistent in the only sense that matters: no contradiction is true (hereabouts, or in any other possible world). But other objections may be put forward, and in fact I think there are arguments which show that a concretist theory of impossible worlds is not viable after all. Here are two such objections.

First, if a contradiction is true in some world, then some contradiction is actually true, viz., some contradiction with completely unrestricted quantifiers.[31] And such a contradiction being true in the actual world *is* a reductio of the concretist theory. It is not the case that consistency hereabouts is all that really matters when it comes to the evaluation of theories. I do not know what *arguments* one might advance for this claim. It may be that the only way one could support the claim is to carefully consider the claim vis-a-vis the contention that consistency in the actual world (or in each possible world) is all the consistency that is required of a good theory, and to make the intuitive judgment that local consistency is not all that matters. Indeed, such a judgment seems reasonable, and though the concretist may disagree with it, he or she is powerless to refute the judgment on anything but similarly intuitive grounds. I myself judge that no contradictions are true (quantifiers completely unrestricted), and so I think that the concretist theory of impossible worlds is untenable.

Second, this argument. Consider an impossible world W such that a large number of the propositions about W that are true in W are incompatible with that world being a concrete object. Could such a world be a concrete object? The answer hinges on how concrete worlds represent propositions as true. If Lewis is right and a proposition is true at W just in case it is true when we quantify only over things in W, then W represents itself as concrete iff W is concrete (quantifiers restricted to things in W). Here the quantifier restrictions do very little work. If W is concrete (quantifiers restricted to things in W), then W is concrete, and vice versa. By hypothesis, W does not represent itself as

[31] I assume here that such a proposition may be said to be true in the actual world. The nature of the proposition prevents us from assessing its truth value at the actual world by asking whether it is true when we restrict the scope of our quantifiers to the contents of the actual world. This procedure would only give us the truth value of a different proposition, one without completely unrestricted quantifiers. If the proposition in question is not true in the actual world, then the actual world is not maximal, and thus not a world after all.

concrete, so W is not concrete. If representation does work this way, then any theory according to which all worlds are concrete is inconsistent.

Must concrete worlds represent in the manner described above? For the concretist, the thing which represents Humphrey as waving is a person very much like Humphrey, waving. How might such a thing represent? Lewis mentions a couple of possibilities. An other-worldly person might represent Humphrey by being Humphrey; that is, Humphrey might be a part of many overlapping worlds. Without going into detail, there are serious difficulties for this view regarding how different worlds can represent Humphrey as having different intrinsic properties. Is Humphrey, overlapped by these various worlds, waving or not? The other, more promising option is counterpart theory, according to which an other worldly person represents Humphrey as waving by being a waving counterpart of Humphrey. Something counts as a counterpart of Humphrey if it is sufficiently similar to him in important respects, whatever these may be.

Now if concrete worlds must represent by way of counterparts, is the above argument successful? What counterpart theory explains is *de re* representation, representation of an object, like Humphrey, as having some property, such as the property of waving. The above argument deals with the conditions under which certain propositions might be true in or according to a concrete world. We need not ask just how a concrete world might represent *de re* of a proposition that it is true (a task that is made more complex by the claims of some that propositions are trans-world objects, such as set of possible worlds). It is enough to note that the proposition *Humphrey waves* will be true in a given world just in case that world represents *de re* of Humphrey that he is waving. (Even if there are worlds according to which Humphrey is waving and Humphrey is not waving, there is agreement between what that world represents *de re* and what is true in that world. Such a world would

represent of Humphrey that he is waving and that he is not waving.) So a world is concrete according to itself just in case it represents itself as being concrete. It represents itself as concrete just in case the counterpart of that world in that world is concrete. Since the counterpart of a world in that world is just that world itself, a world represents itself as concrete iff it is concrete. So counterpart theory gives us the same result about representation as thinking about representation in terms of restriction of quantifiers.

The long and short of it, then, is that the above argument against concrete impossible worlds is successful; concrete worlds must represent in a way that is not compatible with the theory that all worlds, possible and impossible, are concrete. The Achilles' heel of a concretist theory of impossible worlds is the fact that there are certain things which concrete worlds cannot represent inaccurately: the concreteness of worlds, for example, and perhaps other facts, such as those regarding what occurs at other worlds, or certain truths about whatever trans-world objects there would be. In contrast, if worlds are thought to be abstract, there is nothing to prevent inaccurate representation on any topic whatsoever. It might be true in a world W that it is concrete (the proposition *W is concrete* might belong to B_w), despite the fact that W is abstract and not concrete.

Does this argument presuppose the intuitive judgment that contradictions regarding what is true at other worlds are sufficient reason to reject a theory? If the judgment were false, couldn't it be that at some world W it is true that W is concrete and not concrete, though in the actual world it is true that all worlds are concrete? I think not. Each of the steps of the argument is proposed as true, and each true proposition is true in the actual world. The supposition that W is not concrete at W thus leads to a contradiction in the actual world, and we have a reductio ad absurdum even by the concretist's standards. Hence the two objections are independent.

In sum, we may agree with Lewis's conclusions: a concretist theory of impossible worlds is not viable, though there is nothing to prevent the abstractionist from recognizing such things.

A Less Controversial Alternative?

Suppose one is willing to grant that the argument from ways is a sound argument. Is it really necessary to adopt a garish ontology of impossible worlds? Might one not instead say that certain classes of propositions -- certain books, as I have called them -- are suited to play the role of ways things couldn't be and that it is superfluous to posit states of affairs in addition?

Let us consider three positions along these lines:

Position 1: All the argument from ways tells us is that something or other plays the functional role of ways things couldn't be. Possible worlds are states of affairs, but there is no need to posit impossible world states of affairs as well since the work of impossible world theory can be done with maximal, inconsistent classes of propositions, which are less suspect ontologically.

Position 2: All the argument from ways tells us is that something or other plays the functional role of ways things couldn't be. The things that play the role of ways things *could* be are concrete objects rather than abstract states of affairs. Since concreta cannot play the role of ways things couldn't be, this role must be filled by something else such as propositions or collections thereof.

Position 3: States of affairs are propositions. Both do the same metaphysical work, and to distinguish them is to multiply entities without necessity. The roles of possible and impossible worlds are filled by similar objects. Specifically, possible worlds are large, consistent collections of propositions (or consistent propositions) and impossible worlds large, inconsistent collections of propositions (or inconsistent propositions).

First, let's examine the claim, shared by Positions 1 and 2, that we need only think that some object or another play the role of ways things couldn't be. The idea is Lewisian in spirit. With respect to ways things might be, Lewis says, "I suppose it is a firm commitment of common sense that there are some entities or other that fill the roles, and therefore deserve the names. But that is not to say that we have much notion what sort of entities those are."[32] And: "All this is a matter of fitting suitable entities to the various rather ill-defined roles that we rather indecisively associate with various familiar names. Don't think of it as a matter of discovering which entities *really are* the states of affairs, or the ways things might be, or the possibilities, or the propositions, or the structures!"[33]

The idea, I take it, is that 'ways things couldn't be' (to return to the case at hand) names a functional kind, like the kind doorstop. Anything which performs the task of stopping a door is a doorstop, and a great variety of things might perform the task. Likewise any things which together perform the requisite tasks are impossible worlds, and it is a mistake to say that certain states of affairs (or propositions or what have you) are *really* the impossible worlds, since many different objects might occupy the roles well enough. The requisite tasks, it is supposed, are those involved in modeling impossibilities.[34] There must be as many elements of the model as there are impossibilities; all elements must share some formal property which we may fairly dub 'maximality'; some element must represent the impossibility that all propositions are true, and that element must therefore be associated with or related to each proposition in a certain way; and so on. It is reasonably clear that if impossible worlds are as I say they are, then the classes of propositions

[32] On the Plurality of Worlds, p. 184.
[33] On the Plurality of Worlds, p. 186.
[34] However, this cannot be a completely satisfactory way of putting it. According to the views in question there are no impossibilities to be modeled aside from the various models themselves. Truth be told, they are not models *per se*, but only collections of elements which are related in the prescribed ways.

that are their books are able to perform these tasks extremely well. As an additional bonus, the reduction of impossible worlds to books has the advantage of clarifying the nature of a proposition's being true in an impossible world: it is simply that proposition's membership in that world.

It is difficult to say precisely what counts as a functional kind. It is not simply that belonging to the kind is a matter of occupying a certain role. Belonging to any kind is a matter of occupying a role of some sort, i.e. of having certain properties. We might do better to say that functional kinds may be instantiated by a much wider variety of things than the variety of those that instantiate non-functional kinds. But, besides being hopelessly vague, this approach does not seem to capture the notion either. Shall we consider the kind dog to be a functional kind? Most would say no, despite the fact that dogs exhibit at least as much variety as, say, television remote controls.

Perhaps we could make some progress here by insisting that belonging to a functional kind is a matter only of having a certain sort of property, whichever properties are appropriately related to function. Whether or not such a strategy would ultimately prove successful, we need not bother ourselves with finding a solution to this puzzle now. The more pressing question is this: why believe the suggestion that 'impossible world' names a functional kind (whatever precisely a functional kind is)? Lewis has a few things to say about states of affairs and other entities being whatever fills certain roles[35], but gives little in the way of explicit remarks that indicate *why* one should think that this is the nature of states of affairs. We do not have much notion what sort of entities fill the roles with which we are relatively familiar, he says, but of course the fact that we encounter some difficulty saying just what a thing is does not itself show that we are dealing with a functional kind.

[35] See On the Plurality of Worlds, pp. 182-7.

I suspect that Lewis's motivation stems from his belief that all that exists are concrete possible individuals and set theoretic constructions of them. For if that is all that exists, then states of affairs, if there are any, could only be concrete objects or sets. The difficulty is that neither concrete objects nor sets are representational objects[36] as states of affairs (and also propositions) must be. Since they cannot *be* states of affairs, the best that concreta and sets can do is to play the role of states of affairs, imitating them as best they can, substituting a similarity-based counterpart relation for genuine representation. Confronted with a good argument for states of affairs, Lewis can only conclude that to be a state of affairs is to belong to a certain functional kind. And so to find what the states of affairs are, "we must survey the candidates according to our best systematic theory of what there is" (184).

Another approach would be to allow the conclusion that there are representational objects to *inform* our theory of what there is. If by the argument from ways or some other means we come to believe that there are states of affairs, we may conclude that an ontology of concreta and sets is too sparse to be true. Those who take this approach or who otherwise do not share Lewis's commitment to an ontology of concreta and sets will not share his motivation for regarding states of affairs as whatever occupies a functional role. And, so far as I can tell, there is little other reason to regard states of affairs in that way.

Position 2, then, I reject because of its concretism about possible worlds and its functionalism about both possible and impossible worlds. I also reject Position 1 because of its functionalism about impossible worlds.

Position 1 has additional problems to contend with. The hope is apparently to avoid controversial ontology, but the questionable entities are

[36] Such objects are not representational by nature, at any rate. Some material objects, such as the sentences on this page, may be said to represent by virtue of being incorporated into a language. But possibilities and impossibilities cannot be representational because of their place in a language unless that language itself exists necessarily.

included in an ontology that recognizes states of affairs, as Position 1 does. There is no ontological advantage in moving the name 'impossible world' from one sort of entity to a more familiar one; the ontological commitments stay the same. The view avoids commitment to the objects I call impossible worlds only if it adds, "Though there are possible states of affairs, there are no impossible states of affairs." As suggested earlier, there is something deeply incongruous about a view so amended. And if this amendment is not added to Position 1, then the differences between the view and my own theory are terminological and not ontological, hence the supposed advantage of Position 1 over my theory is illusory.

Position 3 may or may not be taken in conjunction with the kind of functionalism that characterizes Positions 1 and 2; here let's consider it apart from that functionalism. Do we commit an offense against Ockham's Razor if we recognize abstract states of affairs (such as *Socrates' being mortal*) distinct from the propositions they seem to correspond to (*Socrates is mortal*)? We say that states of affairs obtain or fail to obtain; propositions, on the other hand, are true or false. But does this difference in terminology reflect a real difference between states of affairs and propositions, or is it a quirk of language that we have two different idioms for describing one kind of object and its ways of standing in relation to the world? One proposal is that there is a real difference in the fact that propositions are possible objects of certain attitudes and states of affairs are not. Chisholm presents this argument against the theory that propositions are states of affairs and his reply:

> '(i) Your theory implies that, if a man believes that a storm is occurring, then that state of affairs which is the occurrence of a storm is the object of his belief. But (ii) the sentence "He believes that a storm is coming" is natural and clearly grammatical, whereas "He believes the occurrence of a storm" is unnatural and not clearly grammatical. Hence (iii) if a man believes that a storm is occurring something other than the occurrence of the storm is the object of his belief.'

> The premises of the argument are certainly true. If we wish to say of a man that he believes that a storm is occurring, we do not *say* 'He believes the occurrence of a storm'. But we may say 'He believes in or suspects, or is counting on, or is mindful of, the occurrence of a storm'. And where we may say of a man that he fears, regrets, hopes or knows that a storm is occurring, we may also say, equally well, that he fears, regrets, hopes for or is cognizant of the occurrence of a storm. Such points of usage may throw light upon various intentional attitudes. But surely they give us no reason to suppose that 'the occurrence of a storm' and 'that a storm is occurring' refer to different things. The argument is simply a *non sequitur*.[37]

Chisholm is right. The fact that our linguistic conventions disallow locutions like "He believes the occurrence of a storm" is not conclusive evidence that *the occurrence of a storm* (the state of affairs) is not the object of the man's belief. However, we also lack conclusive evidence that 'the occurrence of a storm' and 'that a storm is occurring' do refer to the same object. Clearly there is a close relationship between propositions and states of affairs; plausibly there is even a 1-1 correspondence. But neither a close relationship nor a 1-1 correspondence is sufficient to establish the identity of one sort of entity with the other. (Compare the correspondence between propositions and their unit sets.) So it appears that neither position has a decisive advantage over the other. The above considerations, taken alone, leave me without a significant inclination toward either view. I weigh my guess that our language reflects reality against my fear of multiplying entities and find that in this instance they are approximately equal. But there is one other consideration which, if not compelling, at least moves me to prefer the theory that states of affairs and propositions are distinct.

The consideration has already been mentioned. Whether or not there is identity between propositions and states of affairs, there is some relation between them. In particular, there is a relation of *being true in*; every situation is such that certain propositions are true in it. (Even if the identity thesis is

[37] Person and Object, p. 125.

true, the relation in question is not identity since in that case many propositions are true in each world-proposition.) States of affairs have books. It would be absurd to suppose that there is only one impossible state of affairs, so the books on states of affairs are not all closed under entailment. Since states of affairs do not subject their books to a closure requirement, books may contain as many or as few propositions entailed by their other members as you like. But then any (non-empty) class of propositions may be the book of some state of affairs. Hence the 1-1 correspondence between states of affairs and classes of propositions.

Is this correspondence inconsistent with the identity thesis? I.e., might propositions be in 1-1 correspondence with classes of propositions? If there were a set of all propositions, the answer would be a swift "No". As it stands, I am not quite sure. But the answer may well be "No", and I am wary of giving an implicit "Yes" to this question, so I operate under the assumption that states of affairs are not propositions.

The plausible reasons one might have for thinking that states of affairs are in correspondence with propositions are easily accommodated on this view. One might say that for each proposition, e.g. *Socrates is mortal*, there is a state of affairs which consists in that proposition's being true--*Socrates' being mortal*. And it is so: for each proposition there is a state of affairs whose book is the unit set of that proposition. This state of affairs exactly consists in that proposition's being true. One might also claim that for each state of affairs there is a proposition which says that things are as they are in that state of affairs. This, we should say, is close: there is some collection of one or more propositions which together say that things are as they are in the state of affairs. (The conjunction of propositions in a given book, which might be taken a single proposition which does the job, may fail to be one of the things true in a non-maximal state of affairs.) This account does justice to the sorts of

intuitions we have about the relations between propositions and states of affairs. There is no clear reason to prefer the 1-1 correspondence account to this one.

Positions 1 and 2 seem to me to be clearly wrong; Position 3 seems unclearly wrong. Position 3 avoids the awkward type-difference between possible and impossible worlds with which Positions 1 and 2 are afflicted, and the case against Position 3 does not compel assent. I am happy to report, though, that if I am wrong about Position 3, the theory of impossible worlds survives in a modified form. Impossible worlds would in that case be certain necessarily false propositions, and the existence of a multiplicity of impossible worlds would be even more evident. Much of the theory would need to be revised, but my central claims would stand. I hope that those who are strongly inclined to identify states of affairs with propositions will make full use of the theory of impossible worlds, suitably revised.

The Analysis of Possibility

Another objection to impossible worlds goes along these lines.

> The view that there are possible worlds but not impossible worlds ('PWO' for 'possible worlds only') gives us a nice account of modality: necessity is truth in *all* worlds and possibility is truth in *any* world. The view that there are possible and impossible worlds ('P&IW') says that necessity is truth in all *possible* worlds. But this view raises a question. What makes the possible ones possible? We can't say without giving some independent account of possibility. All we can say is that possibility is truth in some of a certain collection of worlds; but this criterion is completely uninformative. We are left without any explication of the notion of possibility. So P&IW makes a mystery of modality. If there are only possible worlds, however, the question 'What makes this world possible?' does not arise.[38]

The objection alleges that P&IW raises certain questions about modality which it ought to answer. In particular, it ought to answer the question "Why are possible worlds possible?" and answer it in an informative

[38] Many thanks to Brian Leftow for helpful articulation and discussion of this objection.

way. I think what the objection really means to require of a world theory is an *analysis* of modality. That is, it assumes that a theory of worlds must provide necessary and sufficient conditions for the necessity, possibility, etc. of propositions without making any ineliminable use of those notions. It is in this sense that the sentence "A bachelor is an unmarried man" gives an analysis of the notion of bachelor in terms of the notions of man and of being unmarried. The analysis of modality, it is assumed, is what enables us to give an informative answer to the question of what makes a given world possibly obtain. The answer will take the form "The satisfaction of condition C," where condition C, taking its cue from the analysans, will make no use of modal terms.

Lewis's account is an example of a reductive analysis of modality. Possible worlds, he says, are spatiotemporally isolated objects, objects that stand in spatiotemporal relations only with their parts. A proposition's being possibly true is thus to be understood as that proposition's being true in some spatiotemporally isolated object, which in turn is to be understood (in typical cases) as being true when the ranges of the quantifiers involved in the proposition are restricted to some spatiotemporally isolated object. Thus we have an attempt (unsuccessful, in my view) to reduce modality to the notions of spatiotemporal relatedness, truth, object, etc.

However, not all possible world accounts purport to provide an analysis of modality. An ontology of possible states of affairs, for example, might make no attempt to explain what a possible state of affairs is without use of the notions of possibility, necessity, or some other modal term. In fact, it is rather commonly thought that any such attempt would be futile because possibility, necessity, and their ilk form, as it is said, a tight circle of interrelated modal notions, none of which can be properly analyzed without recourse to some element of the circle. If this is so, we may point out useful relations between

modal notions (e.g., whatever is possibly true is not necessarily false, and vice versa), but there is no *more* informative analysis of modality to be found.

So is an analysis of modality in non-modal terms a *sine qua non* for P&IW or not? The objector assumes so, but does not argue for this claim. It is at least plausible that there is no further explication. Analysis must come to an end somewhere, and our failure to produce such an analysis so far (Lewis notwithstanding) gives us some reason to suspect that this is the place. At the very least, some argument for the necessity of a more informative analysis would have to be given before we had a substantial objection here. As it currently stands, the objection merely assumes a premise denied by many modal theorists, and which may well be false.

In any event, a PWO theory which does not actually supply an informative analysis also fails to meet the requirements of the objection. For the question "Why is world W possible?" may also be asked of PWO. If the proposed answer is that W is possible because it exists, we may ask why W exists. One who holds PWO will say that 'a situation in which Mars is colonized' succeeds in referring to various entities but that 'a situation in which Mars is a divisor of 7' does not. What accounts for this difference? In order to meet the objection, the proponent of PWO must answer the question without any ineliminable use of modal terms. Otherwise she, too, "makes a mystery of modality".

Conceivably those who say that modal notions form an irreducible circle are mistaken and there exists an answer to the question. But if so, we do not yet know what that answer is. Until it is *shown* that there is a more informative analysis of modality (and, furthermore, one that cannot be used by the impossible worlds theorist to explain the difference between possible and impossible worlds) the objection puts PWO in a position no better than P&IW. The advocate of impossible worlds, then, has little to fear from this worry.

The Fine-Grainedness Objection

Yagisawa mentions an interesting objection against his Lewis-style theory of impossible worlds. Part of the motivation for adopting Lewisian (i.e. concrete) possible worlds to begin with, he says, is the resources this gives us for an extensionalist theory of properties. Any theory that identifies a property with the objects that actually instantiate it falls to familiar criticisms. *Having been born with a heart* and *having been born with a kidney* are distinct properties with the same instantiations, so such a theory is too coarse-grained. Concrete possible worlds allow us to distinguish these properties by the possibilia that instantiate them, their extensions in all possible worlds, since it is possible that something be born with a heart but not a kidney.

We naturally hope along similar lines, Yagisawa continues, to use impossible worlds to distinguish between distinct but necessarily coextensive properties such as triangularity and trilaterality. Some impossibilia will have triangularity but not trilaterality, and other impossibilia will have trilaterality but not triangularity, and so the two are distinct. But isn't such a proposal *too* fine-grained?

> It even appears that no property is ever identical with any property whatever, according to the above proposal. For any property P and any property Q, either it is possible for P and Q not to be coextensive or it is impossible. If it is possible, there is a possible world where P and Q are not coextensive. If it is impossible, there is an impossible world where P and Q are not coextensive. Either way, the set of all possibilia and impossibilia having P is different than the set of all possibilia and impossibilia having Q. Therefore, according to the above proposal, P and Q are not the same property. This is true for any P and Q whatsoever, including P and P. So according to the proposal, no property is the same as any property, including itself! This is certainly an unwelcome consequence of any proposal ("Beyond Possible Worlds", 195).

Unwelcome indeed! Does this objection weigh against an actualist, abstractionist theory of impossible worlds as well?

First of all, we may note that the objection objects not to impossible worlds *per se* but to a theory of properties formed in the wider context of a theory of impossible worlds. Conceivably, it could turn out that the theory of properties fails but that there are impossible worlds nonetheless.

Second, it is no part of the actualist program to promote an extensionalist theory of properties. According to the actualist there are no impossible or merely possible entities that might fill out the extensions of properties as the extensionalist theory requires; nothing belongs to the "set of possibilia and impossibilia having *P*" but what actually has *P*. As it stands, then, the objection does not apply to actualist theories.

But suppose we set aside these points and ask whether the property theories that are most natural for the actualist friend of impossible worlds are susceptible to a similar objection. If so, then perhaps impossible worlds are not quite as useful as we would have thought. (This isn't very impressive as an objection to impossible worlds, I realize, but discussion of it will be instructive.)

How might the actualist take advantage of impossible worlds when faced with the problem of distinguishing between different but necessarily coextensive properties? The natural approach is to duplicate the identity conditions suggested by the extensionalist view without making reference to non-actual individuals. So one might say that property P is identical to property Q just in case this biconditional holds:

> (IC) For every possible or impossible world W, an entity X has P in W iff X has Q in W.

For the actualist, an entity X has property P in W whenever the proposition *X has P* is true in W. The objection could be fitted for this account in this way. For any properties P and Q, there is an impossible world (if not a possible world) such that some X has P in W but X does not have Q in W. But then IC fails and P and Q, whatever they are, are distinct-- even if P *is* Q.

The actualist may reply that IC *does* hold. For suppose that both 'P' and 'Q' are names of redness. Then there is indeed some world in which some X has P and lacks Q, which is to say there is some world in which X has redness and lacks redness. Now if X both has redness and lacks redness, then X has P and lacks P, and likewise X has Q and lacks Q. But then the biconditional IC does hold, since X has P in W and X has Q in W.

Yagisawa rejects a similar line of argument in the context of his extensionalist theory. The quoted passage continues:

> It is a mistake to maintain that the above proposal does not really have this unwelcome consequence, by arguing as follows: "It is not true that the set of all possibilia and impossibilia having *P* is different from the set of all possibilia and impossibilia having *P*. It is certainly true that in some impossible world an object has *P* and does not have *P* at the same time. But such a world is not a world in which *P* and *P* are not coextensive; the extension of *P* contains all objects that have *P*, including those that have *P* and do not have *P* at the same time." This is a mistake because since it is impossible for *P* not to be coextensive with *P*, it follows, on the extended modal realism, that there is an impossible world in which *P* is not coextensive with *P*; such a world is one in which the extension of *P* is not identical with the extension of *P*. Such a world is more than a world which merely contains something that is and is not *P*; it is a world in which the law of identity fails for the extension of *P*. Thus the above, natural proposal cannot be sustained ("Beyond Possible Worlds", 195).

According to an actualist theory, too, every proposition is true in some impossible world or another, so it may look as if we have the raw material for a reply to the actualist as well as to the extensionalist. I think, though, that Yagisawa's remarks succeed only within an extensional theory of impossible worlds, whereas the actualist avoids the objection.

Rephrased, the actualist abstractionist's claim is that whenever '*P*' and '*Q*' name the same property, 'X has *P*' and 'X has *Q*' express the same proposition, and so IC is satisfied. (And whenever '*P*' and 'Q' name different properties, 'X has *P*' and 'X has *Q*' express different propositions, and there will

be some impossible world such that one of the two is true in it and the other is not.) The supposed difficulty involves worlds in which *P* is not coextensive with *Q*, or (in actualist translation) in which the negation of *X has P iff X has Q* is true, even though '*P*' and '*Q*' name the same property. From an abstractionist perspective, however, such a proposition's being true in certain impossible worlds is entirely irrelevant. We would have a problem if ~(*X has P iff X has Q*)'s being true in a world W somehow prevented *X has P* and *X has Q* from being true together in W, but it does not. That ~(*X has P iff X has Q*) is true in W does not change the fact that *X has P* and *X has Q* are the same proposition. This is so even if the book on W reports (falsely) that ~(*X has P iff X has Q*) or reports (falsely) that ~(*X has P in W iff X has Q in W*) or reports (falsely) that *X has P* and *X has Q* are distinct propositions. So the identity criterion IC stands, regardless of whatever propositions turn out to be true in a given impossible world. The abstractionist is successful in holding a position of the sort that Yagisawa rejects.

This kind of position produces much more difficulty for the extensionalist, because the extensionalist is also a concretist. The concretist regards other worlds as places spatiotemporally unrelated to us, and so regards truth in a world as a species of truth, namely, truth regarding some particular domain. So if it is true in W that the extension of *P* is not identical with the extension of *P*, then for the concretist it is true *simpliciter* that the extension of *P* is not identical with the extension of *P*. Hence the fact that the extension of *P* is not identical with itself in some impossible worlds leads to intolerable difficulties for the concretist who identifies properties with their extensions. The abstractionist, though, is never obliged to think that what is true in an impossible world is true, even with respect to some limited domain. Instead, she regards truth in a world as a matter of a proposition characterizing part of the content of the abstract world in question. The propositions that are true in

an impossible world may simply be *false* propositions, and so what they say about the extension of P or about the identity criteria of properties or about that world itself is quite beside the point.

We are left with the question of how the extensionalist might deal with the fine-grainedness objection, if not along the lines available to the abstractionist. Yagisawa offers what he calls an "incomplete" solution to the difficulty. His attempt is to identify each property not with its extension in all worlds, but in some smaller group of possible and impossible worlds. His choice is the "analytically familiar worlds", i.e. worlds which share all analytic facts with the actual world. It is not an analytic truth that triangular things are trilateral, so triangularity and trilaterality have different extensions in the analytically familiar worlds. On the other hand, it is analytically true that vixens are female foxes, so in every analytically familiar world, everything that has the property *being a vixen* has the property *being a female fox*, and *vice versa*. We get the result that these properties are identical.

This solution is not *ad hoc*, Yagisawa says, since "we naturally expect two synonymous (i.e. analytically connected) predicates to express the same property" (197). Perhaps he means here that if one has already decided to identify a property with its extension in some but not all possible and impossible worlds, then the analytically familiar worlds are a natural choice. It still seems, however, that to identify a property with its extension in *any* group of worlds other than all the worlds is a significant departure from the extensionalist program. Among the unattractive aspects of Lewis's structuralist account of properties, Yagisawa lists the "striking feature ... that it abandons the basic modal extensionalist insistence that a property is to be identified with the set of things which have that property" (193). But Yagisawa, too, it seems, is willing to abandon this "basic modal extensionalist insistence" in favor of a different insistence, identifying each property with

something other than the objects which have that property. The reason for this shift is clear enough: the usual extensionalist claim is open to devastating objection, whereas the revised account seems to yield the desired property-identifications. But if this is the reason, then the proposed solution is *ad hoc* after all; it is motivated only by conditions of adequacy and not by the purported extensionalist insight. It seems an adequate theory of property individuation can be attained only if one *gives up* the central extensionalist claim, and this gives us reason to give up the central extensionalist claim.

Thus neither the proposal that Yagisawa rejects nor the one he advocates rescues a concretist theory of impossible worlds (coupled with an extensionalist theory of properties) from the fine-grainedness objection, but a version of the former does enable the abstractionist to refute the objection.

A closing remark: we should not think that an impossible worlds-based theory of property (or proposition) individuation will give us any new information about which properties (or propositions) are identical and which distinct. Like the world-based theory of modality, the purpose of such an account is to explicate certain relationships--in this case between properties and propositions, or between properties and states of affairs--not to arbitrate unclear instances. Our best guide in particular cases will remain conventional usage of the words which refer to the property or properties in question. It is usage which will tell us (or fail to tell us) that *being a vixen* and *being a female fox* are identical, and we may conclude from this that the proposition *Mary is a vixen* is true in a world iff the proposition *Mary is a female fox* is true in that world, because "Mary is a vixen" and "Mary is a female fox" express the same proposition.

The Specter of Set-Theoretic Paradox

The last objection I'll consider in this chapter alleges that my theory of impossible worlds engenders set-theoretic paradox. In particular, it alleges that the Theory of Unrestricted Books (TUB), which I said was needed to ensure that no impossibilities are overlooked, falls prey to the diagonalization trick of Cantor's Theorem (or--what amounts to the same thing--Russell's paradox).[39]

Let f be a function mapping each proposition to a book. We will show that f does not map the propositions onto the books. Let D = {x : x is a proposition and x is not a member of f(x)}. D is a class of propositions, so by TUB there is a state of affairs S whose book is D.

Now suppose that D is in the range of f, so that f(p)=D for some proposition p. Is p a member of D? If it is, then by D's definition p is not a member of f(p), i.e. of D. So p is not a member of D. But then p is a proposition and not a member of f(p), so p must belong to D after all. We've reached the contradiction.

So D must not be in the range of f. There is no mapping of propositions onto books; there must be more books than propositions. But it seems clear that for every book B there is a distinct proposition *B is a book*, so that there are at least as many propositions as books. Here again we have reached a contradiction, and it seems the only premise left to deny is TUB.

How shall we respond to such a forbidding charge? In the first chapter I said that we should think of books in general as classes rather than sets precisely because of a powerful Cantorian diagonalization argument against the assumption that worlds' books are sets. However I didn't say how we should conceive of classes or how classes could escape the diagonalization argument; and now it seems another incarnation of that very argument has appeared.

[39] I owe Michael Thrush and Tom Crisp many thanks for offering versions of this objection and for extensive discussion of the issues they raised.

Let's clarify the notion of class, then, as follows. Like a set, a class is a totality which (in general) has certain objects as members. In fact, sets are classes which obey all the axioms of ZF set theory; the class of sets is a subclass of the class of classes. But classes in general do not obey (analogues of) those axioms. In particular, we cannot assume there is an analog of the axiom of comprehension:

> Let P be a property. For any class A, there is a class B such that $x \in B$ if and only if $x \in A$ and x has P.

This axiom was implicitly used in the above argument against TUB. It is this axiom that allows us to suppose there is a class D = {x : x is a proposition and x is not a member of f(x)}. The property P, in this case, is given by the expression after the colon, i.e. *being a proposition which is not a member of its image under f*. The class A is the class of propositions, we may suppose, and B is D, the class whose existence is asserted.

Since there is no axiom of comprehension for classes, we cannot simply deduce (as we can in Cantor's Theorem) the existence of the diagonal D from the assumption that there is a class of propositions and a mapping of propositions into books (in Cantor's Theorem, that there is a set X and a mapping from X into its power set).

We might naturally be suspicious of classes, as above conceived--these purported objects which so resemble sets without being sets. Are there such things? I do not know how to give a direct proof, but several considerations may take the edge off our suspicion.

First, it is standard mathematical practice to distinguish between sets and (proper) classes. One speaks of the class of all sets, for example, and the class of ordinal numbers. Naturally, the results of set theory cannot be assumed to apply to classes as well as sets. Cantor's Theorem tells us that every set has more subsets than members, but each subclass of the class of

sets is also a member of that class, so the analog of Cantor's Theorem fails. Furthermore, classes are sometimes taken as the domains of functions. For example, a successor function maps the ordinals into the ordinals. It is not as if the failure of classes to satisfy the axioms of ZF renders the notion of class too obscure to be of use in mathematics.[40]

Second, set theory recognizes more specific kinds of setlike non-sets, such as second-order sets. There is no set of all sets, but we suppose that all (first-order) sets may be gathered into some other kind of totality, namely, a second-order set V. Since the axioms of set theory are so intuitive, we assume that similar axioms govern second-order sets. So, for example, there is a second-order axiom of comprehension, which tells us that for any property P and second order set A, there is a second-order set B such that $x \in B$ if and only if $x \in A$ and x has P. If we let A=V and P be the property of being non-self-membered, we can deduce the existence of a second-order set R of all first-order sets which are not members of themselves. Since R is not a first-order set, R is not a member of itself, and no contradiction follows. Because the axioms of second-order set theory parallel those of first-order set theory, there is no second-order set of all non-self-membered second-order sets, though these can be gathered into a third-order set.

The point is simply that there are setlike non-sets--objects which have members, are individuated by their extensions, have unions and "subsets", etc.--and these are well-enough understood. (But the classes referred to in TUB cannot all be sets of any particular order, since for no ordinal *n* is there a *n*th-order set of all truths, and the book on the actual world is the class of truths.)

[40] Paul Bernays went so far as to say, "This distinction between sets and classes is not a mere artifice but has its interpretation by the distinction between a set as a collection, which is a mathematical thing, and a class as an extension of a predicate, which in comparison with the mathematical things has the character of an ideal object" (Axiomatic Set Theory, pp. 56-7). I am suspicious, however, of the implied gulf between mathematical objects and ideal objects.

So the Cantorian argument does not apply in any straightforward way to classes or to states of affairs. The argument relies on premises taken from axiomatic set theory, but these axioms (or their analogues) cannot be assumed to hold for classes or states of affairs in general. Still, some difficult questions remain. Though we cannot take for granted axioms from which the existence of a diagonal class can be deduced, it may seem that there ought to be such a class. In claiming there is a class of sets, don't we assume there is a class of any objects whatever? On Bernays's account of the distinction, for any property there is a class of objects having that property. From this assumption the contradiction would seem to follow as before. So the difficulty can be avoided only if one adopts a view of classes on which not just any objects may form a class, or restricts TUB in some way, or denies the existence of the diagonal property of which the diagonal class is the extension. None of the three is particularly attractive, and I do not know how to decide between them. If it should turn out that TUB does need to be restricted, then so be it, but it is very difficult to see how this could be done plausibly. What propositions are such that there isn't even a state of affairs that consists in those propositions being true?[41]

Whatever the merits of this partial reply, the objection does not have anything in particular to do with the introduction of impossible worlds. A very similar objection challenges possible world theory, according to which every truth is true in the actual world. By the diagonalization argument given in the note on p. 16, there is no set of truths. And if there is a diagonal class of the

[41] Tom Crisp has argued that even if we had a satisfactory solution to this difficulty, the paradox could be reasserted without reference to classes. The argument would then proceed by way of plural quantification and multi-grade relations between propositions which duplicate the effect of mappings. I have chosen not to address this version of the argument in the body for two reasons. (1) The plural quantification version would mimic the class version, so a solution to the latter would suggest a solution to the former. No doubt the "solution" would strain against some intuition or another, but I seriously doubt it would do so much more than a solution to the class version. (2) Both versions of the argument alike present as much a problem for possible worlds theorists as for impossible worlds theorists.

sort mentioned above, the assumption that there is a class of truths appears to blossom in contradiction. (The problem arises in the same way for other possible worlds, since by maximality just as many propositions are true in one possible world as in any other.)

If no reply to the earlier objection is successful, no reply will be successful in this case either. If the only workable reply to the earlier objection is to restrict TUB, possible world theory will still have a problem, since we don't want to restrict TUB in such a way that there is no actual world (or that any possible world is "missing"). The trouble for possible worlds thus seems at least as serious as it is for impossible worlds.

Russell's paradox is persistent, and I have certainly not solved it here. However, the fact that it afflicts possible world theory (to say nothing of the other contexts it plagues) should assure us that the problem, deep and puzzling as it is, does not give us reason to think possible and impossible states of affairs are not on an ontological par. Rather, it gives us reason to think there is something we do not understand about sets and classes, or about infinite cardinals, or about mappings. Perhaps there is something wrong with our conceptualization of classes. If a solution is ever discovered, no doubt many of our theories will need to be subtly revised. But that does not imply that a theory which encounters this difficulty cannot be rather close to the mark.

To sum up, we have seen one objection which depends on false assumptions about the logical closure of books, another which does not apply to abstractionist theories, another which has only a superficial plausibility, and some alternative theories which are either problematic or else congenial to spirit of the proposed theory. The only objection which strikes me as seriously worrisome is also an objection against possible worlds, and so gives the possible worlds theorist no additional difficulties. If there are more potent objections

against adding impossible worlds to the theory of possible worlds, I do not know what they are. The next chapter, which describes in greater depth the nature of impossible worlds, will hopefully dispel any lingering doubts or confusions about impossible worlds that may remain.

CHAPTER 3: THE NATURE OF IMPOSSIBLE WORLDS

What are impossible worlds like? Told that there are such strange beasts, one inevitably wonders what things would be like if some impossible world were actual. What if 1 and 1 made 5? How would things be? In addition, many questions arise concerning the worlds themselves and their place in the metaphysics of modality. This chapter aims to provide a brief tour of the universe of impossible worlds, pointing out some salient features, interesting results, and connections with more familiar modal notions. We will begin with the natural question of how things would be if the impossible were actual.

Imagining the Impossible

There are at least three reasons why it is challenging to say what a world in which 1 and 1 make 5 is like, each of which should be familiar from similar difficulties with possible worlds. First, there is no single world in which 1+1=5. There are a great many worlds in which 1+1=5, and these differ significantly in character. Let W be a possible world in which French colonies came to dominate North America in the last few centuries. Do the denizens of North America have a greater tendency to eat French bread in W than they in fact do? Is the English language widely spoken in W? The answer, of course, is that we cannot say: there are possible worlds of the type in question in which French bread is eaten much more frequently than in actuality, and there are other possible worlds in which it is not eaten at all. Some of these worlds are such that English is rarely spoken outside of England, and others are such that English is spoken all around the world. No single description of how things are

covers all the possible worlds in which the French colonies came to dominate North America. Likewise in the case of impossible worlds there is too much variety among the worlds to allow any one answer to the question of what a world with a given inconsistency or necessary falsehood is like.

Second, it is not immediately clear that a world is the sort of state of affairs we are capable of imagining. The issue is size (or else "size"); one might plausibly think that too many things are true in a world (whether possible or impossible) for us to consider each of them being true at once. However, this way of putting it almost suggests that the act of imagining involves the the generation of a mental list of propositions, along with mental check marks to indicate which of the propositions are true in the imagined situation. The familiar fact of the matter is that we normally imagine in a variety of other ways.

Often one *pictures* how things would be; imagination may involve an image of what is imagined. Sometimes (and perhaps always) there is something more than the mere image; there is also a mental act of identifying the image as an image of something. Paul Tidman cites the Wittgensteinian example of a person imagining King's College on fire: "It would be absurd to ask, 'Are you sure it's King's College? Maybe you are just imagining a building that looks like King's College.' Whether one is imagining King's College depends not on the image, but on what we take the image to be an image of."[42] There may be other non-pictorial qualifications of our images. For example, we might imagine an individual water molecule which has existed intact for over one hundred years. Whatever image we produce need not differ from an image of a newly-formed water molecule. The difference is rather the non-pictorial thought--a caption of sorts--"The molecule has existed for over one hundred years." Sometimes, too, our imaginings may involve little or no imagery,

[42] p. 301, "Conceivability as a Test for Possibility".

consisting only in non-pictorial thoughts. If I am asked to imagine a state of affairs in which numbers are sets, the mental act I perform seems to be simply the supposition that numbers are sets, or else mere consideration of the proposition *numbers are sets*--and perhaps this is nothing other than a kind of mental check mark next to the proposition.

The question we need to consider is whether we may be said to imagine worlds in any of these ways. I think we can. For example, we can form images of worlds (that is, images of things which would exist if some particular world were actual). But isn't there (usually) too much that exists in a world for us to picture it all? Yes and no. Yes, since it seems that we are unable to produce mental images of enough detail and complexity that each thing extant in a typical world is represented by some part of our image. No, since we may picture a thing, or group of things, without picturing all of it. Our mental images of visible objects are indistinguishable from images of the visible surfaces of these objects from some perspective. Nonetheless, we succeed in imagining more than mere surfaces. There is a sense, then, in which the size of worlds does not keep us from picturing them (though perhaps the image must be accompanied by the thought that the pictured state of affairs is a maximal state of affairs, a way *everything* could or couldn't be).

The third snag concerns other limits of human imagination. There are possible worlds which, considerations of size to the side, we find extremely difficult to imagine in certain respects. Indeed, the actual world is just such a world. Even physicists need to make a special effort in order to visualize the universe as operating according to relativity, to say nothing of quantum mechanics. I am not sure precisely why we find this difficult; the difficulty decreases over time, so perhaps it has something to do with unfamiliarity or our conceptual resources. At any rate, something makes it difficult to hold even certain actual states of affairs in our heads.

Imagination also fails us in the case of impossible worlds, and fails at least as badly. Take the impossible world λ, which is such that each and every proposition is true in it[43]. What would things be like if λ were actual? Such a state of affairs is difficult even to begin to imagine. The best I can do is to (rather indistinctly) picture a handful of different events superimposed on each other and proceeding independently. The image is quite a far cry from what things would be like if every proposition were true, issues of size notwithstanding. How does one imagine that each of

all ravens are black,

there are thousands of white ravens,

there are no ravens

are true? Even worse are those propositions not about any easily-visualized material object, such as

all numbers are sets,

all sets are souls,

no souls are numbers.

The only thing one can do, it seems, is to remember which propositions are supposed to be true in the state of affairs (in this case, all of them), or else to hold the thought that all propositions are true, perhaps adding some decorative imagery. Attempts to *picture* the state of affairs with clarity and completeness are all but fruitless.

So for a variety of familiar reasons, there may be no satisfying answer to the question of what a certain impossible world is like. I linger on this point in part because I think some will take the absence of a satisfying answer as an objection to my theory. The dialectic goes something like this:

[43] Stalnaker gives the name 'λ' to the "absurd world" in his semantics for counterfactuals, and I imitate him here.

Doubter: "What would it be *like* for 1 and 1 to be 5? I don't think you can give any coherent answer to that question."

Me: "Well, if the question is about how things would seem to the inhabitants of certain worlds, then no single answer will do. There are a wide variety of ways it might seem. In some of the relevant worlds the proposition *when asked what the sum of 1 and 1 is, people generally say it is 5* is true, and in some worlds *when asked what the sum of 1 and 1 is, people generally say it is 2* is true."

D: "I guess my question isn't really about how it would *seem*, but about how it would *be*. For example, if Rebekah has one son, and then has another, would she have five sons?"

M: "Again, there is no single answer. Some worlds' books include each of the propositions *1 + 1 = 5*, *Rebekah had one son and then had another*, and *Rebekah had five sons*, whereas others include the first two but not the third, and some of these include the proposition *Rebekah had exactly three sons* or even *1 + 1 is not 5*. The best I can do in telling you what the various impossible worlds are like is to tell you which propositions might be true in them."

D: "But then it seems to me that you haven't said how things would be at all--you've just listed a bunch of inconsistent propositions rather than describing any coherent situation. You haven't managed to refer to any situation at all."

There are a number of different ways in which we might frame the doubter's thoughts, some of which are suggested by the preceding chapters. For now I want only to point out that this sort of doubt could be caused by the idea that everything that could count as a situation must be, in a certain sense, imaginable. Hearing that there are worlds in which 1 and 1 make 5, one wants to know how this works, and what sort of consequences this

arithmetical anomaly has. And though I will claim in the next chapter that (in certain contexts, at least) there may be an answer to the question of what would be true if 1 and 1 made 5, it seems there is little that can be said to make these situations "coherent" or to help us grasp *how* 1 and 1 could be 5.

But as the possible world cases illustrate, we do not have reason to think that states of affairs are in general easily imaginable. When we find that we have difficulty grasping an impossibility of one sort or another, we ought not suppose that our imaginative failure is indicative of a problem with the theory of impossible worlds, and we certainly ought not conclude that there are no such worlds. Imaginative failures are familiar enough from the case of possible worlds, and in any event they have little bearing on questions of existence.

That said, there are impossible states of affairs which we can imagine as well as we can imagine possible states of affairs. Considering the question whether imagination is any guide in distinguishing the possible from the impossible, Peter van Inwagen asks, "Can we imagine a world in which there is transparent iron?"

> Not unless our imaginings take place at a level of structural detail comparable to that of the imaginings of condensed-matter physicists who are trying to explain, say, the phenomenon of superconductivity. If we simply imagine a Nobel Prize acceptance speech in which the new Nobel laureate thanks those who supported him in his long and discouraging quest for transparent iron and displays to a cheering crowd something that looks (in our imaginations) like a chunk of glass, we shall indeed have imagined a world, but it will not be a world in which there is transparent iron. (But not because it will be a world in which there *isn't* transparent iron. It will be neither a world in which there is transparent iron nor a world in which there isn't transparent iron.)[44]

We have already affirmed the point that certain imaginative acts seem beyond our abilities. Van Inwagen makes the additional point that if any imaginings could confirm the possibility of what is imagined, these kinds of imaginings are

[44] "Modal Epistemology," p. 79.

normally quite beyond us. We cannot imagine a world in which there is transparent iron--not in any useful sense.

However we may ask whether there is any sense at all in which we can imagine such a world, and here the answer is that we can. Just as we can picture something without picturing all of it, we can picture a world with transparent iron (whether possible or impossible) by picturing the part with the award ceremony. It is in more or less the same way that we imagine a world in which there is *opaque* iron: we visualize a dark, heavy chunk, and someone says it is iron (or we think, "That's iron"). Even knowing what we do about iron, our imaginings are insufficient to show that opaque iron is possible. If any imaginings could show this, it seems they, too, would have to be at a level of detail comparable to that of condensed-matter physicists. But we can, in some sense, imagine opaque iron, and in this sense we can imagine transparent iron as well.

Certainly not just any imaginative act is an imagining of transparent iron. The question is what counts. When have we imagined something poorly, and when have we failed to imagine it? In some cases it may be a judgment call. I am inclined to be generous about what counts, since all our images are partial, unclear, and indistinct. Yet we do often succeed at imagining one thing or another. In fact, the ability we have to imagine so much is precisely what what makes the inference from conceivability to possibility dubious.

A Menu of Impossibilities

If TUB is at all close to the truth, impossible worlds are best understood via their books. Usually it will be more helpful to ask what is true in an impossible world than it is to look for an image or a feeling of "how things go" there. The best way to get a feel for the features of impossible worlds is to

examine a menu of examples. Below, therefore, is a sampling of books along with some comments on questions that arise along the way.

One preliminary notion: The books of all impossible worlds (with a single exception) are not closed under entailment. Each impossible world, to some degree or another, compartmentalizes whatever necessary falsehoods or inconsistencies are true in it. To make this idea more precise, let us say that an impossible world W has a *locus of impossibility* L just in case

(1) $L \subseteq B_w$,

(2) $B_w - L$ is a consistent class of propositions (i.e., possibly the conjunction of all its members is true),

(3) no proper subclass L^* of L is such that $B_w - L^*$ is a consistent class,

and

(4) no subclass L^* of B_w has cardinality less than that of L and is such that $B_w - L^*$ is a consistent class.

(The last two clauses are not redundant, as one of our menu items should make clear.) Informally put, a locus of impossibility is the least that needs to be removed from an impossible world in order to make it possible. The definition allows for the possibility that a given world has more than one locus of impossibility, and it has the consequence that every necessary falsehood true in a given world must belong to each of that world's loci of impossibility.

λ: Every proposition is true in λ. Its book is the class of all propositions. λ has *no* compartmentalization; each consequence of every necessary falsehood (that is, every proposition) is true in it. What are the loci of impossibility of such a world? They are the classes of propositions which contain all propositions but those true in some possible world. Hence λ has as many loci of impossibility as there are possible worlds.

ω: Of course there are other worlds which, due in large measure to multiplicity of inconsistent propositions true in them, are all but impossible to

imagine. For example, let ω be the impossible world whose book contains all and only the propositions that are actually false. If A is the class of all propositions and α is the actual world, then $B_\omega = A - B_\alpha$. The world ω is thus a kind of photonegative of actuality. Each of the following propositions is true in ω: *Napoleon was born in 1 A.D.*, *Napoleon was born in 2 A.D.*, *Napoleon was born in 3 A.D.*, *et cetera*, excluding *Napoleon was born in 1769 A.D.* A similar proliferation of propositions about every other topic will be true in ω, so it is clear that ω cannot possibly obtain. Since, for every proposition P, either P or its negation is false, it is also clear that ω meets the maximality requirement for worlds and is a world.

It is somewhat tempting to characterize ω as a world with a low degree of compartmentalization, or as being such that all but relatively few propositions are true in it, but this characterization is misleading. The cardinality (if we may speak loosely) of B_ω is no greater than the cardinality of B_α, since every proposition in B_ω has a unique negation in B_α.

δ: Consider the impossible worlds whose locus of impossibility has a small, finite cardinality. For example, let δ be the world such that $B_\delta = B_\alpha \cup$ {*Mars is blue*}. In other words, every proposition that is actually true is also true in δ, the additional proposition that Mars is blue is true in δ, and no other propositions are true in δ. Since every true proposition belongs to B_δ, we are dealing with a maximal state of affairs, and since *Mars is blue* is inconsistent with other propositions in B_δ (e.g., *Mars is not blue*), δ is impossible.

The only locus of impossibility of δ is the single-membered {*Mars is blue*}. It is easy to see that {*Mars is blue*}--M, for short--satisfies conditions (1) and (2) of the above definition. Conditions (3) and (4) are satisfied because M's only proper subset, and the only subset with a lesser cardinality, is the empty set, and $B_\delta - \emptyset$ is not consistent. δ has no loci of impossibility which are proper

superclasses of M because of condition (3). δ has no loci of impossibility which are disjoint from M, since such a locus would have to contain all true propositions inconsistent with Mars is blue, and each such class is ruled out by condition (4). Hence M is δ's only locus of impossibility. (Were it not for condition (4), some much larger classes would also be loci of impossibility, so (4) is not a superfluous part of the definition. Condition (3) might then appear to have no function, but it may yet be needed in other cases to rule out infinite classes satisfying (1), (2) and (4) and having proper subclasses of the same cardinality which also satisfy (1), (2) and (4).)[45]

Worlds like δ immediately enable us to prove certain results about what kinds of impossible worlds there are. For instance, we might wonder whether there are any impossible worlds in which no contradictions are true, or whether there are any impossible worlds in which no necessary falsehoods are true. The answer in each case, perhaps surprisingly, is yes. δ is our example. The proposition *Mars is blue* is not a necessary falsehood, and every other proposition true in δ is true in the actual world. Since no necessary falsehood is true in the actual world, none of the propositions true in δ is necessarily false, and *a fortiori* none is a contradictory proposition. (Recall that the books of impossible worlds are not closed under entailment, so the fact that both of *Mars is blue* and *Mars is not blue* are true in an impossible world does not entail that *Mars is blue and Mars is not blue* is true in that world.) The reason δ is

[45] Del Ratzsch has observed that this argument apparently assumes 'This world is actual' and 'α is actual' express the same proposition. By 'This world is actual' we do not express a proposition true in all possible worlds but differing in context, and so differing in reference. There is no such proposition. If there were, then it seems δ would not have a single-membered locus of impossibility. When *Mars is blue* was added to the actual world's book, 'this world' would shift reference to δ, so that the proposition expressed by 'This world is actual' would become a necessary falsehood.

The argument in the text does assume that 'this world' does not shift reference in this way. But I am comfortable with this assumption. Briefly, I am inclined to think that sentences can express different propositions in various contexts of assertion, but that propositions cannot shift in meaning or reference, the context already having been taken into account, so to speak. Propositions *are* meanings or contents, and so a proposition's reference (part of that content) is intrinsic to it.

impossible is that some of the propositions true in it contradict each other, and so they cannot possibly be true together. Nonetheless, no necessary falsehood is true in δ.

In Counterfactuals Lewis says,

> What is meant by the counterfactual ['*If kangaroos had no tails, they would topple over*'] is that, things being pretty much as they are--the scarcity of crutches for kangaroos being pretty much as it actually is, the kangaroos' inability to use crutches being pretty much as it is, and so on--if kangaroos had no tails they would topple over.
>
> We might think it best to confine our attention to worlds where kangaroos have no tails and *everything* else is as it actually is; but there are no such worlds. (9)

δ and its ilk are the worlds whose existence Lewis denies. The world in which kangaroos have no tails and everything else is as it actually is has the book

$B_\alpha \cup \{$*kangaroos have no tails*$\}$,

or perhaps

$(B_\alpha - \{$*kangaroos have tails*$\}) \cup \{$*kangaroos have no tails*$\}$.

The existence of such worlds does not imply that Lewis's semantics give the wrong truth conditions of '*If kangaroos had no tails, they would topple over*'. Worlds with finite loci of impossibility, by the mere fact that they are impossible worlds, are not nearby worlds. They are very dissimilar from the actual world, though some of them have all but identical books. Therefore it remains true that in the worlds most similar to the actual world in which kangaroos have no tails, kangaroos topple over.

The existence of a multiplicity of impossible worlds does, however, present a strong challenge to the Lewisian/Stalnakerian thesis that all counterfactuals with impossible antecedents are vacuously true. This, however, is a topic for a later chapter.

π: Given these results, we might go on to ask whether there are any impossible worlds in which no two propositions contradict each other, that is,

whether there is an impossible world such that for any two propositions true in that world, their conjunction is possibly true. Remarkably, there are. Our example may be regarded as an embodiment of the paradox of the preface. If an author says in the preface of her book that some claim made in the book is false, then her beliefs are inconsistent if she believes all the claims made in the book. Nonetheless each individual claim she makes may be consistent with her claim that some part of the book is in error; it is only the conjunction of all the other claims of the book that is inconsistent with the claim of the preface. The world to be presented is structurally similar to this scenario. Let us call it π.

The book B_π contains all the propositions that are true in α, the actual world, except the proposition *α is actual* and any propositions necessarily equivalent to it. Among the excluded propositions will be the conjunction of all truths[46], the conjunction of all contingent truths, and the negation of the disjunction of all falsehoods[47]. Naturally, B_π will contain the negations of *α is actual* and its necessary equivalents. This specifies all propositions which are true in B_π. π is a world, since for each true proposition that B_π does not contain, its negation does belong to B_π. And π is impossible: each possible world W is such that the proposition *W is actual* is true in it, but no such proposition is true in π since the only proposition of the sort that is true in α (*α is actual*) is stipulated not to be true in π.

[46] If indeed there is such a proposition. I do not know of any good reason for thinking that there are not infinite conjunctive propositions. (Cf. Jaegwon Kim's remark about properties in "Concepts of Supervenience": "such operations as infinite conjunctions and infinite disjunctions would be highly questionable for predicates, but not necessarily for properties -- any more than infinite unions and intersections are for classes" (Supervenience and Mind, p.73) -- though Kim is defending infinite conjunctive properties against a charge of complexity and artificiality, not of nonexistence.) But even if we are comfortable with infinite conjunctions in general, we may have special reservations about a conjunction of all truths. Is this proposition one of its own conjuncts? The conjunction of all truths is true, so it would seem that it must be (assuming now that conjunctive propositions have conjuncts, *contra* theories according to which sentences but not propositions exhibit the relevant sort of structure). An unusual proposition! Still, there is nothing wrong with being unusual, and I am hard pressed to find any other charge to bring against it.

[47] If indeed there is such a proposition distinct from the conjunction of all truths.

To see that no two propositions true in π are inconsistent, let T be the proposition *α is actual*. ('T' is for 'Truth, the Whole Truth, and Nothing But the Truth'.) Its negation ~T is itself possibly true, and is of course consistent with any proposition necessarily equivalent to it. Naturally each proposition that is true in α is consistent with every other proposition true in α. Hence there are two inconsistent propositions in π if and only if one of the propositions true in both α and π is inconsistent with ~T.

Now suppose that some proposition P and ~T are inconsistent; it is not possible that both P and ~T be true. ~T is true in every possible world but α, so there is no possible world aside from α in which P is true. Either α is the only possible world in which P is true or P is true in no possible world. If the former, then P is necessarily equivalent to T, so $P \notin B_\pi$. If the latter, then P is necessarily false and $P \notin B_\pi$. In either case $P \notin B_\pi$, so every member of B_π is consistent with ~T. Thus ~T and its equivalents function as the preface which denies that all of the other propositions of π have it right, but without contradicting any one of them.

Of course, there is nothing special about our choice of α as the possible world from which π inherits most of its contingent propositions. Take the book on any possible world W, replace *W is actual* and its necessary equivalents by their negations, and you will have the book on an impossible world without inconsistent pairs. The world is W's *preface world*. This world's preface will be a locus of impossibility.

ι (?): Is it possible to generalize this result? That is, for what n is it the case that there is an impossible world such that no n propositions true in that world are inconsistent with each other? I am not sure. My guess is that the greatest such n is either 2 or some infinite cardinal. If the former, then every impossible world has some inconsistent triple of propositions true in it.

Certainly π and its ilk have this feature: the proposition *α is not actual* and any two propositions (not necessarily equivalent to *α is actual*) whose conjunction is the conjunction of all truths together form an inconsistent triple in B_π.

How might we attempt to construct an impossible world whose book contains no inconsistent triple? Lewis, though no fan of impossible worlds, mentions what he takes to be an example in "Counterfactuals and Comparative Possibility". Presumably in some possible world Lewis is exactly seven feet tall. If we are inclined to deal in impossible worlds, says Lewis, we may also say that there is an impossible limit-world ι in which *Lewis is over 7 feet tall* is true, and so are each of *Lewis is less than 7.1 feet tall*, *Lewis is less than 7.01 feet tall*, *Lewis is less than 7.001 feet tall*, and so on. Though it is impossible that all the propositions of this world be true, says Lewis, any finite subset of them is true in some possible world. If he is right, ι not only lacks inconsistent triples, but also inconsistent *n*-tuples for all finite *n*.

It is clear enough that any finite subset of {*Lewis is over 7 feet tall*, *Lewis is less than 7.1 feet tall*, *Lewis is less than 7.01 feet tall*, *Lewis is less than 7.001 feet tall*, ...} is a consistent set. The difficulty is that these are not the only propositions true in the world ι, if such a world exists. Can we guarantee that *all* finite subsets of B_ι are consistent? Can we guarantee that all finite subsets of any maximal class of propositions are consistent? It is not clear.

To begin we need to see that any impossible world whose book contains no inconsistent pair (and *a fortiori* any impossible world whose book contains no inconsistent triple) must have at least this much in common with π: no proposition of the form *W is actual* (where W is a possible world) nor any proposition necessarily equivalent to one of these is true in that world. Let us call a world with this property "anonymous". The present claim, then, is that every world whose book contains no inconsistent pair is an anonymous world.

Proof: Let W* be an impossible world such that B_{W^*} contains no inconsistent pair of propositions. Let W be a possible world, and suppose that proposition P is necessarily equivalent to *W is actual* and is a member of B_{W^*}. (I.e., suppose that W* is not anonymous.) Then for any proposition Q true in W, Q is true in W*. (If not, then ~Q is true in W*; but ~Q is inconsistent with P, so ~Q and P would form an inconsistent pair in B_{W^*}.) So $B_W \subset B_{W^*}$. But then for any proposition R true in W* but not true in W, its negation ~R is true in W (since W is maximal), and so the inconsistent pair R and ~R belongs to B_{W^*}, contrary to supposition.

If an anonymous world is to avoid having an inconsistent triple, it needs somehow to get around inconsistent triples of the sort that plague π. One kind of inconsistent triple was mentioned above. Here is another. Each of the propositions *α is not actual*, *Humphrey did not win the election*, and *α is actual or Humphrey won the election* is true in π, and the three form an inconsistent triple. Since any world whose book has no inconsistent triple is anonymous, there is no question of removing the triple by replacing *α is not actual* with *α is actual*. Might we stipulate that *Humphrey did not win the election* (along with propositions necessarily equivalent to it) is not true in our anonymous world, its negation being true there instead? Well, we might, but this does not lead to a promising general strategy. *Every* contingent truth yields a similar inconsistent triple in π, and clearly a world such that each contingent falsehood is true in it will have many inconsistent pairs. Likewise it will not work to replace *α is actual or Humphrey won the election* and it necessary equivalents by their negations. Unfortunately for this strategy, every contingent truth (aside from *α is actual* and its necessary equivalents) is equivalent to the disjunction of *α is actual* with some contingent falsehood. The general strategy of replacing all propositions of this form, then, would force us to replace each contingent truth with its negation, again resulting in many inconsistent pairs.

So if the alleged world ι is meant to be one whose book differs minimally from that of some possible world, it will contain many inconsistent triples like those found in π. If ι is to avoid all these inconsistent triples, its book must be adjusted to differ from B_π (and any similar book) at a great many points, and so far it is not clear whether this can be done without creating new inconsistent triples.

It turns out we can say a bit more about what an impossible world without inconsistent triples would have to be like. Let τ be such a world. As noted above, for every possible world W, the proposition *W is actual* and its equivalents will not be true in τ. So if α and β are both possible worlds, *α is not actual* and β *is not actual* are true in τ. What of the proposition *either α or β is actual*? If it were true in τ, it would form an inconsistent triple with *α is not actual* and β *is not actual*, so it is not, and its negation is. In general, then, if W_1 and W_2 are possible worlds, the proposition *either W_1 or W_2 is actual* and its equivalents are not true in any impossible world without inconsistent triples.

But we may go beyond the two-world case: if γ is a third possible world, then the proposition *either α or β or γ is actual* cannot be true in τ since it is inconsistent with *neither α nor β is actual* and *γ is not actual*. And so on: the general result is that for any finite *n*, no proposition that is true in exactly *n* possible worlds is true in an impossible world without inconsistent triples. Let us say that impossible worlds without inconsistent triples are thus "finitely anonymous". What we have shown is that the members of a certain class of contingent propositions cannot be true in impossible worlds without inconsistent triples, viz., those true in only a finite number of possible worlds.

Finally, one last result which may be of interest in the search for an impossible world without inconsistent triples. So far we have seen examples of worlds with inconsistent pairs but no necessary falsehoods, and of worlds with

inconsistent triples but no inconsistent pairs. Might there also be worlds with inconsistent quadruples but no inconsistent triples, and so on? Here we may give a firm 'No'. For any finite $n>3$, if a world has an inconsistent n-tuple, then that world also has an inconsistent triple.

Proof: Suppose for reductio that some world W has no inconsistent triple but does have an inconsistent n-tuple (where n is finite and greater than 3). Let $\{P_1, P_2, P_3, ..., P_n\}$ be one such n-tuple. If B_w did not contain $P_1\&P_2$, then by maximality it would contain its negation $\sim P_1\&P_2$, which forms an inconsistent triple with P_1 and P_2. So B_w does contain $P_1\&P_2$. Then W has an inconsistent (n-1)-tuple, $\{P_1\&P_2, P_3, ..., P_n\}$. We have proved an inductive principle: every world with an inconsistent n-tuple also has an inconsistent (n-1)-tuple (where $3<n<\omega$). By repeated applications of this principle, we can prove that W has an inconsistent triple, contrary to supposition. Hence no world has an inconsistent n-tuple ($3<n<\omega$) and lacks an inconsistent triple.

If there are impossible worlds without inconsistent triples to be found, then, they must be worlds with the property that Lewis claims for his impossible limit-worlds: no contradiction can be derived from the propositions true in such a world, since every finite set of propositions true in it which might serve as premises is a consistent set. As it stands, we have neither proof that all impossible worlds have inconsistent triples, nor a construction of a world which lacks them. The only hope for the latter would seem to be a world which differs from π not by systematic removal of its contingent truths, but by removal of some and preservation of others. In chapter 5 we will see a strategy for proving that some impossible worlds lack inconsistent triples, but I must forego the details until then.

So much (for the nonce) for the mathematics of impossible worlds. Besides the results themselves, what lessons can we glean from all of this? Well, for one thing, there are impossible worlds with very unusual and

unexpected properties. For another, certain worlds with relatively small loci of impossibility do indeed turn out to be helpful in proving results about impossible worlds; we will see later that it is important in a number of contexts to keep these impossibilities in mind. And for another, worlds with finite loci of impossibility illustrate as starkly as possible the thesis that the books of states of affairs are not all closed under entailment. In B_δ the proposition *Mars is blue* stands alone; many of the immediate implications of *Mars is blue* are not true in δ. This makes δ rather unlike most of the impossibilities we consider, and, but for this example, we might well have overlooked many of the impossible worlds which must exist if books in general lack closure.

Easily Imagined Worlds

The previous section uses a precise and, in important respects, the most illuminating way of characterizing impossible worlds. But it will also be useful for us to consider a much more "natural" group of worlds which cannot be described quite so exactly. These are the worlds most easily imagined. The sorts of inconsistencies in their books may be much better hidden than rather stark discontinuities of worlds like δ. The parts of these worlds to which our attention is drawn may have large areas of "local consistency" which give at least the feel of a possible world.

Science fiction often calls our attention to impossible situations, even if not to situations complete enough to be worlds. However, we may consider the impossible worlds which fill in the gaps. These worlds supply what fiction omits and supply it in a way that is, so to speak, as possible as possible.

In Douglas Adams's novel Dirk Gently's Holistic Detective Agency, for example, we find this informal theory of time travel:

> "But that can't work, can it?" said Richard. "If we [undo a past event via time travel], then this won't have happened. Don't we generate all sorts of paradoxes?"
>
> Reg stirred himself from thought. "No worse than many that exist already," he said. "... It's like a human body, you see. A few cuts and bruises here and there don't hurt it. Not even major surgery if it's done properly. Paradoxes are just the scar tissue. Time and space heal themselves up around them and people simply remember a version of events which makes as much sense to them as they require it to make.
>
> "That isn't to say that if you get involved with a paradox a few things won't strike you as being very odd, but if you've got through life without that already happening to you, then I don't know which Universe you've been living in, but it isn't this one." (228)

Let's suppose that what Reg says regarding time travel and its paradoxes is, according to the novel, the truth of the matter. Then some of the propositions true according to the novel are these: *it is possible to change the past via time travel*, *some contradictions are true*, and *a time-traveller may remember events which, due to changes of the past, never occurred*. Of course many other propositions will be supplied by the plot of the novel. Not all propositions will be supplied, however, since even propositions that are entailed by propositions true according to a work of fiction need not be true according to that work. Rather, the situation is like Frege's. Not everything is true according to Frege, even though one necessary falsehood ("All the axioms of this system are true") is true according to him.

Whether or not time travel is possible, the theory sketched in the quotation cannot possibly be true; here we have a case of impossible fiction. How might an impossible world fill in the details that the novel does not express? There are many ways. Of each proposition-negation pair, we might select one at random to be true in the world if neither is already expressed by the story. Or, to preserve consistency as best we can, we might select some possible world and stipulate that what the story leaves unspecified is to be

specified by this possible world. But if what the story specifies is inconsistent, this method may give us many glaring inconsistencies between what the story specifies and what the possible world specifies. So we might instead specify what the story does not inconsistently, with a result that is more subtly inconsistent than the worlds of the previous method.

The best ways of filling in the details will be difficult to specify with any precision. What is wanted is a world without any *obvious* inconsistencies. Even better: we want a world with as few obvious inconsistencies as we can manage, and with as little obviousness as we can manage. Why, exactly, is obviousness important here? When fiction calls our attention to a state of affairs, we want to enjoy whatever paradoxes the author meant to include while suspending our disbelief in whatever other contradictions pop up. If these contradictions aren't obvious, if they are contradictions relatively unlikely to spring to mind, then in imagining such a state of affairs one's disbelief may not even arise in the first place. In general, the inconsistencies of the most easily imaginable states of affairs will be non-obvious inconsistencies.

Consider the story told in the novel mentioned above. The story provides us with a list of propositions (not necessarily those most naturally expressed by the declarative sentences of the novel) true in or true according to the story. A few of these are mentioned above. Suppose we were to fill out the list of propositions true according to the story by supplying, where needed, a true proposition. Samuel Taylor Coleridge is a character in the novel, but it is neither true in the novel that Coleridge had blue eyes or that Coleridge did not have blue eyes. Since one of these two propositions must be supplied if we are to end up with a world, we stipulate that the world under construction contain whichever of the two is true. If we continue in this way, we will end up with an impossible world.

Though it is clear that the collection of propositions true according to an inconsistent story is not closed under entailment, it is plausible that the propositions clearly entailed by the propositions explicitly asserted by the story are also true according to it. *It is possible to change the past via time travel* is, perhaps, not explicitly expressed by any sentence of the above story, but it is rather clearly true according to it. *The past has occasionally been changed* is a similar example. However, *cigarettes reproduce by mitosis when no one is looking* is evidently not true in the story, even though it is entailed by *some contradictions are true*. The entailment is non-obvious, at lest in the sense that it does not immediately spring to mind, and nothing in the story calls attention to it. So our procedure of adding true propositions to those true according to the story dictates that we add *cigarettes do not reproduce by mitosis when no one is looking* rather than its negation. This procedure clearly yields a world more easily imagined than the impossible world closed under entailment, i.e., λ.

Still, we can do better. The world that results from this procedure is not the one we are most likely to regard as the world of the story. Consider the proposition *Dirk Gently has two eyes*. Suppose no sentence of the novel expresses this proposition, and suppose no proposition true according to the story entails it. Then, since there is in fact no Dirk Gently, the proposition is false and the procedure will add it negation to the world. The negation of *Dirk has one eye* will be added for similar reasons. The story does entail *Dirk is not blind*, and so this proposition will be true in the constructed world. Now, there's nothing *wrong* with these particular propositions being true in an impossible world. It's just that when we read the novel, we tend to imagine a state of affairs in which Dirk has two eyes, and so this sort of state of affairs seems a better match for the story.

We can handle this fact in two ways. First, we could claim that (contrary to what we supposed above) *Dirk has two eyes* is true according to the story (and thus true in our target world) despite the fact that the proposition is neither expressed by any sentence of the novel nor entailed by any proposition so expressed. The idea would presumably be that certain of the author's or the audience's background assumptions are true according to the story. The idea has some plausibility, though of course it would be difficult to say exactly which background assumptions are true according to the story and which are not. The other approach is simply to give up the procedure we have been discussing in favor of some method more accommodating to our imaginations. In either case, we lack (so far) any precise way of characterizing those worlds which seem best suited to fill out inconsistent stories in as plausible a way as possible.

The point is worth making in part because such worlds are apt to turn up in applications of impossible worlds. Without going into too much detail, the account of counterfactuals discussed in the next chapter considers the impossible worlds most similar to the actual world, where what counts as similar depends a great deal on context. It will frequently turn out that "obvious" inconsistencies tend to count against similarity. So the impossible worlds most similar to the actual world in the relevant ways will be those with the fewest "obvious" inconsistencies--i.e. the easily imagined impossible worlds.

A Word About Individuation

States of affairs, impossible worlds included, are to be individuated by content, and the content of states of affairs varies with the propositions that are true in them. How we individuate states of affairs, then, depends on how we individuate propositions. That propositions aren't to be individuated by

entailment relations alone is, I think, quite plain. There is no shortage of distinct, necessarily equivalent propositions--to borrow Perry's example,

George is sleeping

and

George is sleeping and Mary is weeping or Mary is not weeping.

If propositions were to be individuated by entailment relations, then there would be only one necessary truth and only one necessary falsehood.[48] One important feature of these examples is their intentional differences. The propositions are about different things.

The Stalnakerian (for one) might protest that here, if anywhere, we have occasion to apply Ockham's razor. We are not to multiply entities beyond necessity, and indeed, we should relish this opportunity to shave off what we apparently can do without. I am convinced that some version or other of Ockham's razor is a principle that must be observed in the formulation of theories. But there are difficulties of application. What counts as beyond necessity? We begin to see an answer if we reword the principle slightly: Do not multiply entities unless it is necessary for an optimal theory. Whether or not Ockham's razor should be used depends on the merits of the rival theories being considered, and of course those merits may well be the subject of disagreement. Earlier we saw that both Quine and Lewis employ the language of a cost-benefit analysis when evaluating an ontology. That metaphor is

[48] It does not follow from this that there is only one impossible state of affairs. If in this case states of affairs had books of the sorts I have described, there would still be some impossible states of affairs whose books contained the necessary falsehood and others whose books did not. However there would be an unacceptable conflation of states of affairs, e.g. of the states of affairs whose books are {*the Superbowl pregame coverage begins during the second half*} and {*there is a Mobius strip with two sides*}. Of course those disinclined to distinguish the propositions *the Superbowl pregame coverage begins during the second half* and *there is a Mobius strip with two sides* are unlikely to find themselves with strong intuitive grounds for distinguishing these states of affairs. In fact the state of affairs case is rather less clear, since it is not obvious that when we speak of "the state of affairs *there being a Mobius strip with two sides*" we refer to the state of affairs whose book is {*there is a Mobius strip with two sides*} and not some larger state of affairs. So although Stalnaker's view of propositions has consequences regarding states of affairs that I do not accept, the stronger reasons for rejecting Stalnaker's view have to do with propositions themselves.

fitting. In particular, it makes it clearer that the razor is not to be used in every case where some theory with fewer entities is available; the alternative theory needs to be competitive, at the least.

In this case, it seems clear to me that a good theory of propositions will sometimes distinguish between necessarily equivalent propositions, since these sometimes differ in subject matter. The Stalnakerian will disagree. Ockham's razor is legitimately applied only if I am wrong about what makes for a good theory of propositions. So no matter who is right, one cannot appeal to Ockham's razor in order to settle this dispute. To use the razor is to have already decided the crucial issues.

But even if we adopt fine-individuationism, a host of questions about proposition individuation remain. There are many cases where different sentences express propositions which are, if not identical, at least necessarily equivalent and (arguably) about the same things.

QUIZ

Part I: Which of the following pairs of sentences express the same propositions, and which express two different propositions?

1.) All ravens are black.
All non-black things are non-ravens.

2.) Figure C is a circle.
Figure C is the locus of points in a plane a given distance from some particular point in the plane.

3.) Jim loves Mary.
Mary is loved by Jim.

4.) Cicero was a Roman orator.
Tully was a Roman orator.

5.) Hesperus is identical with Phosphorus.
Phosphorus is identical with Phosphorus.

6.) Mars is red.
It is true that Mars is red.

7.) It is true that Mars is red.
The proposition *Mars is red* is true.

Part II: Which of the following pairs of propositional forms yield a single proposition when propositions are substituted for P and Q, and which yield two different propositions?

8.) P
~~P

9.) P&Q
Q&P

10.) P
P&P

11.) P
P∨P

Extra Credit: Is the proposition *α is actual* identical to the conjunction of all truths?

Which answers to the quiz questions are correct clearly has bearing on the individuation of states of affairs. Is *Jim's loving Mary* distinct from *Mary's being loved by Jim*, or not? Suppose that Jim does love Mary, and let μ be the impossible world such that $B_\mu = B_\alpha \cup \{$*Jim does not love Mary*$\}$. If ν is such that $B_\nu = B_\alpha \cup \{$*Mary is not loved by Jim*$\}$, are μ and ν identical, or do we have two distinct impossibilities?

So questions of proposition individuation are important for one's views on the individuation of impossible worlds. Nonetheless, there is a sense in which these difficult questions need not be answered at all. For although a reasonably complete theory must, perhaps, say *something* about the identity conditions of impossible worlds, it need not give a ruling on every difficult case that arises. It may say instead that if the proposition *Jim does not love Mary* is identical with

Mary is not loved by Jim, then μ and ν are identical; and if not, they are distinct. If a theory specifies that the individuation of states of affairs depends on the individuation of propositions in the manner illustrated above, it says enough. As far as I can tell, nothing about the theory of impossible worlds itself enables us to say exactly which way of individuating propositions is best, and so the course of wisdom may be to remain silent, allowing the opinionated to fill in the details as they will. The quiz will be ungraded.

Impossibility and Nonsense

One general strategy for arguing against my view of impossible worlds is to attempt to produce an impossibility which is not among the impossible state of affairs I describe. For example, if one could make a convincing case or unveil a sufficiently strong intuition that there were a certain impossibility θ and show that no collection of propositions whatsoever could be true in θ, then my view would have to be amended (since I say some collection of propositions is true in any given state of affairs and that every impossibility is an impossible state of affairs).

Vagueness, for example, might lead to such an objection. Arguably, since there are states of affairs such that *John is bald* is neither true nor false in them, neither collections which include the proposition *John is bald* nor those which exclude it could be the books of such states of affairs. Rather, books would need to be fuzzy sets.[49] I don't really want to address this objection here, in part because it objects to a certain way of thinking about states of affairs in general, and not only impossible ones.

The objection I do want to address has to do with nonsense. Suppose someone argued as follows:

[49] Jonathan Kvanvig has suggested in conversation that my theory might usefully be amended so that it does not assume bivalence. The vagueness objection might be one motivation for doing so.

> 'Brillig', 'slithy', and so on, though they have connotations, are nonsense terms. In contrast, 'unicorn' and 'octarine' (a fictional color in Terry Pratchet's Diskworld novels) have sense, and they can be parts of meaningful sentences like "There is an octarine unicorn." Arguably, this sentence is possibly true. But "the slithy toves did gyre and gimble in the wabe" is meaningless, and so cannot aspire to possibility. What is not possible is impossible, so it is an impossibility that "the slithy toves did gyre and gimble in the wabe." According to your theory, every impossibility is a state of affairs, and every state of affairs has a book. But there is no collection of propositions that could serve as the book of this impossibility. Any suitable collection would have to contain a proposition expressed by "the slithy toves did gyre and gimble in the wabe," but there is no such proposition.

My hope is that no serious philosopher is at all tempted by this argument. Let's have a look at what's wrong with it.

The fundamental problem is that the "what is not possible is impossible" line of reasoning leads to the conclusion that "the slithy toves did gyre and gimble in the wabe" (like "yanga langa furjeezama" and all other nonsense phrases) expresses a necessary falsehood--even though the argument explicitly says the phrase is meaningless and so, presumably, expresses no proposition at all. Just as "oskjj jgtmw" is neither true nor false (i.e., it expresses neither a true proposition nor a false one), it is neither possible nor impossible (i.e., it expresses neither a possibly true proposition nor a necessary falsehood). And of course the same point holds for states of affairs. From the meaninglessness of a phrase we cannot infer that it denotes an impossible state of affairs, since it might not pick out any state of affairs at all.

That's the easy case. The same lesson can be applied to the slightly more difficult cases where it is unclear whether a statement is meaningful or not. Interpretation of Russell is sometimes difficult, but on at least one understanding of the *Principles*, Russell says that the denoting concept *any man* denotes one man, but doesn't denote a particular man. That is, *any man* denotes an ambiguous man. If this is in fact what Russell is saying, there are a

couple of ways we might diagnose his claim's shortcomings. We might say that since each man is a particular man, and necessarily so, what Russell says is necessarily false; necessarily, there is no such thing as an ambiguous man. Perhaps less generously, we might say Russell's confusion is so great that his "theory" is simply nonsense, and fails to express any proposition at all about what *any man* denotes. It is as if he had said, "*Any man* denotes floo mo pum baa." The words "ambiguous man," though they may initially sound meaningful, do not in this case combine with others to form a meaningful sentence.

Whichever of these criticisms one prefers, one ought not be fooled into thinking that Russell's theory is true only in a state of affairs which cannot be assigned a propositional content. On the one hand, if the theory is meaningful, then it expresses certain propositions, and exactly those propositions are true in some (possible or impossible) state of affairs. On the other hand, if the theory is meaningless, then it lacks propositional content (or at any rate the problematic part of the theory does). But then there is little reason to think that "*the denoting concept* any man*'s denoting an ambiguous man*" picks out a state of affairs. The phrase inherits the meaninglessness of the theory, and so it fails to pick out anything.

In short, meaning goes hand in hand with propositional content, and propositional content goes hand in hand with states of affairs. The result is that meaningless phrases do not pick out states of affairs and thus do not constitute counterexamples to the theory of states of affairs and propositional content that I have presented. Even if we are unsure about a particular case--we don't know, say, whether the sentence "Sean is dancing about architecture" is meaningful or not--we must not think that perhaps the sentence is meaningless but that "*Sean's dancing about architecture*" is meaningful and thus picks out a state of affairs.

CHAPTER 4: COUNTERPOSSIBLES

This chapter and the next address some of the philosophical applications of impossible worlds, beginning with what may be the most important one: the use of impossible worlds in counterfactual semantics. The first four sections below elaborate and defend my theory. The remainder of the chapter raises some issues surrounding the context-sensitivity of counterfactuals and an objection based on the idea of similarity. The theory to be offered owes a great deal to David Lewis's work--even the replies to objections, since the objections are much like those Lewis's theory first met.

False Counterpossibles

A counterpossible is a counterfactual with an impossible antecedent, such as *if God were vicious, then the world would contain much more evil than it does* or *if Hitler had time-travelled into the past and killed himself as an infant, then World War II would not have occurred*. Counterpossible assertions are common both in philosophical arguments and in everyday speech; I have already used several in this dissertation. On a standard account of counterfactuals, counterpossibles are all trivially true. One representative of this view is Lewis, who argues that if we were to suppose that an unentertainable proposition were true, we might just as well suppose that anything at all is true. There are counterpossibles which we would not normally assert in conversation, such as *if there were a largest prime p, then there would be six regular solids* or *if there were a largest prime p, then pigs would fly*. But,

says Lewis, we would not confidently deny them either. And so he is content to let all counterpossibles come out true.[50]

Lewis's analysis of counterfactuals (of which my account will be a variation) is equivalent to the following.

> (1) A counterfactual is (nontrivially) true iff some possible world in which both the antecedent and the consequent are true (an A C world) is more similar to the actual world than every possible world in which the antecedent is true and the consequent is false (an A ~C world).

Since the antecedent of a counterpossible is not true in any possible world, (1) itself does not return a truth value of 'true' for any counterpossible. Lewis says that all counterpossibles are *trivially* true, for the reasons given above. Hence counterpossibles must be treated as a special case. Different reasons are given for the assignment of the truth value of a counterpossible than are given for the assignment of the truth value of a counterfactual with a possible antecedent.

Linda Zagzebski argues in "What If the Impossible Had Been Actual?" that a correct account of counterfactuals ought not assign all counterpossibles the value 'true'. She presents a number of counterpossibles that appear quite clearly to be false, such as *if I were to go backwards in time and change my lecture last week, then I would not reach the same moment of time twice* (166). To Zagzebski's examples I add the following sad tale. Some time ago my checking account balance was (roughly) $400. One day my credit card bill arrived, so I wrote the Visa Corporation a check for $150 and, because my mind was occupied with less mundane matters, recorded a new balance of $350. The next

---

50 Though Lewis does not think that we want to assert that any counterpossibles are false, he does admit that his reasons are less than decisive and offers an alternative analysis which makes all 'would' counterpossibles false and all 'might' counterfactuals true (Counterfactuals, p.25). Elsewhere he suggests another approach, viz., extending the similarity relation between possible worlds to include "impossible worlds where not-too-blatantly impossible antecedents come true" ("Counterfactuals and Comparative Possibility", p. 19). I will propose an account much like the latter but which includes all manner of impossible worlds and not merely the "nice" ones.

week I received the overdraft notice and along with it the inevitable $25 service charge. At that point it would have been correct for me to remark, "If 400 minus 150 had been 350, I wouldn't have overdrawn my checking account." Not only would this have been correct; the remark would have been nontrivial. For suppose I had said, "If 400 minus 150 had been 19.95, I wouldn't have overdrawn my checking account." Surely this is false! If 400 minus 150 had been 19.95, I would have overdrawn my checking account for sure. (We'll consider an objection to this argument in the next section.)

There does seem to be something right about Lewis's supposition that we might just as well suppose that anything is true if we suppose the unentertainable is true, but it also seems that even obviously impossible propositions are entertainable. The above is an example of how one might very reasonably entertain such a proposition. In support of his assignment of 'true' to all counterpossibles, Lewis adds that we would not confidently deny counterpossibles, and so there is little harm in supposing that they are all true. But when we consider counterpossibles like: *if I were to create,* ex nihilo, *one duck every morning, then I would create only five ducks each week,* or: *if the number five grew wings and webbed feet, it would be no different than it is now*, then the truth of all counterpossibles does not seem so harmless after all.

And not all of the examples are silly ones: philosophers frequently use counterpossibles in arguments and treat them as nontrivial truths. Galen Strawson, for example, has argued that human freedom is impossible since if one ever acted freely, that act would be preceded by an infinite number of choices. Whether or not we agree with Strawson about freedom, it is clear that he regards the conditional *if a person ever acted freely, that person would first have made an infinite number of choices* as a counterpossible. And Strawson presumably takes the conditional to be nontrivial (in the semantic sense at issue, as opposed to an epistemic sense). Since he wants to highlight a

relationship between freedom and an infinite series of choices, he would deny the conditional obtained by negating the consequent.

Another example: this one borrowed from Zagzebski, who gets it from van Frassen's discussion of Kant's First Analogy. Van Frassen writes:

> Suppose, however, that all substances cease to be and other substances whose states are not simultaneous with any states of the former come into being. The way in which we have phrased this supposition suggests that the other substances exist after the former. But close scrutiny will show that this is not entailed: there is no ground for asserting *any* temporal relation between the states of the former and those of the latter, except nonsimultaneity. So there would be no way of ordering them all together into a single world history. Since we suppose that such an ordering is always possible, this supposition is absurd.[51]

Zagzebski comments:

> A crucial step in this argument is the assertion of the counterpossible: *If all substances ceased to be and other substances came into being, there would be no way to order them all into a single world history*. ... [T]he counterpossible is neither denied nor reinterpreted in a trivial sense once it is concluded that the antecedent is impossible, and its nontrivial truth is necessary to get the desired conclusion. (181)

Whether or not we agree that the given examples are counterpossibles, a large number of philosophical positions and arguments presuppose that counterpossibles are sometimes nontrivially true, a fact which testifies at least to the significance of the question. Lewis notwithstanding, the idea that all counterpossibles are true has some serious ramifications.

Nonetheless it is true that there are many counterpossibles that we would not confidently deny, especially among those whose antecedents and consequents are unrelated. I think the reason is this. Counterfactual sentences

[51] An Introduction to the Philosophy of Time and Space, p. 48.

are especially sensitive to context.[52] It may be true that if Caesar had been in Korea he would have used catapults--but not if we are discussing how various personalities of history would have deployed the various weapons *actually available* during the Korean War. Before it can be determined whether a particular counterfactual is true or not, the evaluator must understand what sort of context the counterfactual is being asserted in; this involves understanding why it is asserted, what features of reality it is meant to point out.

Often the context is clear from what has already been said. At other times, the counterfactual sentence itself is the only clue as to what sort of context it assumes. In such circumstances a principle of charity takes effect: the one who asserts a counterfactual presumably means it to be true, and so the listener assumes that the context is one which makes the assertion come out true, provided the assertor might plausibly have had some such context in mind. When someone makes the lone statement, "If that guy had been any taller he would have hit his head on the ceiling fan," the charitable listener assumes that the assertor is primarily concerned with the height of the guy in question, discounting the likelihood that if he had been any taller the guy would have been more careful about where he was standing. In effect, the listener chooses which similarity relation is to be operative within the analysis (1) in a manner that makes the statement express a truth.

Now when a counterpossible sentence is uttered, there is often a plausible context which would make it true. It is frequently reasonable to think that the assertor may have in mind the principle that if some unentertainable proposition were true then anything you like would be true. This is especially

[52] Or perhaps not context per se, but something that frequently varies with context, such as the intentions of the speaker or the speaker's assumptions about which aspects of reality are now most relevant. I will continue to use the word 'context' to maximize the fit with Lewis's terminology, but this may sometimes require us to take the word in a fairly broad sense. For example, on this usage it is possible for "context" to change suddenly in the middle of a conversation with a slight change of subject or emphasis.

the case when the antecedent and consequent do not appear to be relevant to each other. If one heard the serious assertion that if there were a largest prime p then pigs would fly, one would tend to take this as a way of saying a largest prime is out of the question, i.e. that the antecedent is, for present purposes, "unentertainable." Since this sort of interpretation of a counterpossible is quite often plausible, we hesitate to deny it; we do not want to deny what is true on a plausible interpretation. And if we think too hard about a given counterpossible sentence or utterance, we begin to lose sight of its context, which may give us the reasons we need for thinking it false. So we do hesitate. Nevertheless, *contra* Lewis, there are contexts in which counterpossibles are simply false. Our reluctance to deny many counterpossibles is not enough to justify assigning the value 'true' to all of them.

In her article Zagzebski is primarily interested in counterfactuals whose antecedents are not *explicitly* contradictory. She reports that she does not have a strong reaction to the claim that both of

> *if it were both raining and not raining here at this moment, then I would be the Pope*

and

> *if it were both raining and not raining here at this moment, then I would not be the Pope*

are true, as they must be if the standard account is correct. So Zagzebski focuses instead on counterfactuals with antecedents that are not explicitly contradictory. It seems to me, though, that even this sort of antecedent need not make a counterfactual trivially true. A contradiction would be true if it were both raining and not raining here at this moment, and to say so is not an empty claim; it is false that if it were both raining and not raining here at this moment then no contradiction would be true. Thus even counterfactuals with explicitly contradictory antecedents may turn out to be false.

The Objection

But there is a fairly obvious objection which needs to be addressed. If P entails Q, isn't that sufficient for the truth of the counterfactual that has P as its antecedent and Q as its consequent? In other words, isn't

(R) $\underline{A \rightarrow C}$
$\therefore A \mathbin{\square\!\!\rightarrow} C$

a logically valid rule of inference? If so, and if a necessary falsehood entails every proposition, then every counterpossible is true, despite the fact that there are a variety of counterpossible statements we want to deny. According to this objection, the proposition *if God were vicious, things would be great*[53] is true. Whatever inclination we might feel to deny the proposition is either (a) simply mistaken, or else (b) a confused apprehension of the fact that the proposition's opposite, *if God were vicious, things would not be great*, is somehow "more revealing".[54]

Lewis, of course, does not share my view of counterpossibles. Nonetheless I am going to propose a Lewisian reply to this objection, since I believe Lewis's theory of counterfactuals provides the resources needed to overcome the objection.

[53] I borrow the example from Jonathan Strand.

[54] Tom Morris and Chris Menzel, "Absolute Creation," 355. Morris and Menzel suggest that an approach similar to this last is an option for those who believe in one-way dependence between necessarily extant beings, for example, a dependence of abstract objects on God. They say one might "acknowledge a *logical* dependence running both ways between God and abstract objects ... and nevertheless maintain that there is a *causal* or ontological dependence running in only one direction ...." (355). On this view, though it is true that if there were no abstract objects there would be no God, the counterpossible *if there were no God, there would be no abstract objects* is "more revealing".

Briefly, the reason I want to reject this path is that the only notion of causal or ontological dependence that I have is one that involves this so-called "logical" dependence. I don't know what it would be for, say, event A to be the sole, sufficient cause of event B if both the occurrence and non-occurrence of A counterfactually implied B. (An exception: we do seem to have this sort of situation in the Frankfurt-style counterexamples to the Principle of Alternate Possibilities, where an alternate cause of B would be triggered by the non-occurrence of A, but such cases are exceptional and the dependence of abstracta on God would not be one of them.) Lewis apparently shares my intuition, since he makes counterfactuals central to his account of causation. Scott Davison, commenting on Morris and Menzel's alternate strategy, remarks that "it is very hard to *understand* the [view that dependence is something other than counterfactual dependence], let alone *argue* for its preferability" (493).

First we need to note what Lewis has to say about context-sensitivity. Let's shift our attention from propositions to sentences and consider this famous pair from Quine:

> (C1) If Caesar had been in command [in the Korean War] he would have used the atom bomb.
>
> (C2) If Caesar had been in command he would have used catapults.

Lewis mentions the pair to illustrate how two kinds of influence of context might be called upon to explain the apparent truth of conflicting counterfactuals. (Let us assume the pair refers to Caesar's exclusive, primary strategy, so that there really appears to be a conflict between the two.) On one theory context serves to supply a part of the antecedent which has been left implicit. The antecedents of the pair might implicitly specify that Caesar has the latest technology available to him, so that (C1) is true and (C2) is false. In another context, (C1) and (C2) might express propositions whose antecedents specify that the weapons available to Caesar are the weapons that were in fact available to him, so that (C1) expresses a falsehood and (C2) expresses a truth.

Lewis prefers the theory that context has another kind of influence.

> [I]nstead of using context to restore the real antecedent from the explicit part of the antecedent, I could say that the explicit antecedent *is* the real antecedent and call on context rather to resolve part of the vagueness of comparative similarity in a way favorable to the truth of one counterfactual or the other. In one context, we may attach great importance to similarities and differences in respect of Caesar's character and in respect of regularities concerning the knowledge of weapons common to commanders in Korea. In another context we may attach less importance to these similarities and differences, and more importance to similarities and differences in respect of Caesar's own knowledge of weapons.[55]

[55] Counterfactuals, p. 67.

So there are several ways in which we might plausibly see context as resolving a tension among our intuitions about counterfactual pairs like the above. Whether we prefer one kind contextual influence or the other or some combination of the two, sentences (C1) and (C2) each express one proposition in one context and a different proposition in another.

Furthermore, as Lewis observes elsewhere, wherever there is dependence on complex features of context,

> [t]here is a rule of accommodation: what you say makes itself true, if at all possible, by creating a context that selects the relevant features so to make it true. Say that France is hexagonal, and you thereby set the standards of precision low, and you speak the truth; say that France is not hexagonal (preferably on some other occasion) and you set the standards high, and again you speak the truth.[56]

Assert (C1), and you speak the truth. Or deny (C1), saying it's (C2) that is true instead, and again you speak the truth. In the case of counterfactuals, context affects not only standards of precision but also the relative weights of various respects of comparison. As a result, the fact that a given counterfactual sentence expresses a true proposition does not preclude the same sentence from expressing a falsehood in another context, as in the ceiling fan example above. And since context tends to select the similarity relations that make a counterfactual true, that false-making context may be harder to find. If I utter sentence (C1) and claim it expresses a falsehood, the objector may be able to claim persuasively that (C1) expresses a truth--and so it does. Nonetheless (C1) does express a falsehood in contexts that downplay the availability of the atom bomb and emphasize Caesar's actual knowledge of weaponry.[57]

---

[56] On the Plurality of Worlds, p. 251.

[57] It is sometimes important to notice that not every similarity between worlds is a matter of overlapping books. Sometimes worlds are similar in part because of some property shared by or relation between propositions true in them. For example, in a certain world the propositions of the form *S is mortal* may be true for each human S that exists in that world. In this respect that world may resemble the actual world--whether or not the proposition *all men are mortal* is true in it. This phenomenon will reappear in this chapter's final section, which deals with an objection based on the notion of similarity.

Another example. My roommate searches the refrigerator in vain for the leftovers. When he eventually finds the container right next to his leg, he says, "If it had been a snake, it would have bitten me," and he's right. And when after a moment he adds, "Of course if it had been a snake it would have been cold and sluggish," he's right again. Here we have two counterfactual statements in apparent conflict. But the conflict is merely apparent; in fact, the sudden shift in what counts as similar to actuality is part of what makes the second remark amusing. Here, as per the rule of accommodation, the first statement selects a context that makes the usual behavior of snakes in close proximity an unusually important respect of comparison. Then the second statement selects a context in which the actual location of the container (the refrigerator) and the effect of coldness on snakes are important respects of comparison.

I suggest that the same rule of accommodation applies to counterpossibles.[58] A single sentence may express one counterpossible proposition in one context and a different counterpossible proposition in another. So when someone tells me that the proposition *400 minus 150 equals 350* entails that I overdraw my checking account so that the proposition expressed by "if 400 minus 150 had been 350, I would have overdrawn" is true, I say, "You're right." Here the assertion comes in a context that calls attention to the entailment relation between the antecedent and consequent.

But that context is a rather different one than the original context of my checking account story. The story emphasized the connection between the result of the subtraction problem and the amount that I felt free to spend; in this context the entailment relation between antecedent and consequent is outweighed by other respects of comparison. In the context of the story, the sentence "if 400 minus 150 had been 350, I would have overdrawn" expresses a falsehood, a different proposition entirely than the one expressed by the same

[58] If we want to get rid of the appearance that the sentences in the snake example express counterpossibles, we may substitute the less idiomatic "If a snake had been where it is, ..."

sentence in the context that emphasizes the entailment of the consequent by the antecedent.

My position may cause the objector some frustration: I say a certain counterpossible sentence expresses a falsehood; the objector says it expresses a truth; and no matter how convincing the case, I say, "No, no, you're thinking of a different proposition." So I'm sorry about that. Still, I think it's the right thing to do. The phenomenon is no different from what happens when one person insists it's false that if Caesar had been in command he would have used the bomb and another insists that it's true. The conflict between them is illusory. And if Lewis is right about counterfactual sentences like (C1) and (C2), I would find it startling indeed if counterfactual sentences with impossible antecedents didn't exhibit a similar sensitivity to context.

One might complain that although shifts in context can bring certain counterfactual sentences to express truths on some occasions and falsehoods on others, there are limits.[59] Some sentences express truths in any context--any context, at least, that we are likely to produce. And, the complainant may add, entailment relations are important enough to us that counterpossible sentences are *all* of this sort. The fact that an antecedent entails its consequent will never be outweighed by some more important respect of comparison; entailment is the metaphysical bottom line.

In reply, I say that if we want to discover the extent of contextual influence, we should look to the data. That is to say, which counterfactuals are true and false will tell us something about how we are willing to rank various respects of comparison, not the other way around.[60] Entailment relations and necessary truths are always or almost always very important respects of comparison between worlds, but whether our purposes sometimes create

[59] Cf. Counterfactuals, p.93.

[60] Cf. Lewis's point in "Counterfactual Dependence and Time's Arrow," in Philosophical Papers, Vol. II, p. 42-43.

contexts in which other respects of comparison are more important than certain of these relations and truths must be determined by the way we want to assign truth values to counterpossibles. And our inclination to call some counterpossibles false is fairly strong; I suspect that even many of those who endorse the rule of inference (R) find the truth of all counterpossibles a somewhat counterintuitive result.

I contend, then, that the rule of inference (R) is invalid. Though in many contexts sentences of the appropriate forms express true propositions, there are contexts that assign the entailment of the consequent by the antecedent a relatively low degree of importance. If in such a case the entailment relation is outweighed by another respect of comparison, the counterpossible may be false.

It follows that counterfactual implication is not an intermediate between strict implication and material implication, i.e., that it is not the case that a strict conditional entails the corresponding counterfactual (the one with the same antecedent and consequent), which entails the corresponding material conditional.[61] The basis for this view is the supposed validity of the inference (R), which gets its appeal from consideration of an incomplete range of cases.

The Modified Account

A satisfactory account of counterfactuals, then, will allow a nontrivial analysis of counterpossibles. Fortunately (as both Lewis and Zagzebski hint) impossible worlds can be used to construct a slight modification of Lewis's analysis that does just that. Consider this revision:

> (2) A counterfactual is true iff some (possible or impossible) A C world is more similar to the actual world than every A ~C world is.

[61] Proponents of this view include Kvart, A Theory of Counterfactuals; Lewis, Counterfactuals, p. 23; and Pollock, "Four Kinds of Conditionals," p.55.

(2) does allow for nontrivial analysis of counterpossibles, as desired. A counterpossible is false just in case no A C world is more similar to the actual world than every A ~C world. Otherwise the counterpossible is true. So the truth value of the counterpossible varies with the similarity relation between worlds rather than being set at 'true' automatically.

The advantages of (2) are that counterpossibles may turn out to be either true or false, and that counterpossibles are assigned truth values on the same basis that all other counterfactuals are assigned truth values--no separate analysis or justification is required. So (2) is superior to (1) in at least two respects. In most other respects, (1) and (2) are the same. (2) does not make the dubious limit assumption Lewis avoids, that is, the assumption that there is always some collection of antecedent worlds closest to the actual world and never an infinite series of closer and closer antecedent worlds.[62] And (2) satisfies as well as (1) does the motivation behind Lewis's account. Lewis begins Counterfactuals saying,

> *If kangaroos had no tails, they would topple over* seems to me to mean something like this: in any possible state of affairs in which kangaroos have no tails, and which resembles our actual state of affairs as much as kangaroos having no tails permits it to, the kangaroos topple over.

This motivating intuition I judge to be right on target, save the word 'possible'. Any state of affairs which resembles the actual state of affairs as much as kangaroos having no tails permits it to is a possible state of affairs, so in this case the word 'possible' is superfluous. And in those cases in which we happen to suppose something impossible, we do not consider only the possible states of affairs that are most like actuality since in none of them does the impossible thing occur. But a one-word revision of Lewis's intuition seems just right.

However, our analysis needs a little chisholming. The formulation of (2) subtly made use of the assumption that in no world are both P and ~P true.

[62] See Counterfactuals, p. 19-21.

But of course we cannot rely on that assumption once we introduce impossible worlds. Consider this sentence:

> (F) If France were a monarchy and France were not a monarchy, then France would be a monarchy.

(F) seems to be true; France could not possibly be both a monarchy and a non-monarchy, but if it were both of those things, one of the things it would be is a monarchy. But the sentence is apparently false according to (2). The antecedent worlds most similar to the actual world are worlds in which both "France is a monarchy" and "France is not a monarchy" are true. (For the antecedent to be true without both of its conjuncts being true would be a gratuitous departure from actuality in this case.) But then no A C world is more similar to the actual world than any A ~C world is, because the nearest A C worlds *are* A ~C worlds.

Since we would like a counterfactual to be assigned 'true' whenever the closest antecedent worlds are all consequent worlds, we need to distinguish the ~C worlds from the nonC worlds, i.e., those which are not C worlds. Thus:

> (3) A counterfactual is true iff some (possible or impossible) A C world is more similar to the actual world than every A nonC world.

This amendment of (2) makes no difference in the evaluation of most counterfactuals, but it makes (F) true, and so avoids an infelicity of (2). Account (3) shares the advantages of (2) over (1) and, I think, will serve nicely as our final analysis.

This last point gives us occasion to note that there is an ambiguity in the notion of falsehood in a state of affairs. Following the wording of (1) more closely, we might have said:

> (4) A counterfactual is true iff some world in which both the antecedent and the consequent are true is more similar to the actual world than every world in which the antecedent is true and the consequent is false,

but such an account would have been ambiguous between (2) and (3). If a proposition is false in a state of affairs just when its negation is true in that state of affairs, then (4) is equivalent to (2). If a proposition is false in a state of affairs whenever the proposition itself is not true in that state of affairs, then (4) is equivalent to (3).

Perhaps the issue is mainly a definitional one, and either stipulation about our use of the word 'false' will do as well as the other. However, I am inclined to think it preferable to regard a proposition as false in a state of affairs if and only if that proposition's negation is true in the state of affairs. On this usage, a proposition P is false in a state of affairs only if the state of affairs addresses the issue of whether or not P. And in every maximal state of affairs (every world) each proposition is either true or false or both; in a non-maximal state of affairs a proposition might be neither true nor false. This seems preferable to the usage on which incomplete states of affairs make both P and ~P false by not addressing the relevant issue at all. On that view states of affairs in which both P and ~P are true make neither of the two false--and this strikes me as a much more awkward convention for the use of the word 'false'. So I am inclined to say that analysis (4) is equivalent to (2) and is in need of the above repair.

'Would' Implies 'Might'

The meanings of the words 'would' and 'might' seem to be such that 'would' implies 'might'. That is, a 'might' counterfactual of the form "If A were the case, then C might be the case" follows from the corresponding 'would' counterfactual, "If A were the case, then C would be the case." Let us call this principle WIM.

Above I claimed that Lewis's account of counterfactuals[63] does not fit the intuitive data, which tell us that some counterpossibles are false. A less obvious--though still awkward--consequence of Lewis's account is the falsehood of WIM. Unfortunately, the problem is not easily solved with impossible worlds; even an account that admits false and non-trivially true counterpossibles must choose between mutually inconsistent but plausible theses. In what follows I will present the problem and propose a solution that differs from Lewis's.

Lewis offers a strongly intuitive definition of the 'might' counterfactual in terms of the 'would' counterfactual. A 'might' counterfactual "If A were the case, then C might be the case" is true if and only if the negation of "If A were the case, ~C would be the case" is true. In terms of the 'would' connective '$\Box\!\!\rightarrow$' and the 'might' connective '$\Diamond\!\!\rightarrow$',

(DEF) $A \Diamond\!\!\rightarrow C = \sim(A \Box\!\!\rightarrow \sim C)$.

It follows immediately from (DEF) and the trivial truth of all counterpossibles that every 'might' counterfactual with an impossible antecedent is false. When A is impossible, the right hand side of the above definition is never satisfied, and so neither is the left hand side. The awkward result is the falsehood of WIM. Sometimes a 'would' counterfactual is true and the corresponding 'might' counterfactual is false.[64] One must affirm "If there were a largest prime p, pigs would fly" but deny "If there were a largest prime p, pigs might fly."

Though Lewis does not comment on this apparently unfortunate state of affairs, in a way his semantics implicitly acknowledges the intuition that 'would' counterfactuals imply their corresponding 'might' counterfactuals. It follows from the account of 'would' counterfactuals and the interdefinability of 'would' and 'might' that each 'would' counterfactual *does* imply the

[63] When the text does not specify, it should be assumed that the word 'counterfactual' refers to 'would' counterfactuals.
[64] Zagzebski takes note of this awkwardness (p. 176).

corresponding 'might' counterfactual if it is non-vacuously true. I suspect that Lewis sees the failure of vacuously true counterfactuals to conform to WIM as a somewhat counterintuitive result that is to be accepted on the strength of the principles that entail it, viz., the interdefinability of 'would' and 'might' and the argument for the trivial truth of all counterpossibles.

In any event, the awkwardness appears even if we do not assume the trivial truth of all counterpossibles. For suppose that (DEF) and WIM are true. Is it ever so that both A $\Box\!\!\rightarrow$ C and A $\Box\!\!\rightarrow$ ~C are true? If it is, then by (DEF) the 'might' counterfactuals A $\Diamond\!\!\rightarrow$ ~C and A $\Diamond\!\!\rightarrow$ C are both false. However, according to WIM, both A $\Diamond\!\!\rightarrow$ C and A $\Diamond\!\!\rightarrow$ ~C are true. So it follows from (DEF) and WIM that A $\Box\!\!\rightarrow$ C and A $\Box\!\!\rightarrow$ ~C are not both true. But there seem to be counterexamples to this conclusion, such as

> (F) If France were a monarchy and France were not a monarchy, then France would be a monarchy

and its opposite

> (F*) If France were a monarchy and France were not a monarchy, then France would not be a monarchy.

So our unpleasant options are to say that (F) and its ilk are false, to reject WIM, or to reject the plausible (DEF). Whichever we choose, we owe some explanation of the resulting oddness. The choice will be a matter of judgment; I do not have conclusive arguments for one option over the others. However, I can present my judgment and say why it is least counterintuitive to me. I reject (DEF).

Rejecting WIM is very unattractive; I find myself with rather firm intuitions which require WIM to be true. The pertinent meanings of the words 'would' and 'might' seem to me to require that propositions expressed by a sentence of the form "If A were true, C would be true" entails the proposition

expressed by the corresponding sentence "If A were true, C might be true". To substitute for the latter its negation leaves us with a contradictory pair.

Only slightly less unattractive is denying either (F) or (F*). If I try, I can generate some confusion about the antecedent. "How am I to envision such a thing?" I ask myself. "What would it be like for France to be and not to be a monarchy?" But none of this confusion helps me escape the conclusion that *if*, impossibly, France both were and were not a monarchy, France would be a monarchy (and, furthermore, it would not be a monarchy).

One might argue that the explicitly contradictory antecedent brings to mind a logically chaotic state of affairs--and who knows whether the two conjuncts will be true in it? After all, there are impossible worlds in which C&~C is true but C is not. But the strong plausibility of (F) and (F*) make it difficult to deny that in at least some (and probably most) contexts such worlds are not the ones most similar to the actual world. And if in any context some pair of counterfactuals with the forms of (F) and (F*) are true, either WIM or (DEF) is false.

(DEF), on the other hand, seems to make a doubtful presupposition. By making A $\Diamond\!\!\rightarrow$ C and A $\Box\!\!\rightarrow$ ~C exclusive of each other, it in effect presupposes that whatever the antecedent, the nearest antecedent worlds will never be ones in which both the consequent and its negation are true. Of course, for the vast majority of antecedents this is so, but we cannot rely on this assumption when antecedents may be necessary falsehoods or even formal contradictions.

With what shall we replace (DEF)? I propose:

(5) A $\Diamond\!\!\rightarrow$ C = no A ~C world is more similar to the actual world than every AC world

or equivalently

(5*) A $\Diamond\!\!\rightarrow$ C = no A non C world is more similar to the actual world than every AC world.

This differs subtly from the account derived from analysis (3) of the 'would' counterfactual and (DEF). That account can be stated as follows:

> (6) A $\Diamond\!\!\rightarrow$ C = no A ~C world is more similar to the actual world than every A world that is not a ~C world.

(5) and (6) yield different truth values in cases where all the A worlds most similar to the actual world are ~C worlds, and some are also C worlds. Here (5) returns the value true and (6) will in general return the value false. (See figure 1, where 'AC' designates AC worlds that are not ~C worlds.) It is never the case that (6) gives the value true when (5) gives the value false, so (5) is a strictly weaker account than (6). According to (5), the right-hand side of (DEF), ~(A $\Box\!\!\rightarrow$ ~C), entails the left, A $\Diamond\!\!\rightarrow$ C, but not vice versa.

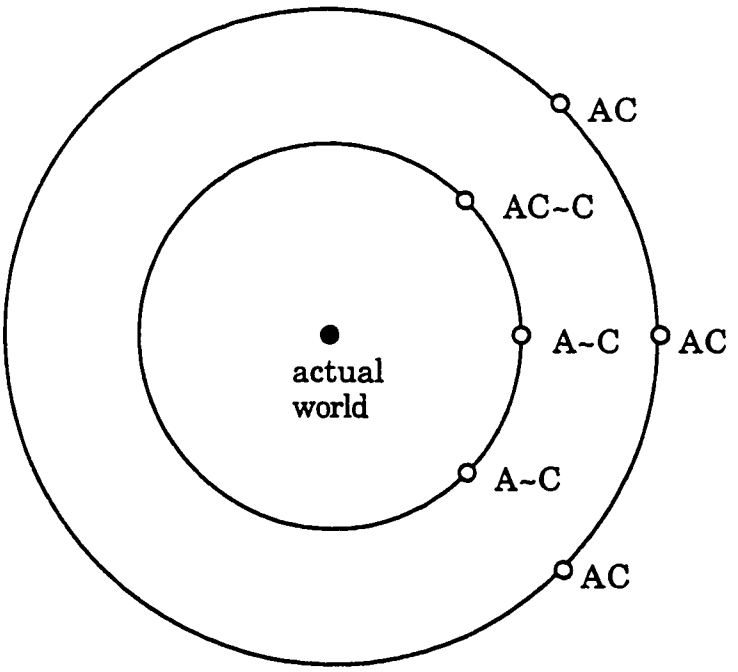

Figure 1: A case in which (5) holds and (6) fails.

If all the A worlds most similar to the actual world are C~C worlds, then (3) assigns the value 'true' to A $\Box\!\!\rightarrow$ C and (6) assigns the value 'false' to A $\Diamond\!\!\rightarrow$ C. I have argued that this is the very situation we find in the case of

counterfactuals like (F) and (F*). This illustrates the falsehood of WIM assuming the truth of (3) and (DEF).

If (5) is correct we do lose the technical convenience of defining the 'might' connective as in (DEF). The relation between 'would' and 'might' counterfactuals can then be expressed by the more complex equivalence

EQ: $(A \Diamond\!\!\rightarrow C) \,\&\, (P) = \neg(A \Box\!\!\rightarrow \neg C)$

where P rules out the scenario in which (5) and (6) yield different truth values. (P) is given by

> (P) No AC~C world is more similar to the actual world than every A non~C world, or some A nonC world is more similar to the actual world than every A non~C world.

The Unentertainable

I said earlier that there seemed to be *something* right about the idea that if the unentertainable were true, then we might just as well suppose anything you like would be true. Then I distinguished between the impossible and the unentertainable and argued that, sometimes at least, not just anything would be true if the impossible were true. But what of those propositions that really are unentertainable? Do each of them counterfactually imply each proposition?

We should pause first to clarify the notion. Lewis does not define 'entertainable', but his usage suggests it is a matter of what can be supposed or considered given a certain conversational aim. A given conversation proceeds on certain assumptions, and some suppositions violate those assumptions, whether by being inconsistent with them or by being unlikely on their basis or by some other means. Thus entertainability, like the truth values of counterfactual sentences themselves, is context-relative. What cannot even be entertained in one context may be the central focus in another. So whenever we speak of entertainability , we must remember that we are speaking, implicitly if not explicitly, of entertainability in some particular

context. With regard to suppositions lying beyond the entertainable, we are inclined to make something like Lewis's claim about the unentertainable.

> (U) For any sentence Q and any sentence P which expresses a proposition that is unentertainable in a given context, P $\Box\!\!\rightarrow$ Q is true in that context.

Lewis argues, "Confronted by an antecedent that is not really an entertainable supposition, one may react by saying, with a shrug: If that were so, anything you like would be true!"[65] As he later points out, there is a similar claim for which a similar justification can be given.

> (U*) For any sentence Q and any sentence P which expresses a proposition that is unentertainable in a given context, P $\Diamond\!\!\rightarrow$ Q is true in that context.

(U*) is a result of Lewis's alternate account of 'might' and 'would', offered to those who find it dubious that every counterpossible is true. He says, "One might perhaps motivate this weakened 'might' in much the same way as I motivated the original, weak 'would': confronted by an antecedent that is not really entertainable, one might say, with a shrug: If that were so, anything you like *might* be true!" (25). (This alternate account makes each 'might' counterpossible trivially true, and, by (DEF), each 'would' counterpossible trivially false. Though the account succeeds in assigning the value 'false' to some 'would' counterpossibles, it fails to assign 'true' to any of them, and so is no more intuitively plausible than Lewis's preferred account. However, one might well accept (U*) without accepting all of the alternate account.)

(U*) strikes me as being a bit more plausible than (U). If nothing else, we may note that if WIM is true, then (U) entails (U*), and so (U) puts us further out on a limb. But in any case both (U) and (U*), as well as the claim that neither of these two are true, are consistent with the possible/impossible world semantics outlined above.

---

[65] Counterfactuals, p. 24.

If (U) is true, then λ, the world in which every proposition is true, is the unique closest world in which an unentertainable antecedent A is true. For suppose that some other world W is as similar or more similar to the actual world given the relevant similarity relation. Then some proposition P is true in λ but not in W. But then by (3), A □→ P is false, contrary to (U).

Several types of world-orderings are compatible with (U*). One kind of world-ordering that is sufficient for the truth of (U*) is that which makes all worlds in which the unentertainable antecedent is true equally and maximally dissimilar to the actual world.

Both accounts of the unentertainable seem to take advantage of the fact that the truth values of counterfactuals depend only on the nature of the nearest antecedent worlds; the relative similarity of the more distant worlds makes no difference (until we shift contexts). Since the worlds in which unentertainable suppositions are true are always less like the actual world than the nearest worlds in which any entertainable supposition is true, the operative similarity relation may just as well be one which, say, makes λ the closest world in which a given unentertainable proposition is true. So long as the relation preserves the order of the closer worlds, those more similar to the actual world than the most similar world in which an unentertainable proposition is true, the truth values of counterfactuals with entertainable antecedents will be unaffected.

As far as I can tell, there is nothing in the notion of unentertainability itself that requires us to say that either (U) or (U*) is true. Whatever support these two have derives from the intuitive force of Lewis's exclamations. As we have mentioned, sometimes in order to interpret a counterfactual assertion charitably we assume a context which makes it true, in effect narrowing the range of similarity relations by which the counterfactual may be evaluated. Perhaps something analogous happens when someone asserts a

counterfactual with an unentertainable antecedent, so that it is evaluated as if λ were the nearest antecedent world, for example. Or perhaps this only appears to be so because we routinely exaggerate when supposing the unentertainable. "He's so stubborn, if he were to change his mind the earth would shake and the mountains would fall into the sea!" The consequent's lack of relevance suggests that any consequent would do nearly as well, maybe because the claim is not meant to be literally true. I do not wish to advance any particular of these theses about the unentertainable; I am content to note that impossible worlds accommodate a variety of stances on this topic.

We have yet to ask whether a possible antecedent can be unentertainable. The above theory does not prohibit it, nor does Lewis's, since $\bigcup\$_i$, the set of worlds in a system of spheres \$, need not include all of the possible worlds (see Counterfactuals, p. 16). And it seems to be a conceptual possibility, since entertainability is a matter of what we language-users are prepared to suppose in a given context. It seems quite likely that some logically possible proposition is or might be unentertainable in some context or another. If (U) is true, the question stands or falls with the question whether there is a context and an antecedent A such that every conditional of the form A $\Box\!\!\rightarrow$ C is true in that context; the affirmative answer has the consequence that at least one impossible world, λ, is more similar to the actual world than some possible worlds in the given context. If (U*) is true, the question stands or falls with the question whether there is a context and an antecedent A such that every conditional of the form A $\Diamond\!\!\rightarrow$ C is true in that context; this has the consequence that some impossible world is at least as similar to the actual world as some possible world (since some among the consequents counterfactually implied by A would be impossible).

Context and Truth

One concern raised by such sentences as the Caesar pair (C1) and (C2) is that such examples cast doubt on the idea that there is any objective fact to which counterfactual sentences correspond. "It may be wondered, indeed, whether any really coherent theory of the contrafactual conditional is possible at all," Quine mused in Methods of Logic. Perhaps instead counterfactual sentences serve a pragmatic function in language (eliciting certain thought-experiments in the listener, say) without expressing truths. Or perhaps counterfactual sentences supply elliptical ways of expressing truths which are not really counterfactual in nature (e.g. Caesar's military strategy often involved novel weapons and techniques).

The worry is not especially plausible in the case of certain counterfactuals. Suppose I am holding a piece of bronze (call it 'Bronze') at arm's length, and consider

(B1) If I were not holding Bronze, it would fall to the floor,

and

(B2) If I were not holding Bronze, it would remain exactly where it is.

We are inclined to say that (B1) expresses a truth, (B2) a falsehood, and that there is a fact of the matter about what would happen if I let go of a suspended piece of bronze. But this may be a bit less clear in the case of counterpossibles. Take these arithmetical hypotheticals:

(A1) If 7 were less than 5, 6 would be less than 5.

(A2) If 7 were less than 5, then the "less than" relation would not be transitive.

One might defend (A2) by pointing out that 5 is less than 6 and 6 is less than 7, and that these facts, together with the proposition that 7 is less than 5, form a counterexample to the claim that the "less than" relation is transitive. But in

this kind of case Quine's worry feels more pressing. Is there really any fact of the matter about such a question?

Lewis's remarks on the influence of context furnish a reply in the case of the more usual counterfactuals, and maybe in the case of the counterpossibles as well. The Caesar examples can be seen as counterfactuals which, whether because of context's effect in filling out partially implicit antecedents or its effect in shifting the relevant similarity relation, are not true in the same contexts. If the antecedents of (C1) and (C2) are implicit and different from each other, then the conflict between the counterfactuals is an illusion due to the similarity of the antecedents' explicit parts. If context's effect is to select a similarity relation, then even identical counterfactual sentences will express distinct propositions in different contexts. So the Caesar example does not pose any threat to the claim that (C1) and (C2) express true propositions; in either case the facts or propositions that they express are not in conflict.

I've argued that in an initially dubious case, certain sentences can express truths in the right contexts. Now let me emphasize the resemblance between the dubious case and apparently clear cases of counterfactual sentences expressing truths. Consider again (B1) and (B2). Here one is inclined to affirm (B1) and deny (B2). The two feel rather unlike the pair (C1) and (C2), between which it is difficult to choose without the prompting of context. However, I think these two pairs are in fact quite similar: in some contexts (B1) expresses a truth and (B2) expresses a falsehood, and in other contexts (B2) expresses a truth and (B1) expresses a falsehood. The difference is not that one pair depends crucially on context and the other does not, but that the contexts in which (B2) expresses a truth arise infrequently relative to those in which (B1) does; whereas the relative frequencies of contexts making (C1) and (C2) true are more or less equal.

Let me attempt to illustrate this claim with an example of a context in which (B2) seems to be true. Imagine I am standing next to a life-size statue of W. C. Fields. The statue is made of bronze, and though it has been well-rendered, that infamous nose keeps falling off. In fact it needs to be held rather firmly in place; I can only keep it in position by obscuring much of the nose with my hand. In this stance I tell you, "This would look just like W. C. Fields if my hand weren't here." And it would, assuming the statue is well-rendered, as I have said. But my claim implies that (B2) is true, since if that piece of bronze didn't stay in place--if, for example, (B1) were true--the statue would not look just like W. C. Fields.

Someone will want to say, "Maybe there is a context in which the assertion of (B2) is reasonable, but *really* it's (B1) that's true. The world itself reflects, and experiment confirms, the truth of (B1), whereas (B2) is a kind of useful fiction." To say that (B1) is true is a change of subject. (B1) really is true, given a more usual context, a normal set of concerns, e.g., one which includes the effects of gravity and the fact that Bronze is not attached to the rest of the statue. And it would be true in that sort of context even if our patterns of thinking and speaking only rarely brought us to think and speak in that context; in that case we would think less frequently about aspects of reality which are in fact important to us. But (B2) is not offered in the usual context, and lacking that context (B1) expresses a rather different proposition, one which may well be false even if the one usually expressed by (B1) is true. It is not that (B2) in the context of the W. C. Fields story fails to reflect reality. It is only that it emphasizes different aspects of reality than those we normally emphasize, viz., the current position of Bronze, and how the position of my hand blocks one's view of the statue. More often we place a relatively greater weight on the effects of gravity and on the fact that Bronze is unattached. But when I say, "This would look just like W. C. Fields if my hand weren't here," the

worlds most similar to the actual are those in which Bronze has its current position; and the proposition that the statue is fully visible is true in such worlds if the proposition that my hand is not in the way is true there.

By way of concession, we may grant that the balance of respects of comparison that makes (B2) true is narrower than the more usual balance, giving significant weight to fewer respects of comparison, and that it is somewhat ad hoc. There are very good reasons for preferring a broader balance in many or most other contexts, e.g., contexts of scientific investigations of the statue's physical properties, or of predicting the short-term future. Perhaps we may say that in this sense the more usual balances are privileged ways of weighing respects of comparison. But none of this suggests that we cannot or do not sometimes use narrow, ad hoc balances in the assertion of counterfactuals, or that such a usage is illegitimate in some way.

The upshot is that counterfactual sentences may be sensitive to context and still express objective truths, facts about the world, when asserted in a given context. In this way they resemble sentences like "I'm happy" and "We're close now"--it is not such a startling result. And this, I think, is how we should regard counterpossible pairs like (A1) and (A2). The apparent conflict between them is merely apparent because the sentences are meant to be asserted in different contexts, in this case one which emphasizes the transitivity of the "less than" relation and another which places greater emphasis on the particular mathematical facts that 5 is less than 6 and 6 is less than 7. Thus in each of the cases considered, context-sensitivity's threat against a fact of the matter is merely apparent.

Are Impossible Worlds Ever Closer Than Possible Worlds?

Above we noted some positions on unentertainable antecedents which imply that impossible worlds are sometimes as similar or more similar to the

actual world than some possible worlds are. Another position with this result is the position that some counterfactuals with possible antecedents and impossible consequents are true. Some have thought it quite clear that such a thing never occurs.

Again we have an issue which may be resolved either way given the above theory; analyses (3) and (5)--in conjunction with the view that there are impossible worlds--do not tell us very much about the nature of the similarity relations generated by the contexts our conversation can produce. However, I think certain other considerations point toward a view on which impossible worlds sometimes are more similar to the actual world than some possible worlds are.

Which similarity relation is to be operative in the evaluation of a given counterfactual depends on the relative importances of various respects of comparison. These relative importances, in turn, depend on the intentions of the assertor, the immediate conversational context of the assertion (if any), and the broader social context that surrounds all conversation. If an assertor can shape his or her intentions and the immediate context in such a way that some respect of comparison presently outweighs some necessary truth (or necessary connection between truths), it may be that the nearest world in which a given antecedent is true preserves similarity in the favored respect at the cost of dissimilarity in respect of sharing the necessary truth. And our intentions, the ways in which we might wish to emphasize one aspect of the world over another, vary very widely indeed. It should not surprise us if some assignment of weights generates a similarity relation which places some impossible world closer to the actual than some possible worlds.

Imagine a computer programmer explaining to a student that a particular machine has been programmed to carry out proofs in a formal system. Since the Gödel sentence of a system, which says essentially "This

sentence cannot be proved in this system", is of special significance, the program contains a command to print "I have proved my Gödel sentence!" if ever a proof of its Gödel sentence is produced. The student asks, "But isn't that instruction moot? Wouldn't the computer have proved something it can't if it were to print that sentence out?"

The answer is yes: it is unnecessary to include the instruction since if the computer were to print 'I have proved my Gödel sentence!' then it would have proved something that it cannot. So here we have a true counterfactual. Its consequent is impossible, so whichever A C world is nearer to actuality than all A nonC worlds is an impossible world. However the antecedent of the conditional is possible -- surely the computer could print out "I have proved my Gödel sentence!" as a result of some logically possible malfunction -- and so there are possible worlds in which the antecedent is true. Since each of these is an A ~C world, each must be less similar to the actual world than the A C world we are considering. Therefore there is an impossible world that is more similar to the actual world than some possible worlds.

Am I right to say that the counterfactual is a true one? I think so; it seems right as it occurs in the above context, and it is that context in which it is meant to be considered. The sentence does not seem to be an exaggeration, such that it is literally false despite its part in a true assertion. Some, no doubt, will beg the question, arguing that the counterfactual is false *just because* it makes an impossible world closer to the actual world than some possible worlds. It is possible, the objector grants, that the computer prints "I have proved my Gödel sentence!" This could be the result of malfunction. So if the computer were to print "I have proved my Gödel sentence!" it would have malfunctioned, since that seems the most likely way in which it might generate the printout, but it would not have done anything impossible like proving something that it could not prove.

This last counterfactual claim may be true, but as in the previous section, we have not a refutation but a change of subject. The point the student is making is that a certain command in the program is superfluous, since the conditions under which it is to be executed cannot obtain. The role of the counterfactual in this claim is to point out the programmed connection between the printout and the proof of the formal system's Gödel sentence. In this context, the worlds most similar to the actual world are those in which this same connection exists, i.e. those in which the computer prints "I have proved my Gödel sentence!" just in case it proves its Gödel sentence, functioning according to its actual design and avoiding unlikely malfunctions. A malfunction of the sort required to make the antecedent true and the consequent false is, in this context, a gratuitous difference from actuality. Hence worlds in which this sort of malfunction occur are not the ones most similar to the actual world. The similarity relation to be used in the evaluation of this counterfactual places greater weight on how the program works (or is supposed to work) than on the possibility or impossibility of the program proving its Gödel sentence. As it happens, the worlds most similar to the actual world are impossible ones.

This argument will not persuade those who believe that context influences a counterfactual only by filling in a largely implicit antecedent. If context acts as a parenthetical addendum to the counterfactual in the present example, we have something like: if the computer were to print "I have proved my Gödel sentence!" (and the computer were working exactly according to its flawless program without malfunction or outside interference), then it would have proved something that it cannot. If this is the way we think of the counterfactual, it is no longer a counterfactual with a possible antecedent, such as is needed to support my claim that impossible worlds are sometimes closer to actuality than some possible worlds.

I suspect that a large part of context's influence is of the second sort Lewis mentions, the specification of a similarity relation rather than the explication of an antecedent. I am not sure how to argue for this claim, though, nor am I certain how to distinguish the two kinds of contextual influence. Let it be enough to note that one is not compelled to think that counterfactuals of the above sort have implicitly impossible antecedents, and so it is reasonable to hold that impossible worlds are sometimes more similar to the actual world than some possible worlds are.

Note that although in the case we are pondering other considerations outweigh whether or not the computer can possibly prove its Gödel sentence, it is not as if considerations of what is possible and what is impossible have no weight at all. Consider the nearest worlds--the worlds that are nearest given the similarity relation that makes the above counterfactual come out true--in which the computer prints "I have proved my Gödel sentence!" What we imagine happening in such worlds is the computer--somehow, impossibly--proving its Gödel sentence, and then finding the instruction on what to do in this eventuality, and then printing out its celebratory exclamation. I think that we would also say that in these worlds it is true that if the computer proves its Gödel sentence then it proves something, and that it is not possible that the computer prove its Gödel sentence, and that 2 + 2 equals just what it always does. If any of these were not the case, the world would not be one of the nearest ones; any departure from these necessary truths would be gratuitous. In this way the nearest worlds, though they are not possible, are yet as possible as possible, so to speak. These worlds are as much like the actual world as they can be without extraneous departure from actuality, given the point being made by the counterfactual and the purposes for which it is asserted.

One might worry that if some counterfactuals with possible antecedents and impossible consequents were true, the reductio ad absurdum argument form would be undermined. If sometimes the possible counterfactually implies the impossible (one species of absurdity), how can we in general rely on the absurd conclusion to serve as an indicator of a faulty premise?

However, I think this worry can be quickly dispensed with. Reductio ad absurdum is really a version of modus tollens. Hence it requires only a material conditional (where the antecedent is the assumption to be disproved and the consequent is the absurdity). If the material conditional is true, which counterfactual conditionals are true is beside the point.

Why, then, are reductio arguments frequently stated with counterfactual conditionals rather than material conditionals? Because $A \mathrel{\Box\!\!\rightarrow} C$ straightforwardly implies $A \supset C$, assuming only that the actual world is maximally similar to itself. Because the implication is clear, it is rhetorically effective to assert the stronger conditional when that conditional is true (and we may assume that it frequently is). No matter if the same counterfactual sentences are false in other contexts. The counterfactuals may be true in the contexts in which they are offered, and in any case all that is needed is the material conditionals they imply.

The Similarity Objection

I close this chapter by considering an objection to a possible/impossible worlds counterfactual semantics. (It is inspired by Lewis's comments about relatively nearby "impossible limit worlds" in "Counterfactuals and Comparative Possibility", though Lewis proposes nothing like the objection to follow.) Far from claiming that impossible worlds are always more distant than the possible worlds, the objection alleges that if the antecedent of a counterfactual is false, the antecedent world most similar to the actual one will

invariably be an impossible world. Among the impossible worlds of the sort I have described, there is one whose book differs from that of the actual world only by containing one extra proposition, the antecedent. If such worlds are always the worlds most similar to the actual one, then our semantics is in serious trouble. For if both the antecedent and consequent of a counterfactual are false, and if, in the antecedent world most similar to ours, the consequent remains false, the counterfactual itself will come out false. And this is a clear indicator that the semantics has gone badly awry; many counterfactuals with false antecedents and consequents are clearly true. If I had jumped from my 13th story window this morning, I would have fallen to my death. Furthermore, since such worlds' books contain both the antecedent and its negation, the sentence "If I had jumped from my 13th story window this morning, I would not have jumped from my 13th story window this morning" will come out true, and clearly it should not.

Are we compelled to judge that "one-proposition difference" worlds are always the antecedent worlds most similar to actuality? I present two natural arguments that answer in the affirmative, both of them faulty. More particularly, each of the two methods of analyzing similarity that yield the above objection can be shown to produce similarity relations very unlike the ones normally operative in a possible worlds semantics. If the central idea of Lewis's semantics is right and there is a notion of overall similarity between possible worlds that allows the semantics to work even reasonably well, then there is no reason to suppose that the similarity notion of a possible/impossible world semantics must be like those that fall prey to the objection.

Let us call the first way of analyzing similarity the cardinality method. The idea is simply that the smaller the cardinality of the class of propositions on which the two worlds differ, the more similar they are to each other. (Two worlds *differ on* a proposition whenever the proposition belongs to the book of

one of them, but not to the other.) The maximum of similarity is the minimum of difference; i.e. two worlds that differ on no propositions are maximally similar. They are, in fact, identical. (The cardinality method thus entails Lewis's assumption that each world is more similar to itself than any other world is. See Counterfactuals, p. 14.) A world differing from the actual on four score propositions will be more similar to the actual world than one differing on a countable infinity of propositions, which in turn will be more similar than a world differing on continuum-many propositions. Worlds that differ from the actual world by the same number of propositions will be equally similar to it. Clearly, according to the cardinality method, a world differing from the actual only by the (false) antecedent of some counterfactual will be the nearest antecedent world.

But the cardinality method does not generate a notion of similarity that is at all useful in comparing the nearness of possible worlds. Two observations are in order. First, there are possible worlds that differ on as many propositions as there are propositions. Since there are at least as many propositions as there are cardinal numbers, and there is no cardinality of the collection of cardinal numbers, the differences between the books of possible worlds cannot always be measured by their cardinalities as the cardinality method seems to imply. Second, even if there were enough cardinal numbers to measure the differences between possible worlds' books, the cardinality method would not yield a usable similarity relation because the cardinality of the differences would *always* be equal to that of all propositions.

This second point may be seen as follows. Let us pretend that the class of all propositions has a cardinality, κ, and that each of its subclasses has a cardinality as well. Now consider two distinct possible worlds, W and W*, which differ at least on some proposition P. For every proposition Q which W and W* share, there is a distinct proposition on which they differ, viz., $Q \supset P$. So if κ_d is

share, there is a distinct proposition on which they differ, viz., $Q \supset P$. So if κ_d is the cardinality of the class of propositions on which W and W* differ and κ_s that of the class of propositions they share, $\kappa_s \leq \kappa_d$. Because W and W* either differ on or share each particular proposition, $\kappa_s + \kappa_d = \kappa$. Finally, since any infinite cardinal added to itself yields itself, $\kappa_d = \kappa_d + \kappa_d \geq \kappa_s + \kappa_d = \kappa$. Clearly κ_d is not greater than κ, so $\kappa_d = \kappa$. So for any two possible worlds, the cardinality of the class of propositions on which they differ is equal to the cardinality of the class of all propositions.

But then the cardinality method evidently fails to distinguish nearby possible worlds from those very unlike the actual world. Each possible world is like any other, as far as this method can tell us. It cannot be that the cardinality method is the manner of measuring similarity implicitly at work in a Lewis-style counterfactual semantics.

Will a related method do the job? The cardinality method attempted to correlate great similarity with small differences between books and to measure the differences by cardinality. But there is another way of comparing the sizes of sets and classes, the way that tells us there are more natural numbers than even numbers, despite the fact that these sets have the same cardinality. Let us call this way the subclass method. The idea is that whenever the class of propositions on which worlds W_1 and W_2 differ is a proper subclass of the class of propositions on which W_1 and W_3 differ, W_2 is more similar to W_1 than W_3 is. Of course, many pairs of worlds will be incomparable, neither class of differences being a subclass of the other, but we may take the subclass method to tell us something about comparative similarity in those cases where it does apply. If an impossible world differs from the actual world only by the antecedent of some counterfactual, that world will be the unique, nearest antecedent world on the subclass method.

The problem this time is that no two possible worlds are comparable. For suppose each of the propositions on which W_1 and W_2 differ is also one on which W_1 and W_3 differ, and that W_1 and W_3 differ on additional propositions as well. Let P be one of the propositions true in W_2 (and W_3) but not W_1, and let Q be true in W_3 but not W_1 or W_2. Then the material conditional $P \supset Q$ is true in W_1, since its antecedent is false there, and in W_3, since both antecedent and consequent are true there, but not in W_2. In W_2 it is $\sim(P \supset Q)$ which is true instead. But then $P \supset Q$ is a proposition on which W_1 and W_2 differ, but on which W_1 and W_3 agree. Contrary to supposition, the differences between W_1 and W_2 do not form a proper subclass of the differences between W_1 and W_3.

So the subclass method, like the cardinality method, fails to show that any possible world is more or less similar to the actual world than any other. Neither method will serve as an analysis of the similarity notion of Lewis's semantics. And so, assuming Lewis's account is not a failure from the beginning, there is a notion of comparative overall similarity between worlds which has nothing whatever to do with the sizes of the classes of propositions on which worlds differ, whether we judge the size of a class by its cardinality or by its sub- and superclass relations (or by both). And if such a notion is available, why shouldn't we be able to use it in a possible/impossible world semantics for counterfactuals?

Still, one might think it likely that the cardinality and subclass methods at least limit a possible/impossible world semantics. In the impossible worlds case, the cardinality and subclass methods do distinguish between worlds. Perhaps we should seek a notion of similarity which combines the cardinality method, the subclass method, and some third method that may break ties between worlds indistinguishable by the first two. Might not some such notion allow us to distinguish between near and far possible worlds, as we are

accustomed, and to maintain the intuitive claim that a world differing from the actual by one proposition is rather similar to it?

If we like, we may grant that there is a way of balancing respects of comparison that makes worlds similar when the size of the differences between their books is small, and we may also grant that this constitutes an intuitive sense in which the actual world and the world with the one-proposition difference are quite similar. But if we do, we should also admit that there are other ways of balancing respects of comparison. In particular, there are ways that place relatively strong significance on the modal status of a world, so that a world is very much unlike the actual world precisely because it is impossible. As we saw earlier, the claim that impossible worlds are very dissimilar to the actual world--perhaps even more so than any possible world--seems to have even greater intuitive support than the claim that some impossible worlds are quite similar to the actual world (in most contexts, at any rate). So even if impossible worlds which differ from the actual by one proposition are similar to the actual world in some sense, we ought not suppose that such worlds are the nearest antecedent worlds in a sense relevant to the evaluation of every counterfactual sentence.

There is another reason why the cardinality and subclass methods cannot govern a counterfactual semantics. The point belongs to Lewis, who takes pains to point out that similarity is vague. The relative importances we assign the respects of comparison between worlds vary widely and frequently. It is fitting that we analyze counterfactuals with a vague notion, says Lewis, since counterfactuals themselves are vague in just the way similarity is. However, neither cardinality nor the subclass relation is vague; neither depends in any way on context. Hence the cardinality and subclass methods are inflexible where similarity is flexible, and the former will not serve to analyze the latter.

The above objection, then, gives us no reason to think that similarity is more problematic in a possible/impossible world counterfactual semantics than it is in a possible world semantics. Unless Lewis's possible world account is much more seriously flawed than has generally been recognized, we need not suppose that the nearest non-actual worlds are always impossible. One lesson to be learned from this is that overall similarity between states of affairs is generally a matter of much more than the size of the collection of unshared propositions. Among other things, it is a matter of the modal properties of the states of affairs involved--something which may have nothing to do with how many or few propositions they differ on.[66]

[66] I owe Patricia Blanchette thanks for some very helpful comments on this section.

CHAPTER 5: IMPOSSIBLE WORLDS IN EPISTEMIC AND RELEVANT LOGIC

When I described the idea of impossible worlds to one philosopher, he worried that there would be no logic which held for all worlds if impossible worlds were as I described them. I replied that indeed no logic would hold for all worlds, since for any two propositions P and Q, there would be some world in which P was true but Q was not. Some caution is required here, since there are rules of inference (of a sort) which hold for all worlds. For example, the trivial inference

$$\frac{P}{\therefore P}$$

holds; in every world in which P is true, P is true, even though ~P might be true there also. Similarly,

$$\frac{\begin{array}{l}P\\Q\end{array}}{\therefore P}$$

and its generalizations hold at every world, so it would not be quite right to say that there are no rules of inference that hold for all worlds. But of course the law of noncontradiction fails, as does modus ponens and our other favorites. So there is a significant sense in which no logic holds at all worlds.

To make this a bit more precise, let us provisionally make a terminological stipulation: a rule of inference *holds for* (or *holds with respect to*) a world W just in case whenever propositions having the forms given by the rule's premises are true in W, the proposition having the form of the conclusion is also true in W. In contrast, a rule of inference is *true in* W just in case the

proposition expressed by the rule (e.g. *propositions of the form Q follow from propositions of the forms P and P⊃Q*) is true in W. Clearly whether a rule is true in a world is independent of whether the rule holds for that world. So given the account of the preceding chapters, it is in these senses that logic does not bind the impossible worlds: no rule of inference is true in every world, and only the trivial inference and its generalizations hold for all worlds.[67]

(There are some "rules" which require a different notion of "holding for" a world. For example, we might want to say that the law of the excluded middle holds for all worlds since by definition each world is maximal, but it has neither premises nor conclusion. For another example, the rule of necessitation one finds in S5 applies only to the system's theorems. That is, either the system is an axiomatic system which takes only theorems as premises, or else the rule is restricted to cases in which the premises are theorems. The same effect can be achieved by building the rule into the axioms, but if it is not, the above notion of "holds for" will fit poorly. We will not take up this complication in what follows; instead we will assume the logics under consideration have rules only of the non-problematic sort.)

But we must not conclude from this that impossible worlds have no interesting uses in logic. The previous chapter provides one important counterexample. In this chapter, I want to provide some evidence of impossible worlds' utility in some other areas by sketching two of these applications.

A number of authors have judged that impossible worlds are particularly well-suited for arenas in which closure under entailment (strict implication) fails. What a given person knows, for example, is not closed under entailment; one may know the axioms and definitions of number theory without knowing that there are arbitrarily large gaps in the sequence of primes. Since

[67] Contrast the theories of Ed Mares, "Who's Afraid of Impossible Worlds?", Chris Mortensen, "Peeking at the Impossible," and Greg Restall, "Ways Things Can't Be." In each of these theories, each impossible world is closed under some paraconsistent consequence relation.

impossible worlds' books likewise are not closed under entailment, we might regard certain worlds with books containing each proposition a person knows as "epistemic alternatives" for that person. We could then say a person S knows P just in case P is true in each of his or her epistemic alternatives, and that P is true as far as S knows just in case P is true in some of his or her epistemic alternatives. That is to say, we could treat "S knows that" and "As far as S knows" as modal operators analogous to necessity and possibility operators, which quantify universally and existentially (respectively) over certain worlds.

Impossible worlds would also seem to have a natural application in relevant logic. The relevantist aims to specify a notion of implication according to which one proposition implies another only if the propositions are relevant to each other in the appropriate way, and so to avoid the so-called paradoxes of implication: that a necessary falsehood implies every proposition, and that a necessary truth is implied by every proposition. The paradoxes are a result of viewing implication as strict implication, that is, as a modalized version of material implication: $\Box(A \supset B)$, where the necessity operator ranges over all possible worlds. One way in which a relevantist might hope to avoid the paradoxes is to replace the usual necessity operator with another necessity operator ■ which ranges over some impossible worlds as well as all possible worlds. If A is said to entail B only when $\blacksquare(A \supset B)$, then a necessary falsehood will not entail everything, since in some impossible worlds in the range of ■ the falsehood is true, but not every other proposition is. Likewise necessary propositions will not always be true in the worlds in question, and so not every proposition will entail them.

We will return to each of these projects in the pages that follow. My aim is not to solve every challenge facing these applications, but only to show that impossible worlds would be well-suited to serve as central features of such

theories, and that we have here fruitful avenues for further research. Before I make that case, though, we need to address a philosophical prerequisite. What is this notion, implicit in our provisional definition of 'holds for', of a proposition's *form*?

Propositional Forms

Though the above definitions make the distinction between being true in a world and holding for a world reasonably clear, there is still something dubious about the definition of 'holds for'. As it stands, that definition presupposes that each proposition has a certain form. Though it may be tolerably clear what it is for *sentences* of a given formal or natural language to have a certain form, it is not entirely evident what it is for a proposition to have a certain form. Indeed, philosophers disagree whether propositions even have different forms--perhaps they are unstructured entities.[68]

It will be useful to have before us some versions of the view that propositions have form. But before we sketch these, we should note that this issue is importantly related to what one believes about the applied semantics of propositional (or "sentential") logic (and other logics).

A pure semantics is a set-theoretic construction of some kind--in the case of propositional logic, typically a set of sentence letters, truth-functional connectives (standardly, ~, &, ∨, ⊃, and ≡), and punctuation marks (such as parentheses); rules for arranging these as well-formed formulae ("wffs" or simply "formulae"); truth values T and F, and rules for determining the truth values of non-atomic formulae given the truth values of their constituent

[68] It is a different question whether propositions have *parts*. Some (Russell of the *Principles*, e.g.) have thought that *Socrates is wise* has Socrates as a part, though it would seem that this makes the proposition as short-lived as Socrates himself. Others have thought *Socrates is wise* has the property *wisdom* and the individual essence *Socrateity* as parts. The proposition does have a close, logical relationship to *wisdom* and *Socrateity*, but I am not sure why this relationship should tell us anything about mereology.

atomic formulae; a definition of 'valid formula' as a formula which has the truth-value T on every assignment of truth values to atomic formulae; and rules for deriving one formula from a set of formulae.

The pure semantics as such is merely of mathematical interest; it tells us only that certain formulae are valid in the given sense, and that certain formulae are derivable from others according to the rules of the system. If we really want a *propositional* logic, i.e., a logic which tells us something about the relations between propositions, we will need a logic with an interpretation of its various formulae, an applied semantics. In particular, we will need an interpretation on which the wffs of the system express or represent propositions.[69]

Let us consider four increasingly strong ways of understanding propositional logic:

(1) Each wff expresses a proposition when the logical connectives are interpreted in the usual ways and each sentence letter is assigned a specific proposition. Since wffs can be combined according to certain rules to form other wffs, for any two propositions there is (at least) one proposition that is the conjunction of the two, and a proposition which is the disjunction of the two, and a conditional proposition of which one of the two is the antecedent and the other is the consequent, and so on. So there is a relation *being a conjunct of* which holds between certain pairs of propositions, and there are analogous relations in the disjunction and conditional cases. Also, each proposition has a negation, which is itself a proposition.

[69] Of course not everyone agrees about this. Some prefer to interpret formulae as sentences (or sentence-types) of some natural language. Though I do not intend to argue this point at any length, I will say that this approach seems to stop short of the mark. Among the things we want to illumine with a logic are implication relations, and these relations hold between propositions, not sentences (unless perhaps in a subordinate sense). Those who quine propositions will object, naturally, but this is not the place to carry out that philosophical debate.

The truth values T and F are, of course, interpreted as truth and falsehood; the propositions represented by formulae assigned T are true, and those represented by formulae assigned F are false. Finally, derivability models deducibility: a formula ϕ is derivable from the formulae in a set Γ just in case the proposition which ϕ expresses is the conclusion of a deductively valid argument whose only premises are the propositions represented by the formulae in Γ.

(2) The claims of the preceding view are true. Furthermore, for any propositions P and Q, if P is a conjunct of Q (i.e. P stands in the relation *being a conjunct of* to Q), P is necessarily a conjunct of Q. And the same is true, *mutatis mutandis*, of the relations *being a disjunct of*, *being an antecedent of*, *being a consequent of*, *being a negation of*, such three-place relations as *x is a conjunction of y and z*, and others. In particular, whether one proposition is a conjunction of another does not depend on our linguistic conventions.

(3) The claims of the preceding views are true. In addition, each proposition has a single logical form, and only formulae which share this form express it. Conjunctive wffs (formulae whose main connective is '&') express conjunctive propositions, negations express negative propositions, and so on. Sentence letters express atomic propositions, i.e., propositions that are neither conjunctive, disjunctive, negative, conditional, or biconditional. The propositions expressed by valid formulae are logical truths.

(4) The claims of the preceding views are true. And not only does each wff express a proposition, each wff expresses a *unique* proposition. Hence distinct wffs always express distinct propositions.

Which understanding of propositional logic is correct, whether one of the above or some other, depends on the answers to substantive philosophical issues, issues I can only treat briefly and incompletely here. Nevertheless, I will say a word or two about each interpretation. I'll go in reverse order.

(4) has some rather implausible consequences: (a) Every pair of propositions has not one but two conjunctions (each having the same conjuncts in a different order). (b) Either the relation *being a negation of* is not symmetric (so that a proposition ~P is the negation of P but not vice versa) or else some propositions have more than one negation (e.g. ~P has both P and ~~P as negations). In each case (and in each of the other, similar cases that could be offered) we seem to have a multiplicity of ways of representing one thing, not a multiplicity of things represented. Clearly it is possible for a language (whether formal or natural) to represent a proposition in more than one way, and formulae like 'p∨q' and 'q∨p' seem as much like duplicators as any pair does. It is difficult to *prove* that the two represent the same proposition. But it is equally difficult to provide an argument to the contrary which does not appeal to the differences in the way these formulae are written or uttered or thought--and these differences are not reliable indicators of a difference between propositions. Why think that the strings of symbols 'P∨Q' and 'Q∨P' correspond to two ways of combining propositions? If we added the notation ∨(P,Q) and ∨(Q,P), would we think that P and Q had four different conjunctions?

(3) suffers from related problems, despite its appealing straightforwardness. One of the consequences of (3) is that each proposition belongs to exactly one of six categories: atomic proposition, conjunction, disjunction, negation, conditional, or biconditional. A proposition is a disjunction, say, because it has a disjunctive nature, to the exclusion of a conditional or negative nature.

Set aside worries about propositional logic's failure to capture the logical form of quantified or modal statements for the present. The concern I want to mention now is another apparent case of a multiplicity of representations of a single proposition. Consider the propositions expressed, in a suitable context,

by "Either this ring is pure gold or not" and "Either this ring is pure gold or we were cheated." If (1) is correct, the propositions so expressed have a conjunction. Call it C. There are also the propositions expressed by "This ring is pure gold" and "This ring is not pure gold and we were cheated." If (1) is correct, the propositions so expressed have a disjunction D. According to (3), C and D are distinct, since C is conjunctive and D is disjunctive.

But are they in fact distinct? Certainly C is logically equivalent to D, and the subject matter is just the same in each. I suppose if someone wished to stick to his guns and insist that C are D distinct, logically equivalent propositions, I would not be able to offer a proof that this is not so. However it becomes very difficult to see what non-question-begging reason could be given to aid the defender. We *think* C and D differently, it might be observed, but is the difference really due to distinct propositions or is it due to distinct ways of internally verbalizing the same proposition? Is it, perhaps, like the difference between how one "thinks a proposition" in English and how one "thinks" the same proposition in German? Non-question-begging reasons can be given for distinguishing between *some* logically equivalent propositions. There is an evident intentional difference between *this ring is pure gold* and *this ring is pure gold and either we were cheated or not*. But no such difference between C and D is apparent, and so (3) seems somewhat dubious.

A related problem for (3) is brought to light by the fact that there are a variety of systems of propositional logic, not all of which share the same connectives. It is well-known, for example, that each formula of standard propositional logic is equivalent to a formula which contains no sentential connectives aside from the tilde and the ampersand. And just as the formulae of a system with only these two connectives also belong to the larger and more common systems, formulae of the common systems themselves belong to

systems with connectives besides the familiar ones. We could introduce '*' as an exclusive disjunction connective, with the introduction rule

From p, ~q derive p*q (or derive q*p)

and the elimination rule

From p*q, ~q derive p (or from p*q, ~p derive q).

The formula p*q would then be provably equivalent to

$p \vee q \ \& \sim(p \& q)$

and to

$(p \& \sim q) \vee (\sim p \& q)$.

Given a suitable interpretation of the atomic formulae of such a system, the formula p*q would express a proposition. From the perspective of (3), p*q would represent a proposition that fell into one of the six categories--though whether a conjunction or a disjunction or something else is unclear. So again (3) is dubious.

The point is that when standard propositional logic is set in a context of systems with different connectives, it appears rather arbitrary to base an ontology of propositions on the standard system in this way. To think that propositions have just the forms that mirror a particular system is to borrow too much from the formal language used to represent them. The ontology of (3) appears no more likely to be true than an ontology based on a system which includes '*' (that is, an ontology which says that some propositions are exclusive disjunctions, and that these are neither inclusive disjunctions nor conjunctions nor negations, etc., though the former may be equivalent to the latter) or an ontology based on a system which lacks '≡'.

If we refuse to adopt an ontology of propositions arbitrarily, then, we must either remain agnostic about which sort of propositional logic cuts the world of propositions at the joints (again, ignoring whatever complications may

be required by quantified, modal, or other propositions whose complexity is not reflected by propositional logic) or else reject the idea that *any* propositional logic cuts the world of propositions at the joints. To choose the latter path is to refuse to go any further than (2) in an ontology of propositions, affirming at most that propositions may be conjunctions or disjunctions or conditionals etc., without claiming that these categories are mutually exclusive. In either event, it seems best not to adopt (3).

Are there reasons for rejecting (2)? One might hold that for a proposition to be a conjunction is merely to be expressed by a sentence suitably constructed from proposition-expressing sentences and the word 'and' (or whatever word plays its logical role). If this is right, *conjunctivity* is a linguistic property; it has to do with what our sentences express, rather than with the nature of propositions. "John is a man" and "John is unmarried" express propositions, so "John is a man and John is unmarried" expresses a conjunctive proposition. But if by some linguistic accident that proposition had not been expressed by any such sentence--if it were expressed only by "John is a bachelor", e.g.--then it would not have been a conjunctive proposition. Whatever proposition was expressed by "John is a man and John is unmarried" in that case would be a conjunctive proposition (assuming 'and' would still play the logical role that it does and that the words "John is a bachelor" did not somehow conspire to accomplish what 'and' does).

What does this understanding of propositions have to recommend it? For one thing, it avoids the pitfalls of (3) and (4). In fact, one might see those very difficulties as casting doubt upon the idea that conjunctivity is a metaphysical rather than linguistic property. A lesson we learn from the failure of (3) and (4) is that we ought not imagine propositional logic mimics the Platonic realm of propositions perfectly. The fact that any two formulae can be conjoined in two ways does not tell us that every pair of propositions has two conjunctions. The

fact that two formulae are distinct does not tell us that the proposition(s) they express are distinct. And the fact that formulae may be conjunctive by nature does not tell us that propositions may be so. Our ways of representing propositions in a formal system sometimes mislead us about the nature of propositions, so we have reason to be suspicious of the idea that properties like conjunctivity and disjunctivity would have been instantiated just as they are no matter what our linguistic conventions had been.

Though the general suspicion that propositional logic occasionally misleads us about propositional ontology seems justified to me, I am not inclined to reject (2) on those grounds. (2) is false if the following sort of instance is possible. Under one convention, Convention A, the sentence "John is a man" expresses proposition M, "John is unmarried" expresses proposition U, and "John is a man and John is unmarried" expresses proposition C. Because it is expressed by a conjunctive sentence, C is a conjunction. According to Convention A, a sentence is conjunctive if it is formed by placing the word 'and' between two other sentences (and making a few adjustments in punctuation). Under Convention B, "John is a man" expresses M, and "John is unmarried" expresses U, as under Convention A. However C is not expressed by any conjunctive sentence but by some other sentence, say, "John is a bachelor", and thus C is not a conjunction. We may further suppose that the conjunctive sentence formed from "John is a man" and "John is unmarried" expresses some proposition other than C. (It makes little difference to the example whether according to Convention B conjunctive sentences are formed as they are according to Convention A or via another connective such as '+' or by concatenation or otherwise.)

I think we can see that such a case is dubious if we consider what it is to be a conjunctive sentence. We've noted that to be a conjunctive sentence under Convention A is to be a sentence suitably constructed from two other

sentences and the word 'and', but that under other conventions conjunctive sentences may be constructed somewhat differently. What is it about a certain way of constructing sentences that makes it a way of constructing conjunctive sentences?

The answer is that the constructed sentence is always stipulated to be true if both of the two component sentences are true, and false otherwise. Indeed, this construction (given the meanings of the component sentences) provides the entire meaning of the sentence. Regardless of whether the convention in question says such sentences are to be formed with an operator like 'and' or '&' or by an organization like concatenation, regardless of whether the conjunct sentences themselves appear in the conjunctive sentence or are referred to by name or description, the conjunctive sentence is designed to be related to the conjunct sentences in the way given by the standard truth tables for '&'. Any sentence whose meaning is not given specifically in this way does not deserve the title of conjunctive sentence.

But if this is so, any conjunctive sentence formed from sentences expressing M and U will express proposition C. The meaning of the conjunctive sentence is entirely a function of the meanings of the conjunct sentences and the nature of conjunctive sentences. So although it may be that, under some linguistic convention, C is expressed only by non-conjunctive sentences, under any convention which does allow conjunctive sentences to be formed the conjunction of the sentences which express M and U will express C. It thus seems rather pointless to regard conjunctivity of propositions as dependent on linguistic convention. We should affirm (2).

Finally we come to (1). As noted earlier, there are serious questions about whether such a view of propositional logic, conservative as it is, is viable. In particular the nominalist will reject the ontology of (1). But to address nominalism in a satisfactory manner would take us far afield from our topic,

and I do not wish to address it in an unsatisfactory manner. I will therefore proceed on the assumption that (1) is correct. (It is unlikely in any event that anyone with strong nominalist sympathies has read this far into the dissertation.)

To recap: we have been looking at a number of different ways of understanding the formulae of propositional logic and their relation to propositions as a means to clarifying what it is for a proposition to have a certain form. We have concluded that positions (3) and (4) are problematic, but that (1) and (2) are relatively safe. Below we will not define 'form' per se. Instead we will define a relation "has the same form as" or "has the form of" which holds between propositions and formulae after we say what it is for one formula to have the same form as another. The definition will say, in effect, that f and g have the same form iff each is a substitution instance of the other.

Let's say that the immediate subcomponents (sub-formulae) f_1 and g_1 (of f and g, respectively) *correspond to each other* just in case f and g have the same main connective, and both f_1 and g_1 occur before the connective or both occur after it. We'll say in general that component f_2 of f and component g_2 of g correspond to each other just in case f_2 and g_2 are f and g, or corresponding immediate subcomponents of f and g, or corresponding components of corresponding components of f and g. Then:

> F-FORM: Formulae f and g *have the same form* iff (i) f and g have the same main connective (if any), and (ii) each component of f has the same form as the corresponding component of g, and (iii) if any two components of f have a component in common, the corresponding components of g share the corresponding components.
>
> The formulae $f_1, f_2, ..., f_n$ (in that order) *collectively have the same forms as* the formulae $g_1, g_2, ..., g_n$ (in that order) iff (iv) each f_i has the same form as g_i, and (v) if any formula is a component of both f_i and f_j, then the corresponding components of g_i and g_j are identical.

This definition is designed to ensure, for example, that (p⊃q)&p has the same form as (r⊃s)&r but not as (l⊃m)&n (as it would if we omitted clause (iii)). Order is important at this stage; (p⊃q)&p does not have the same form as p&(p⊃q). The definition of 'collectively have the same forms as' allows us to say that the formulae p and p&q do not collectively have the same forms as r and s&t, even though p and r share the same form, as do p&q and s&t.

> P-FORM: Supposing that the language in question has been interpreted so that each of its atomic formulae expresses a proposition, a proposition P *has the form of* a formula f iff P is expressed by f when the logical connectives are interpreted in the usual ways, or P is expressed by a formula g which has the same form as f.
>
> The propositions P_1, P_2, ..., P_n (in that order) *collectively have the forms of* the formulae f_1, f_2, ..., f_n (in that order) iff, given some one interpretation of the atomic formulae, the usual interpretation of the logical connectives, and some formulae g_1, g_2, ..., g_n, each P_i is expressed by g_i, and g_1, g_2, ..., g_n collectively have the forms of f_1, f_2, ..., f_n.

In offering this definition we do not assume that a proposition has a unique form or that a proposition is expressed by at most one formula. To take our earlier example, suppose that on a certain interpretation g expresses the proposition *this ring is pure gold* and h expresses *we were cheated*. It is consistent with our definitions that both (g∨~g)&(g∨h) and g∨(~g&h) express the same proposition, even though these formulae do not have the same form. In this case we may say that the proposition so expressed has more than one logical form. (And it is also consistent with our definition that these formulae do not express the same proposition, and with the claim that no proposition has more than one logical form.)

For the same reason, order is not necessarily important at this stage. If we are inclined to think that g∨h and h∨g express the same proposition, then that proposition will have the form of g∨h and the form of h∨g. If, on the other hand, we are inclined to think the two formulae express different propositions,

we may suppose that one of the propositions has the form of g∨h and the other has the form of h∨g.

I have tried to clarify the notion of propositional form in terms of sentential form, but I do not mean to suggest that propositions somehow depend on sentences for their form. If propositions truly can be said to have forms, they have their forms essentially. If there were no languages, propositions would have whatever intrinsic characteristics they have now. However, depending on our ontological assumptions, the above relational notions may mislead us about what kind of forms propositions would (and do) have. For example, if we have the first of the inclinations mentioned in the previous paragraph, then we will think disjunctive propositions have (at least) two different forms--one corresponding to g∨h and another corresponding to h∨g. But perhaps disjunctive propositions have only one disjunctive form, in virtue of which they are usually expressed by (at least) two different disjunctive sentences of a given language. In such a case we need to remember that P-FORM tells us about a relation between propositions and contingently extant sentences, and does not actually define the notion of a proposition's "form." It is designed to help us give some sense to the idea that a certain logic characterizes the relations between certain propositions. We should not regard it as a precise account of whatever forms propositions may have in themselves.

Digression on Logical and Absolute Possibility

The issues discussed above are importantly related to the question whether there is a genuine difference between logical and "absolute" or "metaphysical" possibility. The answer 'yes' seems to assume that propositions have by nature the kind of structure our formal sentences do, as

(2) entails. If this is right, then there are logical truths in the sense that there are disjunctions of a proposition and its negation, and there are conditional propositions whose antecedents are also their consequents. And aside from these propositions there are "non-logical" necessary truths. In terms of the preceding definitions, we could express this view as follows: some necessary truths do not have the form of any tautologous formula (or of any theorem of our preferred logic). Perhaps the likes of *God exists* and *there are numbers* will be offered as necessary truths without the requisite logical structure.

But to deny (3) is to leave open the possibility that a necessary truth have more than one logical form, some of which may be tautologous. In this case it will be misleading to speak of "the" logical form of a proposition. And if it is admitted that a proposition may have the same form as several formulae which do not share their forms, it may be far from clear which propositions will count as logical truths, or whether any necessary truths will fail to count as logical truths. Suppose the proposition *there are numbers* is expressed by an atomic formula n. Might this proposition also be expressed by some formula $p \vee \sim p$, or by $(r \& \sim r) \supset q$? How can we tell? If *there are numbers* is expressed by some tautologous formula, then it is a logical truth, despite the fact that "there are numbers" does not have a clearly tautologous structure. To take a particular case in which (1) and (2) are true but (3) is not, the theory that necessarily equivalent propositions should be identified says there is only one necessary truth, so there can be no distinction between logical truths and necessary truths. Other theories will provide more necessary truths, but it may remain unclear whether each necessary truth covertly takes the form of some tautology or another.

I don't really want to answer the question of whether we can distinguish between necessary truths which are logical truths by nature and those which

are not. I raise the issue here only to point out how it is related to the questions we have been considering. One who is willing to adopt (3) will say each necessary truth has a single logical form; if that form is tautologous, the proposition is a logical truth, and otherwise it is not. One who believes that propositional forms mirror the forms of some logic or another but is agnostic about which one will agree that each proposition has one logical form; so if we know there are numbers to have an atomic form, we know it is not a logical truth. But one who believes that necessary truths may have a number of different forms will find it more difficult to say whether there are necessary truths without the form of some tautology or another, and so will find it difficult to say whether there are necessary truths which are not logical truths.

The above definitions are not meant to imply the last of these views. They are meant to be neutral with respect to them. They allow a proposition to "have the form of" a variety of formulae without implying that propositions have more than one "form."

When a Logic Holds For a World

We may now say more clearly when a rule of inference holds for a world, and we may expand our definition so that it also says when axioms and logics hold for a world (keeping in mind that what we say here applies only to a certain class of rules and logics).

Assume that a rule of inference of a given logic consists of a set of formulae of the logic, all but one of which are *premises*. The remaining formula is the *conclusion*. A rule of inference holds for a world W iff, for any propositions P_1, P_2,..., P_n which collectively have the forms of the premises and the conclusion, if the propositions expressed by the premises are true in W, then the proposition expressed by the conclusion is true in W.

Assume that an axiom is a formula. An axiom holds for a world W iff every proposition which has the form of the axiom is true in W.

Finally, we may say that a logic holds for a world W iff each of its rules of inference and axioms holds for W.

(In the same vein we might also add that the law of noncontradiction (LNC) holds for a world W iff for no proposition P is it the case that P and ~P are both true in W. Note that we can't capture this notion by saying that the axiom ~(f&~f) holds for W, since in some impossible worlds each of P and ~P and ~(P&~P) are true. Nor can we capture the notion by saying that the rule f / ~~f or its converse holds for W, since in some impossible worlds each of P and ~P and ~~P are true, whether or not P and ~~P are distinct. Given this definition, LNC holds for each world which lacks an inconsistent pair, in the terminology of Chapter 3, as well as for many worlds which have inconsistent pairs.)

Again, these definitions do not presuppose either of the ontological views (3) and (4). The issue whether the formulae p&q and q&p express the same proposition, for example, is not decided.

Some Instructive and Useful Results

Result 1: If a logic holds for a world, the propositions expressed by its theorems are true in that world.

Proof: Suppose logic L holds for world w (under some interpretation of its atomic formulae and the usual interpretation of its logical connectives) and that T is a theorem of L. T is deducible from the axioms of L via the rules. Now we show inductively that T expresses a proposition true in w.

Base case: T expresses a proposition with the form of an axiom A. A holds for w, so each proposition with the form of A is true in w, by definition.

Inductive step: $T_1 \ldots T_n$ are theorems expressing propositions true in w, and T is deducible from $T_1 \ldots T_n$ via rule R. Since R holds for w, $T_1 \ldots T_n$ and T collectively have the forms of R's premises and conclusion, and the propositions expressed by $T_1 \ldots T_n$ are true in w, the proposition expressed by T is also true in w.

Result 2: The converse of Result 1 is false.

Proof: Let L be a logic in the standard language with no axioms (and hence no theorems) and modus ponens as its only rule of inference. Consider the world w whose book lacks only some proposition P, so that $B_w = B_\lambda - \{P\}$. Then the propositions expressed by L's theorems are true in w, trivially, but L does not hold for w, since modus ponens does not hold for w. If the formula p expresses P, there is a formula q such that the propositions expressed by q and $q \supset p$ are true in w, but the proposition expressed by p is not.

Result 3: There are impossible worlds for which modus ponens holds.

Proof: The world λ, in which every proposition is true, is such a world.

Result 4: Every logic holds for some impossible world.

Proof: We need only assume that each formula of each logic, when interpreted, expresses a proposition. Then it follows immediately that since every proposition is true in λ every rule and every axiom hold for λ.

Result 5: There are non-trivial impossible worlds (i.e., impossible worlds other than λ) for which modus ponens holds.[70]

[70] Thanks to Michael Thrush for helping me see this result could be proved. Thanks also to Michael Kremer, to whom the given proof is due.

Proof: Let P be a contingent proposition consistent with both propositions N and ~N. Let B_μ be the class of propositions not entailed by P. We will show that μ is non-trivial, impossible, and maximal, and that modus ponens holds for μ.

Because P entails P, P does not belong to B_μ, so μ is non-trivial.

N belongs to B_μ, and so does ~N, so μ is impossible.

Consider any proposition Q. As P is contingent, it is true in some possible world w. In w, either Q is true or ~Q is true. Hence either P does not entail Q or P does not entail ~Q. Hence either Q or ~Q belongs to B_μ, and μ is maximal.

Finally, suppose that propositions Q and $Q \supset R$ belong to B_μ. Then P does not entail $Q \supset R$. It follows that P does not entail R, since if it did, it would entail $Q \supset R$. Hence R is in B_μ. So modus ponens holds for μ.

It would also be helpful to prove that there are non-trivial impossible worlds for which modus ponens holds and whose books contain all necessary truths. (The semantics to be considered shortly uses such worlds.) Unfortunately, it is not as clear as we might like that there are worlds of this sort. Two results are relevant.

Result 6: There is no non-trivial impossible world w such that (i) modus ponens holds for w, (ii) every necessary truth is true in w, and (iii) w has an inconsistent triple.

Proof: We will show that any world satisfying (i), (ii), and (iii) is trivial. Let P, Q, R be an inconsistent triple of world w. Since every necessary truth is true in w,

$P \supset (Q \supset (R \supset (P\&Q\&R)))$

is true in w. Since modus ponens holds for w, (P&Q&R) is also true in w. Further, for any proposition S, the proposition

$(P\&Q\&R) \supset S$

is a necessary truth. Thus S must be true in w. Since this is so for any S, w is trivial.[71]

Clearly, similar arguments allow us to rule out non-trivial impossible worlds satisfying (i) and (ii) and having inconsistent pairs or necessary falsehoods. Inconsistent n-tuples (n a natural number greater than 3) can also be ruled out by similar arguments, or we may recall from chapter 3 that every impossible world with an inconsistent n-tuple (n>3) has an inconsistent triple. Thus we may strengthen Result 6, replacing (iii) with

(iii*): w has an inconsistent n-tuple (n>0).

Note that the question whether any non-trivial impossible worlds at all satisfy (i) and (ii) is related to a question I posed in chapter 3: are there impossible worlds that lack inconsistent triples? If there are, there are also non-trivial impossible worlds for which modus ponens holds and whose books contain all necessary truths. Every necessary falsehood is inconsistent with every other proposition, so a world which lacked inconsistent triples would also lack necessary falsehoods. Such a world would thus be non-trivial and every necessary truth would be true in it. Any world for which modus ponens does not hold has an inconsistent triple Q, $Q \supset P$, and ~P, so modus ponens holds for worlds which lack inconsistent triples. Given Result 6, any non-trivial impossible worlds satisfying (i) and (ii) lacks inconsistent triples. So:

Corollary: There are impossible worlds without inconsistent triples iff there are non-trivial impossible worlds for which modus ponens holds and such that every necessary truth is true in them.

[71] Thanks to Tom Crisp for the insight behind this proof.

As I indicated in chapter 3, I do not know of a proof that there are impossible worlds without inconsistent triples. However, some progress in this direction has been made. It is plausible that there is an inconsistent set of propositions with no finite, inconsistent subset. We might take { *Lewis is over 7 feet tall*, *Lewis is less than 7.1 feet tall*, *Lewis is less than 7.01 feet tall*, *Lewis is less than 7.001 feet tall*, ...} to be such a set. It is also plausible that, for at least some sets of this kind, there is a proposition not entailed by any of its finite subsets. In this case, *Socrates is wise* seems to do nicely. If we also assume that the propositions can be well-ordered, the right side of the corollary follows. The result is due to Michael Kremer.

Result 7: If there is an inconsistent set of propositions $P=\{P_0, P_1, P_2, \ldots, P_n, \ldots\}$ and none of its finite subsets entails Q, and if the class of all propositions can be well-ordered, then there is a non-trivial impossible world in which every necessary truth is true and for which modus ponens holds.

Proof: We will construct a world in which each P_i is true and Q is not true. Since the propositions can be well-ordered, they can be indexed by the transfinite ordinals. We need not assume that the propositions form a set when we say they can be well-ordered. For simplicity, though, let's assume they do form a set of cardinality k and that they can be indexed by the ordinals a<k. So let

$$A_0, A_1, A_2, A_3, \ldots, A_\omega, A_{\omega+1}, A_{\omega+2}, \ldots, A_{\omega\cdot\omega}, \ldots$$

be a transfinite sequence of all propositions.

Now we define a sequence of sub-books by recursion on the ordinals:

$B_0 = P \cup \{A : A \text{ is a necessary truth}\}$

$B_{a+1} = B_a \cup \{A_a\}$ if no finite subset of $B_a \cup \{A_a\}$ entails Q

(a is an ordinal <k)

$B_{a+1} = B_a$ if some finite subset of $B_a \cup \{A_a\}$ entails Q

B_l = the union of the B_a's for a<l, when l is a limit ordinal (<k).

We let B_k be the union of all the B_a's, a<k. That is, A belongs to B_k iff A is in B_a for some a<k.

Lemma: For all a<k, no finite subset of B_a entails Q. Further, no finite subset of B_k entails Q.

Proof: By induction on a. For a=0 this is obvious by our choice of Q. Next, assume that for some a, no finite subset of B_a entails Q. Then by the definition of B_{a+1}, the result follows: either $B_{a+1} = B_a$ or $B_{a+1} = B_a \cup \{A_a\}$ and no finite subset of $B_a \cup \{A_a\}$ entails Q. For the limit case, assume no finite subset of B_a entails Q for all a<l (where l is a limit ordinal). Then every finite subset of B_l is a finite subset of B_a for some a<l and the result follows. By the same reasoning no finite subset of B_k entails Q.

Now we only have to verify that B_k is the book we are looking for.

Non-triviality: Q is not in B_0. Q is A_a for some a, and since Q entails Q, it would not have been added at stage a+1 (and clearly not at any other stage). So Q is not in B_k.

Inconsistency: P is a subset of B_0, so its members belong to B_k, so B_k is inconsistent.

Maximality: By reductio. Take any proposition C, and suppose that neither C nor ~C is in B_k. Then C is A_a for some ordinal a, and ~C is A_b for some b. Without loss of generality, suppose a<b. C is not in B_{a+1}, so some finite subset of $B_a \cup \{C\}$ entails Q. That is, for some finite subset X of B_a, $X \cup \{C\}$ entails Q. Similarly, for some finite subset Y of B_b, $Y \cup \{\sim C\}$ entails Q. Now B_a is a subset of B_b. So $X \cup Y$ is a finite subset of B_b such that $X \cup Y \cup \{C\}$ entails Q and $X \cup Y \cup \{\sim C\}$ entails Q. It follows that $X \cup Y$ entails Q. But this contradicts the lemma. Hence, either C or ~C is in B_k, and B_k is maximal.

All necessary truths are in B_k: Because all necessary truths are in B_0.

Closure under modus ponens: By reductio. Suppose R and R⊃S are in B_k, but S is not in B_k. R is A_a, R⊃S is A_b, and S is A_c for some ordinals a, b, c. WLOG assume that c is the greatest of these three. Then some finite subset of $B_c \cup$ {S} entails Q. Thus for some finite subset X of B_c, X ∪ {S} entails Q. But then X ∪ {R, R⊃S} is a finite subset of B_k which entails Q, contradicting the lemma. Hence if R and R⊃S belong to B_k, S also belongs to B_k. Q.E.D.

I do not know how to show that the propositions can be well-ordered; I am not certain that they can be. But I do not know how to show that the propositions cannot be well-ordered, and I don't know that they cannot be. So I still must withhold judgment about whether there are impossible worlds without inconsistent triples. However, Result 7 strikes me as an extremely interesting one, and it gives us an idea of what such worlds are like if there are any.

If it should turn out that there are no worlds of this sort after all, the semantical project of the next section cannot always be carried out precisely as it stands. I will mention how the semantics might be changed to get around this problem.

Impossible World Semantics for Propositional Attitudes

Jaako Hintikka has long advocated a possible worlds semantics of epistemic logic. One of the chief criticisms of his approach is that it seems to entail what has been called "logical omniscience." It appears that according to Hintikka's semantics one knows every proposition that is a logical consequence of a proposition that one knows.

This result seems unfortunate, since evidently one often does *not* know every consequence of what one knows. It normally takes some time to work

through an arithmetic problem, or one of Raymond Smullyan's logic puzzles, or a complicated modal argument. And if it takes time--even if only a short time--then there is some moment at which one knows something which entails the solution of the problem or the conclusion of the argument (if it is sound), without knowing the solution or conclusion itself. To take a particular case, many students of geometry know what is to be known about how to perform a geometric construction with straightedge and compass but do not yet know that it is impossible to trisect an arbitrary angle with these tools. (In fact, over the years there have been quite a number of people who have continued to believe that the angle could be trisected--and in many cases, that they themselves had trisected it--even after reputable mathematicians told them otherwise. The efforts of some of these hardy and foolhardy souls have been entertainingly chronicled by Underwood Dudley in A Budget of Trisections.)

For these reasons, it has been widely agreed that the set of propositions known by a given person is not closed under entailment.[72] And of course the same goes for the propositional attitudes of believing, hoping, imagining, etc. Hence the pickle for epistemic logic and logics of many other propositional attitudes. It looks as if the world-based semantics of epistemic logic & co. is fatally flawed.

In "Impossible Possible Worlds Vindicated," Hintikka argues that this apparent flaw is merely apparent since it depends on the mistaken assumption that every epistemically possible world is a logically possible world. Once we jettison that assumption, the path is clear for a world-based semantics for epistemic logic of the sort that Hintikka proposes. An example of a semantics which adopts this strategy is provided by Veikko Rantala in his "Impossible Worlds Semantics and Logical Omniscience."

[72] For Stalnaker's dissenting opinion, see his Inquiry.

Rantala wants to show that one can "develop an axiomatized logic for a propositional attitude such that it (a) indicates what is valid about the attitude (i.e., holds in all circumstances), (b) is sufficiently weak so that it would correspond to our intuitions in that it does not presuppose logical omniscience ... (c) is logically adequate in the sense of being consistent and complete" (107). Rather than limiting his account to some particular attitude, Rantala adopts the more general goal of satisfying (a)-(c) for a variety of propositional attitudes. As an example, he presents a logic whose modal operator L stands in for a propositional attitude not "epistemically weaker than" knowledge. The method used in producing this logic, Rantala says, can be adapted to produce logics of other propositional attitudes which also satisfy (a)-(c).

Hintikka's strategy of providing a semantics for such a logic involves the notion of epistemic alternatives. (Or as he puts it, epistemic a-alternatives. I'll occasionally omit the prefix.) They are "the contingencies which are left open by whatever a knows in W" (476), or the worlds "compatible with everything a knows in W" (475). The idea, then, is that a knows P if and only if P is true in each of a's epistemic alternatives, that is, just in case P is true in all of a certain class of worlds. This suggests that we may represent knowledge formally by a restricted necessity operator. The analysis of knowledge will turn out to be a semantics for a modal logic who necessity operator ranges over the epistemic a-alternatives.

It should be immediately clear that however Hintikka is using the word 'compatible' here, it cannot mean 'consistent with'. If it did, every consequence of what a knows would have to be true in the epistemic a-alternatives (since each alternative is maximal), and so a would be logically omniscient after all. Likewise, the contingencies "left open" by what a knows cannot be only those that are consistent with what a knows, for the same reason. Rather, what is true in a's epistemic alternatives will sometimes be inconsistent with what a

knows; this is precisely the reason for using an impossible worlds semantics to deal with the problem of logical omniscience. Unfortunately it is not immediately clear in what sense we ought to understand 'compatible' and 'left open by' if we wish to provide a Hintikkean semantics for epistemic logic. A number of options are available, and we will have to see which best serves our purposes.

The example logic Rantala considers, T_Ω, is an adapted form of the modal logic T; the only difference is that T_Ω has a restricted rule of necessity. Let the language of T_Ω include propositional variables, the connectives ~ and & (with $\vee$, $\supset$, and $\equiv$ defined in terms of these) and the operator L. Let p, q, r, ... be formulae of the language of T_Ω, and Form be the set of these formulae. Ω is to be a subset of Form which restricts *a*'s logical omniscience. Intuitively, it is the set of logical truths *a* knows. The axioms and rules for T_Ω are:

(PC) All propositional tautologies of the language.

(M_1) $Lp \supset p$

(M_2) $(Lp \,\&\, L(p \supset q)) \supset Lq$

(MP) $p, p \supset q / q$

(N_Ω) If $p \in \Omega$, p / Lp

Rantala notes that since each theorem of T_Ω is a theorem of T also, T_Ω is consistent.

Axiom M_1 and rule MP together require that whatever proposition is the object of the attitude L is true if T_Ω is an actual logic, so it makes sense that we take L to represent knowledge. If we wished to provide a logic for a weaker attitude such as belief, we would need to omit M_1; that is, we would need a base system even weaker than T. Alternately, M_1 could be restricted in a way analogous to the way in which N_Ω is restricted.

We should also note at this point that axiom M_2, though it idealizes our notion of knowledge to some extent, is not enough to guarantee the logical omniscience of the subject a. All it insures is that a knows the immediate consequences of what a knows when a knows that those propositions are immediate consequences of what a knows. It is entirely possible that a knows some proposition P which implies Q, but that a does not know that P implies Q, so that a is not logically omniscient and M_2 is not violated.

Below is Rantala's semantics for T_Ω. (Do not spend too much time on its details now; we will make some amendments to it momentarily.)

A 4-tuple M = (W, W*, R, V) is a T_Ω-model iff

(1) W is non-empty;

(2) $R \subseteq (W \cup W^*) \times (W \cup W^*)$ such that R is reflexive in W;

(3) V is a function V: Form $\times$ (W $\cup$ W*) $\rightarrow$ {0, 1} such that:

(i) For all $w \in W$,

$V(\sim p, w) = 1$ iff $V(p, w) = 0$

$V(p \& q, w) = 1$ iff $V(p, w) = V(q, w) = 1$

$V(Lp, w) = 1$ iff $V(p, w') = 1$

for all $w' \in W \cup W^*$ such that wRw',

(ii) For all $w^* \in W^*$,

if $V(p, w^*) = V(p \supset q, w^*) = 1$, then $V(q, w^*) = 1$,

(iii) For all $p \in \Omega$, $w^* \in W^*$,

if $V(p, w) = 1$ for all $w \in W$, then $V(p, w^*) = 1$.

We define formula p to be true in a T_Ω-model M = (W, W*, R, V) if and only if $V(p, w) = 1$ for all $w \in W$, and p is T_Ω-valid if and only if p is true in every T_Ω-model. (Note that being "true in" a model is a distinct notion from being "true in" a world. *Propositions* are true in worlds, whereas *formulae* may be true in

models if they are assigned the value 1 with respect to each member of the model's W-element.) Rantala goes on to prove completeness, i.e. that a formula is provable in T_α iff it is T_α-valid.

My original hope was to point out that normal and non-normal worlds (which Rantala declines to interpret, since, he says, appropriate interpretation will vary quite a bit from logic to logic) may be understood as certain possible and impossible worlds. Certainly Rantala's title hints at this kind of applied semantics. However, the members of W* need not be inconsistent. In fact, the completeness proof Rantala gives defines W* in such a way that it members must be consistent if the L operator is interpreted as 'α knows that'. Furthermore, the semantics does not actually require the members of W* to be maximal; we might just as well think of them as incomplete states of affairs or situations. If we add a maximality condition to the semantics, the given completeness proof fails, and it is not clear that there is any way of repairing it. Rantala's semantics can do its work without anything like impossible worlds, so his title is oddly misleading.

What I will attempt to do here, then, is to give a logic which differs slightly from Rantala's and a corresponding semantics whose elements are fittingly interpreted as possible and impossible worlds. Let $T_\alpha 2$ be like T_α except in that it omits T_α's (M_2). Of course $T_\alpha 2$ is consistent if T_α is. We will adapt the formal semantics by amending (3 ii) to

(3 ii*) For all $w \in W$, if $V(p, w) = V(p \supset q, w) = 1$, then $V(q, w) = 1$,

so that it applies only to possible worlds, and requiring maximality:

(4) For all $p \in$ Form, $w \in W \cup W^*$, $V(p, w) = 1$ or $V(\sim p, w) = 1$.

Now we want to prove completeness: a formula is $T_\alpha 2$-valid iff it is provable in $T_\alpha 2$.

Right-to-Left: The propositional tautologies are valid because of (3 i). (M_1) is valid by reflexivity of R, i.e. (2), and the last clause of (3 i). (MP) preserves validity by (3 ii*). (N_Ω) preserves validity by (3 iii).

Left-to-Right: Consider the 4-tuple <W, W*, R, V>. Let W be the family of maximal, $T_\Omega 2$-consistent extensions of $T_\Omega 2$. Let W* be the union (for all w) of all Alt(w), where

Alt(w) = {A | for all p, if Lp $\in$ w, then p $\in$ A and A is maximal}.

Define R by

$(w_1, w_2) \in R$ iff $w_2 \in Alt(w_1)$.

Define V as follows:

(y) For every w* $\in$ W*, V(p, w*) = 1 iff p $\in$ w*.

For every w $\in$ W,

V(p, w) = 1 iff p w (if p is a propositional variable),

V(~p, w) = 1 iff V(p, w) = 0,

V(p&q, w) = 1 iff V(p, w) = V(q, w) = 1,

V(Lp, w) = 1 iff V(p, w') = 1 for all w' such that w' $\in$ Alt(w).

First we show by induction on the complexity of a formula that for every p $\in$ Form and w $\in$ W,

(x) V(p, w) = 1 iff p $\in$ w.

This is straightforward when p is a propositional variable or when p is a negative or conjunctive formula. For the remaining case, let w $\in$ W, and suppose (x) holds when p=f. We show (x) holds when p=Lf. If V(Lf, w) = 1, then V(f, w') = 1 for all w' $\in$ Alt(w), i.e. f $\in$ w' for all w' $\in$ Alt(w). Now if a formula q does not belong to {p | Lp $\in$ w}, then there is some maximal set which does not contain q and does contain every p such that Lp $\in$ w. That is, there is some

member of Alt(w) which does not contain q. So if f $\in$ w' for all w' $\in$ Alt(w), then Lf $\in$ w.

Conversely, if Lf $\in$ w, then f $\in$ w' for all w' Alt(w). By (x) and (y), V(f, w') = 1 for all w' $\in$ Alt(w), so by the definition of V, V(Lf, w) = 1.

Now we show that <W, W*, R, V> is a T_Ω2-model.

Clearly W is non-empty.

R is reflexive in W: Since every member of W is T_Ω2-consistent and maximal, if w $\in$ W contains Lp, w contains p, so w $\in$ Alt(w). (Note that W is therefore a subset of W*.)

(3 i) is satisfied: Obvious from the definition of V.

(3 ii*) is satisfied: Follows from maximality and T_Ω2-consistency of members of W.

(3 iii) is satisfied: Let f $\in \Omega$, w* $\in$ W*, and V(f, w) = 1 for all w $\in$ W. By (x), f $\in$ w for all w $\in$ W. So f belongs to every maximal T_Ω2-consistent set, so it's a theorem. Since f $\in \Omega$, Lf $\in$ w for all w $\in$ W by (N_Ω), and thus Lf is also a theorem. So (for every w $\in$ W) f $\in$ w' for every w' $\in$ Alt(w). Since each member of W* belongs to Alt(w) for some w $\in$ W, f $\in$ w* for every w* $\in$ W*. So V(f, w*) = 1 for every w* $\in$ W*.

That's it. It should be clear that the elements of W $\cup$ W* are very naturally interpreted as possible and impossible worlds. All such elements are maximal (and must be if the argument for (x) is to succeed), some (the members of W) are consistent, and some (such as Form itself) are inconsistent.

The sketch above, I hope, serves to show that with impossible worlds the problem of logical omniscience is surmountable. World-semantics for logics

of the propositional attitudes ought not be deemed infeasible on its account. And so in this arena we see how impossible world theory shows the promise of sharing in the philosophical fruitfulness possible world theory has enjoyed.

Affinities and Tensions with Relevant Logic

Impossible worlds quite naturally come into view when the subject is the semantics of relevant logic. A number of authors (e.g., Routley et al. in Relevant Logics and Their Rivals) have suggested that impossible worlds could and ought to be used in just this setting. However, there are a couple of respects in which relevant logic and the impossible world theory I've offered make strange bedfellows. Both the affinities and the tensions ought to be mentioned before we sketch the proposed application.

Relevantists battle a recent tradition in philosophical logic that has often been opposed to intensions and meanings. The Stalnakerian way of distinguishing propositions by their entailment (strict implication) relations is such an extensional view. On this view, propositions are distinguished by their extensions in possible worlds, or even identified with the set of possible worlds in which they are true. Meaning, or content, is often thought to distinguish necessarily equivalent propositions (for reasons rehearsed in the first chapter), so clearly extensionalists must reject *that* sort of meaning, the sort that goes beyond a proposition's extension in possible worlds. Relevantists, in denying that a necessary truth is implied by any proposition, insist that implication involves a meaning or content other than the propositions' extensions. Relevantists will thus be sympathetic to my earlier remarks on propositional content and, if it is granted that states of affairs bear a strong resemblance to propositions, to the idea that the states of affairs are not to be distinguished only by their extensions. This sympathy marks an important similarity

only by their extensions. This sympathy marks an important similarity between the relevantist's project and my own.

However I have assumed that relevantism is false. My reasons for doing so do not spring from my ideas about impossible worlds. Rather, I think there is something to the criticism that relevance is an epistemic notion. A relation deserving of the name 'entailment' or 'implication', according to the relevantists, can only hold between propositions relevant to each other. Anderson and Belnap give the example of a mathematician who makes a conjecture about Banach spaces and claims in a footnote, "if [the conjecture] is false, it implies that Fermat's last conjecture is correct" (Entailment, 17). In the absence of any reason to think the conjecture and (now) Fermat's last theorem are connected somehow, they say, we have no reason to think the mathematician's claim true. This is so even if every possible world in which the conjecture is true is a possible world in which Fermat's last theorem is true.

The criticism, briefly, is that the relevantist position confuses implication with obvious implication, or easily demonstrable implication, or some such notion. The necessarily false conjecture *does* imply Fermat's last theorem, though there may be no straightforward, non-question-begging argument from the one to the other. The lack of such an argument is what makes the mathematician's assertion misleading. According to this criticism, the notion of relevance, like the notion of question begging, is a relative notion. What is relevant depends on which inferences are obvious. If there were an extraterrestrial race which, somehow, knew instinctively that the conjecture in question strictly implies Fermat's last theorem (without knowing that the conjecture was necessarily false or that the theorem was necessarily true), then it would surely be relevant to their mathematical investigation of Fermat's last theorem whether the conjecture about Banach spaces were true. If our concern is the *logical* relationship between the two propositions, we need

If our concern is the *logical* relationship between the two propositions, we need not and ought not take into account relevance, which may hold for some thinkers and not for others (and in any case will hold only contingently if one proposition strictly implies the other). Logic books quite correctly warn us against fallacies of relevance, just as they warn us against question begging, but that is no reason think that the idea of implication itself must involve relevance.

What does any of this have to do with my theory? For one thing, the relevantist might be inclined to challenge my claim (implied by TUB) that the books of states of affairs are not closed under *entailment*. It may be reasonably clear that books are not closed under strict implication--that, for example, *the Eiffel tower is in Germany* needn't be true in a state of affairs in which *12 has more prime factors than it has divisors* is true. But it is somewhat less obvious that the book on such a state of affairs needn't contain *12 has more than 5 prime factors* or *some number has more prime factors than divisors*. Specifically, it is not obvious that the claim that states of affairs are closed under relevant implication (and not under strict implication) lacks any motivation. Nonetheless, there are still a number of reasons for thinking that states of affairs are not closed under relevant implication. Chief among these: the theory that all states of affairs are closed under relevant implication seems to omit certain impossibilities, such as the impossibility that 12 have more prime factors than divisors and that no number have more prime factors than divisors.

One might also worry that my rejection of relevantism undermines the motivation for this part of the chapter. Why provide a semantics for a mistaken logic? Well, I'm happy to lend the relevantists a hand, mistaken or not. If we disagree about the logic of entailment, I hope there will be no hard feelings. But more to the point, the fact (if it is a fact) that the logic of

entailment does not require relevance does not show that relevance is of no use at all. One might very well want a relevant logic of (non-question-begging) assertability, say. The formalization of relevance *is* a useful enterprise, and one to which impossible world theory can contribute.

Impossible World Semantics for Relevant Logic

Relevant logic is generally motivated by the (so called) paradoxes of implication or entailment. On the prevailing notion of entailment, it is true that a necessary falsehood entails every proposition and that every necessary truth is entailed by every proposition. The relevantist aims to specify a notion of entailment according to which one proposition entails another only if the propositions are relevant to each each other in the appropriate way, and so to avoid the paradoxes.

Standardly, entailment is a modalized version of material implication: $\Box(A \supset B)$, where the box represents "metaphysical" or "absolute" necessity. One way in which a relevantist might hope to repair this conception is to replace the usual necessity operator with another necessity operator '■' which ranges over some impossible worlds as well as all possible worlds. If A is said to entail B only when $\blacksquare(A \supset B)$ is true, then a necessary falsehood will not entail everything, since in some impossible worlds in the range of '■' the falsehood is true, but not every other proposition is. Likewise necessary propositions will not always be true in the worlds in question, and so not every proposition will entail them.

My use of '■' here is merely heuristic; the relevantist needn't have a modal logic in mind at all. The idea is simply that there is some collection of "relevant" worlds which relate to relevant implication as possible worlds relate to strict implication. For a relevantist who favors a particular relevant logic L, the most natural candidate is the collection of worlds for which L holds.

One of most serious obstacles to such an approach is the difficulty of showing that there are impossible worlds for which the logic in question holds. We saw earlier that whether there are non-trivial impossible worlds for which modus ponens holds and in which every necessary truth is true is a difficult question. This tells us, at least, that despite the variety of impossible worlds, not just any condition is met by more than one world. (Any proposition of the form *many impossible worlds meet such-and-such a condition* is true in many worlds, but of course that is a rather different matter.)

It is an interesting question whether there is a consistent logic which holds only for possible worlds and the trivial world. Is there, for example, a non-trivial impossible world for which, say, S5 holds? As noted at the beginning of the chapter, our notion of "holds for" does not apply straightforwardly to S5 since its rule of necessitation applies only to the system's theorems. So let's suppose we have a version of S5 which absorbs the effect of this rule into its axioms: the usual axioms of S5 are axioms, as are the necessitations (and necessitations of necessitations, etc.) of the axioms, and modus ponens is the only rule of inference. Call this S5*. Does S5* hold for any non-trivial impossible world?

S5* does not hold for the preface world π, which has inconsistent triples and every necessary truth so that, by Result 6, modus ponens does not hold for it. S5* also does not hold for the world μ, which we constructed to be a non-trivial impossible world for which modus ponens holds. Since μ excludes every proposition entailed by a proposition P, μ excludes every necessary truth, so the axioms of S5* do not hold for μ. If the propositions can be well-ordered, Result 7 tells us that there are impossible worlds for which S5* holds, assuming S5* to be the sober metaphysical truth of the matter, so that its axioms are necessary truths.

The >-fragment of the relevant logic R ('R', for short) has the following rule of inference and axioms:

RR: p, p>q / q

RA1: p>p

RA2: (p>q) > ((r>p) > (r>q))

RA3: (p > (q>r)) > (q > (p>r))

RA4: (p > (p>q)) > (p>q)

I do not know how to show that R holds for a variety of impossible worlds. If R lacked its rule of inference, this task would be easy; there are, of course, many worlds in which all of the propositions with the form of one of the axioms are true. But a logic with no rules of inference is profoundly uninteresting. As with S5*, though, we can show there are such worlds if the propositions are well-orderable and R's axioms are necessary truths.

So the thought that R holds for certain non-trivial worlds is not implausible. Suppose we were able to overcome this obstacle for a relevant logic L (either R or some other), and call the worlds for which L holds the L-worlds. Then our semantic strategy would be to define L-models (W, W*, V) where W is the class of possible L-worlds (i.e. presumably all possible worlds), W* is the class of impossible L-worlds, and V is the assignment of sentence letters to propositions. The aim is to show that a formula is provable in L iff, under V and the standard interpretation of logical connectives, the formula expresses a proposition true in all worlds of all L-models.

The left-to-right direction is pretty easy; it is basically Result 1. The right-to-left direction is not quite that obvious, but it looks promising. Suppose f expresses a proposition P and is not a theorem of L. Formula f is not an axiom of L, so P is not true in some impossible worlds for which the axioms of L hold. The question now is whether this continues to be so when we add the requirement that L's rules of inference hold for the worlds under consideration.

It appears that it does. Why would P be true in *all* L-worlds if it were not expressed by a theorem of L? If the right-to-left direction is true in the case of L, then L is complete. Moreover, impossible worlds play an intuitive and helpful role in the semantics of L.

In Closing

Other applications of impossible worlds naturally suggest themselves. We might investigate, for example, how best to use impossible worlds in the semantics of paraconsistent logics, i.e. those logics with non-trivial, inconsistent extensions. The Notre Dame Journal of Formal Logic's 1997 symposium on the topic includes proposals to use impossible worlds not only in paraconsistent logic, but also the theory of information, the theory of counterpossibles, and the analysis of inconsistent fiction.

A great deal remains to be done. Both of the main examples of this chapter deserve more detailed treatment. Furthermore, different accounts are bound to have somewhat different ontological assumptions, so it will be important to see what is presupposed by the various accounts, and whether these accounts can be adjusted to contrary assumptions. There are also the unanswered questions of Chapter 3 and 5, and questions in a similar vein. But what we have seen so far is enough to substantiate the claim that impossible worlds are of significant philosophical interest. We can expect the potential uses of impossible worlds to be no less widespread than modality itself.

WORKS CITED

Adams, Douglas. Dirk Gently's Holistic Detective Agency. Simon and Schuster, 1987.

Anderson, Alan Ross and Belnap, Nuel D. Entailment. Princeton University Press, 1975.

Barwise, John. "Information and Impossibilities," *Notre Dame Journal of Formal Logic*, **38**: 488-515, 1997.

Bernays, Paul and Fraenkel, Abraham. Axiomatic Set Theory. North-Holland, 1958.

Chisholm, Roderick. "Identity Criteria for Properties," *The Harvard Review of Philosophy*, 14 - 16, Spring 1994.

_____. "An Intentional Explication of Universals," *Conceptus* **66**: 45 - 48, 1992.

_____. On Metaphysics. University of Minnesota Press, 1989.

Dudley, Underwood. A Budget of Trisections. Springer-Verlag, 1987.

Goble, Louis F. "Grades of Modality," *Logique et Analyse* **51**: 323 - 34, 1970.

Grim, Patrick. "Logic and Limits of Knowledge and Truth", *Noûs* **22**: 341-367, 1988.

_____ and Plantinga, Alvin. "Truth, Omniscience, and Cantorian Arguments: An Exchange," *Philosophical Studies* **71**: 267 - 306, 1993.

Hintikka, Jaakko. "Impossible Possible Worlds Vindicated," *Journal of Philosophical Logic* **4**: 475 - 484, 1975.

_____. "Surface Information and Depth Information," in Hintikka, Information and Inference. D. Reidel Publishing Company, 1970.

Hrbacek, Karel and Jech, Thomas. Introduction to Set Theory. Marcel Dekker, Inc., 1978.

Kim, Jaegwon. Supervenience and Mind. Cambridge University Press,

Kim, Jaegwon. Supervenience and Mind. Cambridge University Press, 1993.

Kripke, Saul. "Semantical Considerations on Modal Logic," *Acta Philosophica Fennica* **16**: 83 - 94, 1963.

_____. Naming and Necessity. Harvard University Press, 1972.

Kvart, Igal. A Theory of Counterfactuals. Hackett Publishing Co., 1986.

Lewis, David. Counterfactuals. Harvard University Press, 1973.

_____. "Causation," "Counterfactual Dependence and Time's Arrow," and "Counterfactuals and Comparative Possibility" in Philosophical Papers, Vol. II. Oxford University Press, 1986.

_____. On the Plurality of Worlds. Basil Blackwell, 1986.

Loux, Michael. "Introduction: Modality and Metaphysics," in The Possible and the Actual. Cornell University Press, 1979.

Mares, Edwin D. "Who's Afraid of Impossible Worlds?" *Notre Dame Journal of Formal Logic*, **38**: 516 - 526, 1997.

McNamara, Paul. "Does the Actual World Actually Exist?" *Philosophical Studies* **69**: 59-81, 1993.

Mortensen, Chris. "Peeking at the Impossible," *Notre Dame Journal of Formal Logic*, **38**: 527 - 634, 1997.

Naylor, Margery Bedford. "A Note on David Lewis's Realism about Possible Worlds," *Analysis* **46**: 28 - 29, 1986.

Perry, John. "From Worlds to Situations," *Journal of Philosophical Logic* **15**: 83 - 107, 1986.

Pollock, John. "Four Kinds of Conditionals," *American Philosophical Quarterly* **12**: 51 - 59, 1975.

Plantinga, Alvin. The Nature of Necessity. Oxford University Press, 1974.

_____. "Actualism and Possible Worlds" in Loux, Michael (ed). The Possible and the Actual. Cornell University Press, 1979.

Priest, Graham. "What is a Non-normal World?" *Logique & Analyse* **130-140**: 291 - 302, 1992.

Rantala, Veikko. "Impossible Worlds Semantics and Logical Omniscience," *Acta Philosophica Fennica* **35**: 106- 115, 1982.

Restall, Greg. "Ways Things Can't Be," *Notre Dame Journal of Formal Logic*, **38**: 583 - 596, 1997.

Routley, Richard. Relevant Logics and Their Rivals. Ridgeview Publishing Company, 1982.

Sharlow, Mark. "Lewis's Modal Realism: A Reply to Naylor," *Analysis* **48**: 13 - 15, 1988.

Stalnaker, Robert. "A Theory of Conditionals," in N. Rescher (ed). Studies in Logical Theory. Blackwell, 1968.

Strand, Jonathan. "The Semantics of Conditionals," dissertation, University of Notre Dame, 1991.

Tidman, Paul. "Conceivability as a Test for Possibility," *American Philosophical Quarterly*, **31**: 297 - 309, 1994.

van Inwagen, Peter. "Two Concepts of Possible Worlds," Midwest Studies in Philosophy, **11**: 185 - 213, 1986.

_____. "Modal Epistemology," *Philosophical Studies* **92**: 67 - 84, 1998.

Vander Laan, David. "The Ontology of Impossible Worlds," *Notre Dame Journal of Formal Logic*, **38**: 597 - 620, 1997.

Zagzebski, Linda. "What If the Impossible Had Been Actual?" in Beaty, Michael (ed). Christian Theism and the Problems of Philosophy. University of Notre Dame Press, 1990.