

**POSSIBLE WORLDS:
WHAT THEY ARE GOOD FOR AND WHAT THEY ARE**

by

Alexander Robert Pruss

B.Sc. (hon.), University of Western Ontario, 1991

Ph.D., University of British Columbia, 1996

Submitted to the Graduate Faculty of
Arts and Sciences in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh

2001

[committee signature page goes here]

Copyright © 2001 Alexander R. Pruss

*For my father,
who taught me also the value of philosophy*

POSSIBLE WORLDS: WHAT THEY ARE GOOD FOR AND WHAT THEY ARE

Alexander R. Pruss, Ph.D.

Advisor: Nicholas Rescher

University of Pittsburgh, 2001

This thesis examines the alethic modal concepts of possibility and necessity. It is argued that one cannot do justice to all our modal talk without possible worlds, i.e., complete ways that a cosmos might have been. I argue that not all of the proposed applications of possible worlds succeed but enough remain to give one good theoretical reason to posit them. The two central problems now are: (1) What feature of reality makes correct alethic modal claims true and (2) What are possible worlds?

David Lewis makes possible worlds be concretely existing universes. Unfortunately, I show Lewis's account involves set-theoretic, ethical, inductive and probabilistic paradoxes, and commits Lewis to an objectionable form of primitive modality that governs the choice of the counterpart relation. The most promising contemporary alternatives to Lewis's theory have been the worlds of Adams and Plantinga constructed out of Platonic entities such as maximal collections of consistent propositions. However, these approaches fail to provide a satisfactory answer to the question of what makes true modal claims true. I also criticize some alternative accounts.

Finally, I discuss and combine two historical approaches. The first is an Aristotelian approach that says a non-actual event is possible is to say that some actual substances could have initiated a causal chain that could lead up to the event in question. However, it can be shown that some plausible global possibility claims can be made true on this account only if there is a necessarily existent first cause (or aggregate of first causes) capable of initiating very different universes. On the other hand, Leibniz made possible worlds be ideas in the mind of an omniscient necessarily existent deity. Leibniz fails to explain what it is that makes these possible worlds *possible*, but if we were willing to combine his story with the conclusion drawn from the Aristotelian one, we could get the following story: Possible worlds are ideas in the mind of an omniscient deity and what makes them *possible* is that this deity has the Aristotelian capability of initiating causal chains that can lead to them being actualized.

Preface

I would like to thank Robert Brandom, James Dreier, John Earman, Gregory Fitch, Richard Gale, Jeremy Heis, John Leslie, Kenneth Manders, David Manley, William McEnaney, Robert Koons, John Norton, Alvin Plantinga, Nicholas Rescher, Ernest Sosa, Donald Turner, Peter van Inwagen and two anonymous referees for encouragement, interesting discussions and/or comments on various parts of this project.

And above all, I would like to thank my wife and parents for all their loving encouragement in my studies.

Table of Contents

[Preface](#)

[Table of Contents](#)

[List of Tables](#)

[Part I. Introduction](#)

[Section 1 Generic definitions and basic modal realism](#)

[Section 2 Metaphysical *versus* logical possibility?](#)

[Section 3 S5](#)

[Section 4 Six kinds of views on possibility](#)

[4.1 Parmenides, Spinoza, Leslie and Rescher](#)

[4.2 Leucippus, Democritus, Meinong, Lewis and Aristotle](#)

[4.3 The linguistic view](#)

[4.4 The propositional primitive modality view](#)

[4.5 Aristotle again and branching](#)

[4.6 Leibniz](#)

[Section 5 Impossibilia](#)

[Part II. Applications and pseudo-applications.....](#)

[Section 1 Modality](#)

[1.1 Box and diamond](#)

[1.2 The global nature of modal claims](#)

[1.3 Supervenience](#)

[1.4 Transworld comparison](#)

[Section 2 Counterfactuals and causality](#)

[Section 3 The direction of time](#)

[3.1 Lewis's argument](#)

[3.2 The pulled plug](#)

[3.3 The button on the laser](#)

[3.4 The general case](#)

[3.5 Another counterexample](#)

[3.6 Conclusions](#)

[3.6.1 The problem with Lewis's approach](#)

[3.6.2 A fix](#)

[3.6.3 McCall's approach](#)

[Section 4 Propositions](#)

[4.1 Unstructured propositions](#)

[4.2 Structured propositions](#)

[Section 5 Properties](#)

[Section 6 Overall assessment](#)

[Part III. The Lewisian ontology of extreme modal realism](#)

[Section 1 The Lewisian account of possible worlds](#)

[Section 2 Identity vs. counterpart theory](#)

[2.1 Arguments for counterpart theory](#)

[2.2 Arguments for identity theory](#)

[2.2.1 General arguments](#)

[2.2.2 Attributions of ability](#)

[2.3 Conclusions about identity and counterpart versions of basic EMR](#)[Section 3 Indiscernible worlds?](#)[Section 4 Lewis's arguments for his ontology](#)[4.1 The analysis of actuality argument](#)[4.2 The cost-benefit argument](#)[4.2.1 What is modality?](#)[4.2.1.a A solution to the Parmenidean challenge and the mystery of modality](#)[4.2.1.b Essential properties, counterpart relations and primitive modality revisited](#)[4.2.2 Other applications and assessment](#)[Section 5 Objections to Lewis's account of actuality](#)[5.1 Common sense](#)[5.2 Indexicality and ordinary language](#)[5.2.1 "The actual world" means "the world in which I exist"](#)[5.2.2 "The actual world" means "the universe that contains *this*"](#)[5.2.3 Gale's general argument](#)[5.2.4 Nor can one simply substitute](#)[Section 6 The possibility of non-spatio-temporally related co-actual entities](#)[Section 7 Cardinality and the "set" of all possible worlds](#)[7.1 Introduction](#)[7.2 There is no set of all possible worlds](#)[7.3 Lewis's proviso](#)[7.4 What should Lewis do about this?](#)[Section 8 Ethical issues](#)[8.1 The ethical objection](#)[8.2 The indexicality of morality](#)[8.3 Problems for counterpart theory](#)[8.3.1 A transworld counterfactual](#)[8.3.2 The first case: saving a life](#)[8.3.3 The second case: self-torture](#)[8.3.4 A general reply](#)[8.3.5 The need for consequentialistic considerations in ethics](#)[8.4 The logical independence of free choices](#)[8.5 Fatalism](#)[Section 9 Induction and actuality](#)[9.1 Introduction](#)[9.2 Neutral ground](#)[9.3 The Humean argument and a reply Lewis cannot give](#)[9.4 Conclusions](#)[9.5 Cardinalities of regular and irregular continuants](#)[9.5.1 The deterministic case](#)[9.5.2 The non-deterministic case](#)[Section 10 Possible worlds and probability](#)[Section 11 The epistemological objection](#)[Section 12 A final assessment of extreme modal realism](#)[Part IV. Ersatz ontologies](#)[Section 1 The general strategy](#)

[Section 2 Linguistic approaches](#)

- [2.1 The basic idea](#)
- [2.2 Objection from alien properties](#)
- [2.3 Primitive modality](#)
- [2.4 The arbitrariness objection](#)
- [2.5 Mathematical constructions](#)
- [2.6 Combinatorialism and Armstrong](#)

[Section 3 Platonic approaches](#)

- [3.1 Platonic approaches](#)
- [3.2 What are propositions](#)
- [3.3 The set-theoretic objection](#)
- [3.4 Kripkean objections](#)
- [3.5 How do propositions represent?](#)
- [3.6 *What* do modal propositions represent? The unmet Parmenidean challenge and the need for an analysis of possibility](#)
- [3.7 Affirmations of necessary existence](#)
- [3.8 What are propositions?](#)

[Section 4 Conclusions](#)[Part V. Rescher's critique of the very notion of possible worlds](#)[Section 1 The argument](#)[Section 2 The individuation principle in the argument](#)[Section 3 Counterexamples to Rescher's argument against possible worlds](#)[Part VI. Aristotelian-Leibnizian theistic ontology](#)[Section 1 Leibniz's approach](#)

- [1.1 Leibniz's argument for the existence of God and the explanation of the nature of abstracta.....](#)
- [1.2 A solution to the problem of how the possible worlds represent](#)
- [1.3 The Parmenidean challenge revisited and an assessment of the Leibnizian approach](#)

[Section 2 Aristotelian possibility and causality](#)

- [2.1 Proximate matter, proximate cause and Parmenides](#)
- [2.2 The basic Aristotelian view](#)
- [2.3 Global possibilities: a difficulty](#)
- [2.4 Primitive causal modality and counterfactuals](#)
- [2.5 The necessary existence of a first cause: a proof](#)
- [2.6 The capability of producing an effect makes it true to say that the effect is possible](#)
- [2.7 The Humean objection from the compatibility of all things](#)
- [2.8 The free will of creatures](#)
- [2.9 Branching worlds all the way down](#)

[Section 3 The challenges to Lewis's ontology no longer a problem](#)

- [3.1 Primitive modality](#)
- [3.2 Counterpart vs. identity theory](#)
 - [3.2.1 Transworld identity](#)
 - [3.2.2 Haeceities](#)
 - [3.2.3 The identity of indiscernibles and a causal-historical account of identity](#)
 - [3.2.4 The divine ideas as differing between worlds](#)

[3.2.5 Conclusions about identity](#)

[3.3 Attributions of ability](#)

[3.4 No set of all possible worlds](#)

[3.5 The ethical and probability-theoretic objections](#)

[3.6 The inductive objection](#)

[Section 4 Rescher's challenges no longer a problem](#)

[Section 5 Is God omnipotent if logical possibility is defined in terms of his power?](#)

[Section 6 How we know what is possible](#)

[Section 7 The mystery and obscurity objections](#)

[7.1 The ontological mystery](#)

[7.2 Obscurity](#)

[Part VII. Final conclusions](#)

[Section 1 Cost-benefit arguments for the Aristotelian-Leibnizian ontology of possible worlds](#)

[1.1 Modality](#)

[1.2 Explanation of what propositions are](#)

[1.3 Not a completely new ontology](#)

[1.4 Connection with theistic arguments for the existence of God](#)

[Section 2 Conclusions](#)

[Appendix: Reinterpretation of functional systems: A sketch](#)

[Bibliography](#)

List of Tables

Table 1: Summary of Conclusions..... 228

Part I. Introduction

Section 1 Generic definitions and basic modal realism

Modal assertions involving possibility and necessity are a part of our ordinary languages as well as of our philosophical patrimony. Confining ourselves to the non-modal, there are many things we could not say. We could not mark the difference between a unicorn^[1], which *could* exist, and the square circle, which *could not*. Modality is a natural way of marking the difference between, on the one hand, the relation of Smith's being a bachelor to Smith's being unmarried, and, on the other hand, the relation of Smith's being fifty-feet tall to Smith's not being a mammal. Someone *could not* fail to be unmarried if he is a bachelor, but he could be a mammal even if he is fifty feet tall—though *in fact* no mammal is that tall.

It is important for ethical purposes to say what *could* have been done but was undone, and what would have happened *were* it done. It is generally agreed that in *some* sense of “possible” a human being can only be held responsible for an act if it was at least logically possible that he avoid it. When we say that moral worth supervenes on actions and non-moral circumstances, we are saying that it *could not* be the case that someone's moral worth was different though his actions and the non-moral circumstances were the same.

When we discuss the problem of evil, we sometimes wonder whether it is *possible* for God and evil to co-exist, a different problem from the *de facto* question of whether the evils of this world make the existence of God probable or not.

When we talk of natural objects, we often cannot specify the kind that the object falls into without talking of dispositional properties, that is properties that *would* be actualized were circumstances different. Something might in fact live all its life just like a horse, but if it is true that *were* it poked in the underbelly, where in fact it never was poked, it *would* suddenly and naturally sprout wings and fly away, it is not a horse.

Our expressive capabilities would be greatly impoverished without *can be, might be, must be, is possible, is necessary, would be* and their ilk. We need these terms to talk of the reality around us. Yet, paradoxically, talk involving possibility often does not appear to be *about* anything real. The unicorn that is possible does not exist, I have not done otherwise than I have, and the actions and non-moral circumstances are as they are.

One way to conceptualize modal notions is to think of a “possible world”, a way (with “way” understood so broadly as not to prejudice the ontological question of *what* possible worlds are) that a cosmos could have been. Different possible worlds are different ways that our world could have been.

The main alternative to thinking of modality in this global sense is thinking of it in a local sense, of thinking of alternative ways that *portions* of this world could have been. Such piecemeal modality is what ordinary language normally engages in. When we say that Hitler might never have been born, we do not generally just mean that there is some possible world in which he doesn't exist—e.g., a world at which the universe is and has always had an unchanging constant energy density. We mean that that portion of this world which corresponds to the birth of Hitler might not have been even though much of the rest of the world, especially at least the distant past prior to Hitler's birth, was pretty much the same, and the laws of nature were those that we have. What exactly is to be kept fixed in this “might never have been born” claim depends on the context. Thus, while apparently speaking only of portions of worlds, the context determines what *whole* worlds we are speaking of, namely what portions of the actual world are supposed to be imagined as remaining in that possible scenario in which Hitler had never been born. To disambiguate our ordinary piecemeal talk of possibility we bring in whole possible worlds.

The need to talk of whole worlds is shown particularly clearly when we make counterfactual utterances. For we are wont to ask questions like: “How might or would have the course of history gone had Hitler never been born?” And on a plausible account of how to answer such questions, we should think of *whole* worlds in which Hitler was not born, and to say what holds in such worlds.^[2] Given what our context fixes, namely most events

prior to Hitler's birth and the laws of nature, we can say certain things about what happens in those worlds. For instance, the course of events in other galaxies is the same as in the actual world—whether the awful events of the 20th century occurred or not is not going to affect what happens in other galaxies, if only because the information about them, traveling at the speed of light, has not yet arrived there. But the course of local history would have been at least somewhat different, and we can speculate about how it might have been different. Our ease in saying in the same breath that events in other galaxies would have been the same but the events here would have been different does indicate that it is appropriate to analyze counterfactual situations holistically.

Moreover, what is possible in a portion of the world may well depend on the rest of the world. For instance, what happens categorially here presumably depends on what the laws of nature are that hold. It is impossible for there to be a world with exceptionless laws of nature like ours but where things don't fall when dropped under appropriate conditions; however, apart from such laws, it is certainly possible. It is impossible that there be unjustified evil in a local portion of the universe if there is an all-powerful, all-knowing and all-good deity in the universe. Moreover, when there is such a deity, then what evils can exist in a portion of the world may well depend on what happens elsewhere in the world, since the justification of some evil in one portion of the world may depend on events elsewhere. Our ordinary modal claims need to be contextually disambiguated, and when thus disambiguated are seen as involve whole possible worlds. Because of all this, possibility and necessity *prima facie* require reference to be made to whole possible worlds, and so one should try to make sense of possible worlds.

In conversation, Rescher has objected that the holistic intuitions apply to physical (or at least, I suppose, causal) and not necessarily logical possibility. However, at least we should leave open the option that logical necessity might involve some holistic aspects. For instance, suppose that the essentialist intuitions are correct that something could not be water were not certain laws of nature in place. Then the claim that it is possible that, say, it is possible for the water in a glass to fail to be gravitationally attracted to the center of the earth might be a *global* claim for it is a claim that there could exist certain counterfactual laws of nature, which could be global ones, and that water could logically co-exist with these laws. (Note, by the way, that the truth value of this claim is unclear.) Of course maybe laws of nature will turn out not to be global, but then specifying that they lack global modal oomph will itself be a global claim.

Given a basic notion of possible worlds, whatever their ontology, we need some correlative notions. By “the (or our) cosmos” I shall mean the aggregate (i.e., mereological sum) of all actually existing things. By “the (or our) universe” I shall mean the aggregate of all actually existing spatio-temporal things. Each world represents or corresponds to a way the cosmos could have been. In what way this representation works is open at this point of the investigation. One of the worlds shall be distinguished as “the actual world”, i.e., the world that represents the way *our* cosmos in fact, or actually, is. An individual “exists in” a world *w* if, were that world actual, that individual would exist, or, equivalently, if *w* represents the cosmos as containing that individual. A proposition is “true at” a world *w* if, were that world actual, that proposition would be true, or, equivalently, if *w* represents the cosmos as satisfying that proposition. Occasionally, the term “domain” will be used for the collection of all possible individuals that exist in a given world.

What the notions of “represents”, “actual”, “exists in” and “true at” really signify will depend on what our ontology of possible worlds is. There are many possible such ontologies. There is the crazy one, which nonetheless will be conceptually useful at times to keep in mind, that there necessarily is a Platonic library somewhere which contains physical books, of infinite size, each of which gives a maximal consistent description of

[3] a cosmos in some fixed language. On this view, a world is one of these books. A world *represents* some possible way of being a cosmos if the book that the world is describes the way that cosmos would be correctly. A world is *actual* if everything written in it is true. A proposition is *true at* a world if it is expressed by some sentence in the book. An individual *exists in* a world if the world describes the individual as existing.

Other theories will have other renderings of the basic notions. For instance, David Lewis thinks that all

possible ways that the universe could be is a way that some concretely existing universe really is. Moreover, cosmos and universes are the same for him. Thus, worlds are concrete universes. A world represents some cosmos if it *is* that cosmos. The *actual world* is the world we inhabit. A proposition is *true at* a world if it truly describes a state of affairs obtaining in that world. An individual *exists in* a world if it inhabits that world.

A propositional Ersatzist may take a world to be a maximal collection of compossible propositions. The *actual world* is the collection all of whose propositions are true. A world corresponds to a cosmos by having as its members propositions true of that cosmos. A proposition is *true at* a world if it is a member of it. An individual *exists in* a world if some proposition in that world says that the individual exists.

Leibniz, on the other hand, thinks that worlds are maximally consistent ideas in the mind of God. The actual world is the idea that God has chosen to actualize. An idea corresponds to a universe by being a mental representation of it. A proposition is *true at* a world if it is a part of, or maybe represented by, that world. An individual *exists in* a world if the idea represents him as existing.

We can now give a possible worlds semantics for possibility and necessity claims. It is possible that p providing there is a world w at which p is true. It is necessary that p providing p is true at every world. Having possible worlds lets us consider “local” and “global” modalities in a uniform way. When I say “Hitler might not have existed” in an ordinary way, I am saying that the proposition that Hitler does not exist is true at some world which matches ours in various relevant respects. When I say “It is logically possible that unicorns exist”, I may just be making the claim that the proposition that unicorns exist is true at some world, without putting any restriction on which worlds are relevant here.

Some further terms will be useful. A proposition is *contingent* providing it is true at some but not all worlds, i.e., providing neither the proposition nor its negation is a necessary truth. An individual x is a *necessary being* if it exists in all worlds. An individual is a *contingent being* if it exists at some but not all worlds. There is no relevant sense that I am aware of in which one can say that there “are” impossible beings, so I shall not define them. [4] Occasionally, I shall use δp and Δp to mean “necessarily p ” and “possibly p ”, respectively.

Section 2 Metaphysical versus logical possibility?

The modality in connection with which the possible worlds are possible is what is often called “metaphysical” possibility, with the paradigmatic example being that if Kripke is right, then it is metaphysically impossible that water fail to be H_2O . Some have argued that there are in fact two *different* kinds of modality. Some propositions, such as the proposition that H_2O contains hydrogen atoms are *logically* necessary since it is logically necessary that anything that has two atoms of hydrogen and one atom of oxygen in each molecule (and that, after all, is the definition of “ H_2O ”) contains hydrogen atoms. But that water *is* H_2O is a different kind of necessity, since it is not one that follows from the logic of the terms involved.

But consider the claims:

- (1) H_2O contains hydrogen atoms.
- (2) Water contains hydrogen atoms.

The defender of a distinction between logical and metaphysical possibilities claims that (1) and (2) have different modal status.

As an opening gambit, the Kripkean can reply that they cannot have different modal status, because modal status belongs to propositions, not to sentences, and (1) and (2) express the same proposition, and hence have the same modal status by Leibniz’s law. The defender of the distinction between necessities can either deny that (1) and (2) express the same proposition, or claim that they differ in modal status as *sentences*. The latter claim I have no need to dispute, since I can simply confine my account to that of the modal status of propositions.

But in fact the claim of a variance in the modal status of the two sentences is dubious. What does it mean? That sentence (2) could have expressed a false proposition? Yes, doubtless, but so could (1): after all, it

might have been uttered in a language where H_2O means “two electrons and one photon”. Nor will it do to say that (2) might have been true in *our* language. For in a language in which (2), or rather the proposition expressed by (2), is true, “water” was defined by pointing to some other liquid, and hence the meaning of the word is different, and hence the language is different.

One might also say that the difference between the sentences is that we can know *a priori* that (1) expresses a true proposition. If this is what is meant by claiming that the modal status is different, I concur, but note that the difference is epistemological not ontological. Perhaps there are truths of arithmetic we cannot know *a priori*. It is not implausible to suppose we have an intuitive grasp of only finitely many axioms of arithmetic, and perhaps we only have first-order resources available to us in connection with arithmetical truths. But then by Gödel’s incompleteness theorem, some truths of arithmetic cannot be known *a priori* by us. But it is by no means obvious that it follows that these truths of arithmetic have an ontological and modal status different from the others. Only their status relative to us is different.

Consider now the alternative of claiming that (1) and (2) express different propositions. I shall argue that nonetheless there is reason to think, if Kripke’s ostensive account of the naming of natural kinds is correct, that the two claims have the same modal status. For, suppose I point to Cicero and say:

- (3) Cicero is Cicero.
- (4) Tully is Cicero. ^[5]
- (5) *This* is Cicero.

Is there a difference in the modal status of what these sentences express? Surely not in the case of (3) and (4). “Cicero” and “Tully” are just as synonymous as “Sh’lomo”-in-Hebrew and “Solomon”-in-English or as “automobile” and “car” are, and hence (3) and (4) evidently express the same proposition. Admittedly, it takes greater ignorance to deny sentence (3). But if one understands the meaning of “Cicero” and “Tully” in context, it involves no greater self-contradiction. Sincere denial of (4) involves a failure to grasp that in this context “Tully” and “Cicero” are synonyms. But sincere denial of (3) involves a failure to grasp that the two inscriptions “Cicero” are synonyms--which in a different context they might not be (“Aristotle [Onassis] is not Aristotle [the Stagirite]”). Thus, the propositions expressed by (3) and (4) have the same modal status.

What, then, of (4) and (5)? Surely they, too, have the same status. For the use of “This” with a pointing finger renders it into a temporary name for Cicero, no different from “Tully” except in respect of the fact that we use the same inscription “This” in connection with many more meanings than we do “Cicero”. But with no change in propositions expressed, we might subscript all our demonstratives with unique symbols, and evidently things would stand no differently with the new version of (5) (e.g., “This₁₇₃₇₃ is Cicero”) than with (4) in terms of the modal status or proposition expressed. Hence, neither does (5) differ from (4), and hence from (3), in modal status. Note that I could imagine someone denying that (3) and (4), and also (4) and (5), express the same proposition, but the claims about identity of modal status do not seem open to question.

But, then, if “water” functions as a demonstrative pointing to the natural kind of that paradigm body of water that was involved in a Kripkean baptism thereof, and if that natural kind just is H_2O , then the difference between (1) and (2) is precisely that between (3) and (5), and hence involves no change in modal status.

Therefore, if we accept a Kripkean account of natural kind names, there is no distinction between logical and metaphysical necessity that could be used to distinguish (1) and (2). But we can do justice to our intuition that there is some sort of a difference between the two *sentences* by adverting either to the epistemological difference or to the following distinction. Some terms in English are defined by ostension and some verbally. “Bachelor” is defined verbally as an “unmarried man”. “Water” is defined ostensively as *that* natural kind. For any sentence S, let V(S) be the sentence obtained from S by first replacing each unquoted word that is verbally defined by its definition, iterating as many times as possible, and then replacing every remaining item of non-logical vocabulary by an undefined logical constant, a different constant for each word defined by a different ostensive act. Then, we can

say that S is *verbally necessary* if and only if $V(S)$ is a tautology. Thus, (1) is verbally necessary. To see this, suppose for simplicity that “ H_2O ” is defined as “a chemical constituted by molecules containing two atoms of hydrogen and one of oxygen” where each non-logical term here is not itself verbally defined.

Then, $V(1)$ is something like “A C constituted by M s containing two A s of H and one of O is a C constituted by M s containing two A s of H and one of O ”, where capital variable letters are logical constants, which is a tautology. But “water” is not a verbally defined term, so $V(2)$ is “ W is a C constituted by M s containing two A s of H and one of O ”, which is plainly non-tautologous. So we *can* do justice to both the intuition that there is a difference in the logical status of (1) and (2) and the argument that the propositions they express have the same modal status. Note that the same approach will show a distinction between the logical status of (3) and (5). $V(3)$ is “ C is C ” while $V(5)$ is “ X is C ”, so that only the former is verbally necessary. Interestingly, if “Tully” and “Cicero” are independently bestowed names, neither being verbally defined, $V(4)$ is “ T is C ”, and hence (4) is not verbally necessary.

I shall not use the concept of verbal necessity further. It depends too much on historical accidents, such as whether a second name was defined expressly a synonym for the first or was independently ostensively bestowed. These are important issues for the philosophy of language, but have little ontological significance in them for the structure of possible worlds or the modal status of propositions. I will talk of logical necessity, necessity *simpliciter* and metaphysical necessity as synonymous, for I do not think useful ontological distinctions can be made between them. None of these necessities are verbal. They are all “real necessity”, to use Kant’s term.

Section 3 S5

The modal logic assumed through most of this dissertation is S5, i.e., a logic satisfying the axioms:

$$(6) \quad \delta(p \dot{\vee} q) \dot{\vee} (\delta p \dot{\vee} \delta q)$$

$$(7) \quad \delta p \dot{\vee} p$$

$$(8) \quad \grave{\text{a}} p \dot{\vee} \delta \grave{\text{a}} p,$$

for all propositions p and q , and where $\grave{\text{a}}$ is short for “it is possible that” while δp is short for $\sim \grave{\text{a}} \sim p$, together with the “rule of necessitation” that if a formula is an axiom or theorem, then that formula prefixed by δ is also an axiom or

theorem. ^[6] This system is known as S5 and is characterized by an accessibility relation that is reflexive, symmetrical and transitive.

The most controversial axiom here is (8) which says that if something is possibly true, then it is impossible for it to fail to be possibly true. In modernity, the axiom goes back at least to Leibniz’s discussion of Descartes’ ontological argument. Descartes had defined God as a being that has all perfections, and one perfection is the property of necessary existence. Leibniz noted that Descartes’ argument was missing a crucial premiss, namely that it was *possible* for God to exist, and argued that once that premiss was added, the argument became valid. To show the essential use of S5, simplify the argument by supposing simply that God by definition a necessarily existent being. Then, Leibniz’s point is that once one adds the premiss that it is possible for God to exist, then it follows that God exists. For, then $\grave{\text{a}} \delta(\text{God exists})$, which implies $\delta(\text{God exists})$. But this implication is of course just an application of the contrapositive of (8).

Now, there certainly are kinds of modality for which (8) fails. For instance, suppose we consider a forward branching temporal structure, and say that p is possible at some point z in the structure providing p is true at z or at some future point that can be reached from z . Then, p is necessary at z providing that p does not fail at z or at any future point that can be reached from z . Then, it is possible (here and now) that I will at some point in my life run a marathon. But it is certainly not necessary that this is possible, because there is a future I can reach where my legs are cut off before I run a marathon and at a point in the future of that accident there will no longer be any reachable points at which I run a marathon.

One way to argue for (8) in our setting, however, would be to start with two intuitions. The first is that

things could not have been such that it would have been impossible for things to have been as they in fact are. However things might have gone, it still would have been true that they might have gone the way they in fact have gone. If things could have gone a certain way, then had they gone that way it would have been true that they could have gone the way they in fact went. This is the Brouwer axiom: $p \dot{E} \delta \dot{\Delta} p$. It tells us that the accessibility relation is symmetric.

The second intuition is that we when we talk about metaphysical possibility, we are talking about “ultimate” possibilities. Now, if we have a possibility operator $\dot{\Delta}$ such that $\dot{\Delta} p$ can hold without $\dot{\Delta} \dot{\Delta} p$ holding, then this operator does not tell us about *ultimate* possibilities. If it could have been that it could have been that p was true, then there is a real sense in which p could have been true. If we then deny that $\dot{\Delta} p$, we are saying that $\dot{\Delta}$ does not tell us of the ultimate possibilities there are, but of possibilities relativized to some way that things have been. Indeed, in such a case there *is* a reasonable more ultimate possibility operator, namely $\dot{\Delta} \dot{\Delta}$. Thus, if we are talking of ultimate possibilities, it is reasonable to require that $\dot{\Delta} \dot{\Delta} p$ should imply $\dot{\Delta} p$. This is the S4 axiom; it tells us that the accessibility relation is transitive.

But of course the Brouwer and S4 axioms, together with (6), imply (8). (Just apply the Brouwer axiom to $\dot{\Delta} p$ to conclude that $\dot{\Delta} p \dot{E} \delta \dot{\Delta} \dot{\Delta} p$; then use (6), S4 and the rule of necessitation to conclude that $\delta \dot{\Delta} \dot{\Delta} p \dot{E} \delta \dot{\Delta} p$.)

Alternately, one can argue that broadly logical possibility cannot have been different, since it is a matter of what propositions follow from what propositions (a proposition is possible if and only if its negation does not follow from it), and what follows from what could not have been different. Therefore, if $\dot{\Delta} p$, then it could not have been the case that $\sim \dot{\Delta} p$, i.e., $\dot{\Delta} p \dot{E} \delta \dot{\Delta} p$.

The S5 system of modal logic will be in the background for most of this dissertation. It is worth noting that the most prominent views of possibility to be considered, namely the Lewisian and ersatzist ones, are such as to leave little room for the denial of S5. The theistic account sketched at the end of the dissertation will also be such. One could thus give the following argument for S5: the best metaphysical accounts of possible worlds that we have require S5. However, I shall not give this argument here. Instead, I shall feel free to use S5 in my arguments for and against various metaphysical views of possibility, on account of the plausibility of S5 holding in the case of a notion of ultimate possibility.

Section 4 Six kinds of views on possibility

4.1 Parmenides, Spinoza, Leslie and Rescher

In his poem *On Nature*, Parmenides learns from the goddess that there are only two

ways of enquiry that are to be thought of. The one, that [it] is and that there is no non-being [*ouk esti mē einai*], is the path of Persuasion (for she attends upon Truth); the other, that [it] is not and that it needful that there be non-being [*esti mē einai*], that I declare to you is an altogether indiscernible track: for you could not know

[*gnoiēs*] what is not [*to ge mē eon*]—that cannot be done—nor indicate it. ^[7]

What is there to be said and thought must needs be: for there is being, but nothing is not [*esti gar einai, mēden d' ouk estin*]. ^[8]

The argument, insofar as it is more than just an assertion, is that non-beings plainly do not exist, and while if we speak and think, we are speaking of *something*.

We can put this argument in a more modern form by thinking about the controversial theory of the *truthmaker* of a proposition, a theory that I will in fact assume in a number of the arguments in this thesis. Realism requires that propositions be *made true* by something real. The proposition that there are horses is made true by the horses of this world. The proposition that Socrates is sitting is made true by Socrates' sitting, or the sitting Socrates *qua* sitting. The item in the world that a proposition is made true by is called its *truthmaker*. What exactly the truthmakers of propositions are depends on one's ontological system. For instance, if one is committed to an

Aristotelian worldview on which all there is are substances and their attributes, broadly construed, then the truthmaker of every true proposition will ultimately be substances and their attributes. An event ontology, on the other hand, may have the truthmakers be mereological sums of primitive events. But whatever the truthmakers are in one's ontology, in the case of propositions giving concrete facts about concrete entities, the truthmakers are made up of concrete things: tables, chairs, dogs, cats, sittings, shoutings, or the like. [\[9\]](#)

Moreover our language provides a way of giving the truthmaker of a proposition in a way that is neutral between ontological systems. To every declarative sentence there corresponds a participial nominalization. To "Socrates is a philosopher, was a war hero and taught Plato" there corresponds "Socrates' being a philosopher, having been a war hero and having taught Plato". To "Brutus betrayed Caesar" there corresponds "Brutus' having betrayed Caesar". To "There are horses" there corresponds "There being horses." If a sentence expresses a proposition, then the denotation of its participial nominalization is the truthmaker of that proposition. But what kind of an item "Brutus' having betrayed Caesar" denotes, whether it is ultimately a complex ultimately of substances and their attributes, or of events, or a fact in a world that is all that is the case, this is left open.

In any case, then, a proposition is true if and only if it has a truthmaker that really exists. This gives us a sense we can attach to Parmenides' cryptic remarks. If we know or speak truly, there must be an object of our knowledge or speech, namely the truthmaker of the proposition we know or express. It is this object that we know or speak of. The assertion that we cannot know or speak of what is not, then, becomes the claim that if we are to be right, there must *be* something we are right *about*: something that makes our affirmations true. Where the truthmaker is not, neither is there anything true.

Of course the notion of a truthmaker is going to be pointless unless we have some substantial theory about *what kinds* of entities can play that role. I can always say that the truthmaker of p is just *its being the case that p* , and if I do this for every true proposition, every true proposition will have a truthmaker in an apparently trivial way. [\[10\]](#)

However, saying this would not be so trivial. It is after all a substantial ontological claim that there are such things as *its being the case that p* . But this trivialization of the truthmaker theory does show that a criticism that some theory cannot provide a truthmaker for some proposition is short hand for an argument that we are not satisfied with just this trivial truthmaker for the proposition. One way to be dissatisfied is if one has certain ontological intuitions that do not fit with the idea of *its being the case that p* being a basic entity not to be further reduced. For instance, if one is an Aristotelian who thinks that all there are substances, their modifications and their relations, one will insist that truthmakers be nothing but substances, their modifications and their relations. This kind of an Aristotelian picture will in fact underlie a number, though not all, of the truthmaker-based arguments I will give. Moreover, it is in general preferable in a philosophical theory of some proposition p that one be able to say more about the truthmaker of p than that it is *its being the case that p* . Being able to say more about this truthmaker is itself a reason in favor that theory. Thus, even if we do not want to insist that *always* more can be said, we will *ceteris paribus* prefer a theory that says more.

While Parmenides did not deal with modality *per se*, we may be able to find in his writings an argument against change, which argument easily generalizes to an argument against modality.

And how could something that is [*to eon*] be in the future? How could it come to be? For if it came into being, it is not: nor is it if it is ever going to be in the future. [\[11\]](#)

A claim about the future must be made true by a truthmaker that is in the future. But then there is nothing *now* by which the claim is to be made true. Hence, when we say something *will be*, we are perforce speaking of something that is not, and thus not speaking truly.

The argument construed in this way may be criticized by a B-theorist for conflating existence *simpliciter* with merely *present* existence, but it is much more interesting in the modal case. We can say that the proposition that there *will be* a sea-fight tomorrow is made true by tomorrow's seafight, which exists *simpliciter*. But there is a

much deeper problem in the case of modal propositions. What makes true assertions of mere possibility? Suppose no sea-fight in fact occurs tomorrow. What, then, makes true the proposition that there *can* be a sea-fight tomorrow? If there will be a sea-fight tomorrow, then maybe the sea-fight that is tomorrow can make the proposition reporting its future occurrence true now. So, in parallel with this, the sea-fight that is merely possible maybe makes true the proposition that there can be a sea-fight. But this will not do, because “is merely possible” is truth-canceling in a way that “is tomorrow” is not. A merely possible sea-fight is not anything that exists. If it is not anything that exists, it cannot make anything true. But *what else* could the assertion that there *can* be a sea-fight be about, one asks, other than the future sea-fight?

Parmenides, not having a clear notion of modality, merely claims that his one reality is atemporally unchanging. But he could have used the same arguments to arrive at the further claim that this one reality *must* be as it is and can be no other, and doubtless if he were asked the modal question clearly, he would say this. And this is the Parmenidean puzzle of modality. It comes as a paradox and a problem. It seems that the proposition that there can be unicorns is, if anything, about unicorns—its truthmaker would have to be comprised, at least in part, of unicorns or their existing. Thus, its truthmaker does not exist, there being no unicorns and no existing of unicorns, and so the proposition is false. But it is paradoxical to admit that only the things that are could be.

If we are to avoid this paradox, we need to explain *what* the truthmakers of modal propositions are, and what it is about these truthmakers that makes them suitable to be such. Otherwise, we cannot have any realistic theory of the truth of modal propositions. It will be the purpose of this thesis to attempt an answer to the problem. Moreover, the attempt will be made within the confines of a broadly Aristotelian ontology, where the basic entities are substances and their modifications (properties and relations), an ontology which thus will tend to be unfriendly to the idea that such entities as *the state of affairs of it being possible that unicorns exist* could be primitive. However, although the intuitions behind this kind of an ontology drive much of the project, they are not presupposed by the individual arguments that will be given for the preferred answer to the Parmenidean problem and against the non-preferred answers.

Parmenides actually has *two* arguments against change. The second argument is an invocation of the Principle of Sufficient Reason (PSR) that says that every true proposition has an explanation. If what is should have come to be, then Parmenides asks:

[W]hat need would have driven it later rather than earlier, beginning from the nothing, to grow? ^[12]

This argument presupposes that the existent must have come from the non-existent, and Aristotle will later dispute this presupposition. But on this presupposition, the argument is a very interesting one. Later, St. Augustine in his *Confessions* would ask in the same way why God created the world when he did and not any earlier, and supply the answer that there was no time prior to the creation of the world so the question is malformed. Augustine’s answer is also a good reply to Parmenides’ argument for the eternity of the world.

But now that the PSR is on the table, one can use it to formulate a similar argument against alternate possibilities. Everything that is has a reason why it is the way it is. For a *reductio*, suppose that there are contingent propositions, and consider the conjunction *C* of all of them. This conjunction then has an explanation. This explanans is a proposition that is either contingent or necessary. But a necessary proposition cannot give a complete explanation of a contingent proposition, since the necessary proposition is just as true at possible worlds at which the contingent proposition fails and hence cannot supply the complete explanation of the contingent proposition. Hence, the explanans must be a contingent proposition. But then the explanans is one of the conjuncts in the explanandum *C*. Hence the explanans must be capable of explaining itself, which is absurd.

This is the PSR-based argument against modality. The argument can be attacked at three points. First, one may reject the PSR, in which case the argument has nothing to stand on. I shall not take this route for the very good reason that the theory of possibility that I shall end up endorsing as the only reasonable candidate theory of possibility entails the PSR. Secondly, one may allow that there be necessary propositions that can explain

contingent propositions. After all, the explanans need not explain the explanandum, so that the mere fact that the explanans holds in worlds where the explanandum does not is no problem. That Smith developed paresis is explained by his having syphilis and by certain statistical laws of nature, but these laws and the proposition that he has syphilis do not entail that he will develop paresis. Or, more to the point, if the libertarians are right, then that Smith made a free choice between alternatives *A* and *B* might well be a complete explanation of Smith having done *A*. If this is so, then it is epistemically possible that should there be a necessarily existent God who necessarily chooses between possible worlds to create, and the proposition that God chose between worlds to create might explain the existence of this world. Or, thirdly, one might allow that there are contingent self-explaining propositions. The only candidate for a contingent self-explainer that I know of is the claim that some person freely chose something. [\[13\]](#)

The PSR-based argument against modality goes hand-in-hand with a deterministic view of physical explanation, since such a view would make plausible the claim the explanans entails the explanandum, which I have rejected above. Something like this PSR-based argument is found in Spinoza's argument for the necessity of everything. It is essential to Spinoza's argument that for any thing, there must be a cause, namely ultimately the action of God, that determines not just that the thing is but the precise manner in which it is. For, otherwise, given the PSR and the insistence that the explanans entail the explanandum, the thing itself would have to explain why it is in the manner it is rather than in another possible manner, which explanation it cannot supply, Spinoza thinks. [\[14\]](#)

Recently, John Leslie and Nicholas Rescher have defended views that entail that there is only one possible world. Leslie proceeds through an "axiarchic principle", or principle of ethical requiredness (Leslie, 1997; Leslie, forthcoming). This principle corresponds to Plato's Form of the Good, and imposes on the world the necessity of satisfying certain conditions that make it be for the best. One argument for this principle would be through the considerations that fall under the head of "the anthropic principle" (cf. Leslie, 1990). The constants in the laws of nature (masses and charges of elementary particles and strengths of basic forces) appear to be calibrated in such a way as to make life possible. If they were somewhat different, and physics gives us no reason to think they could not be, life like we know it would not be possible. This provides evidence for the axiarchic theory, in that if the axiarchic theory is true, such finetuning is unsurprising, while if it is false, it is more surprising. However, obviously, this also provides evidence for other alternate theories, such as traditional theism, or Carter-type theories that claim that there is an infinite number of universes so it is probable that in some life would exist and those ones are the only ones that we can observe.

Rescher, on the other hand, has argued for a metaphysically necessary principle of optimality as a theory that explains why we find orderly laws of nature that can be mathematically formulated and understood by us. This principle ensures that, necessarily, things are for the best, understood in a Leibnizian sense as a balance between variety and lawlike unity. Of course, there are other theories that, if true, explain the same explanandum. Theism provides one such theory. [\[15\]](#)

Another would be a more limited version of Rescher's theory that merely claims that the *laws* of nature are necessarily for the best, while at least some of the contents of the world are contingent. This more limited version does not overturn modality, and rather than counting as a general view of possibility *simpliciter*, I shall simply take it to be a view about what possibilities are in fact there.

Rescher's view in its unlimited form appears to be subject to the following objection. First of all, if there is only one possible world, then saying that our world is the best of all possible worlds is not saying anything interesting. One could say with equal propriety that it is the worst of worlds. Consequently, the optimality principle cannot explain why the laws of nature are orderly, because if, *per impossibile*, the only possible world were one where they were disorderly, that world would also be the best.

Rescher's own reply (personal communication) is to distinguish a notion of logical possibility from a notion of metaphysical possibility. There is more than one logically possible world, and of these the best one is the one

that is metaphysically necessary. One might make non-trivial sense of the claim that the one and only possible world is optimal, e.g., by considering worlds that are metaphysically impossible recombinations of things in this world (cf. Armstrong, 1989) but nonetheless are modeled by mathematically coherent structures and hence capable of comparison to our world. But then the problem of evil shows its ugly head. The argument from evil against the existence of an omnipotent, omniscient and omnibenevolent deity, difficult enough as it is, takes a particularly difficult form if it is claimed that this world is in fact not just worthy of being made by such a God, but is the *best* conceivable world. Even if we could answer the original argument from evil, ^[16] claiming that this world is the best one is a yet further task. Moreover, the evidence from the apparent non-optimality of this world weighs against the evidence from the lawlike orderliness of the world. And there *are* theories that are better supported by the conjunction of these two pieces of evidence than Rescher's full theory: e.g., the more limited theory that says that the laws of nature are necessary and optimal, but the events in the universe, including freely done human actions, ^[17] are contingent.

Rescher also has another argument, not against possibility *simpliciter*, but simply against the existence of possible worlds. That argument is based on the impossibility of us individuating possible worlds, and I shall discuss it in due course in Part V.

4.2 Leucippus, Democritus, Meinong, Lewis and Aristotle

One extreme holds that merely possible worlds do not exist in any way, because our world is the only possible one. The other extreme view is that all possible worlds must exist. Leucippus' and Democritus' atomism could be an early representative of this view.

Leucippus holds that the whole is without bound...part of it is full and part void... Hence arise unboundedly many [*apeirous*] worlds, and are resolved again into these elements. ^[18]

If one takes the "unboundedly many" in the most extreme sense as involving all possibilities, then indeed we do get a view that all possible worlds exist.

This view would have been of interest merely to historians were it not for Alexius Meinong and David Lewis. Meinong sought to explain the intentionality of thought by invoking objects that correspond to all of our ideas, even those ideas that in our world are not exemplified. Thus, there are some things that don't exist.

David Lewis does this, too, at least for *possible* objects, but further organizes the things that don't actually exist into worlds. More precisely, Lewis posits that every possible world exists, and that these worlds are ontologically on par with one another. What makes two entities be a part of the same world is that they are spatio-temporally related. Thus, the different worlds are not spatio-temporally related, presumably unlike the worlds of Leucippus and Democritus. But, because of the ontological parity thesis, as in Leucippus and Democritus, the worlds are concrete entities just like our world, albeit ones that we could not possibly come in contact with. The horses in the other worlds are horses in exactly the sense in which the horses we know are, except that they are not spatio-temporally related to us.

Material reality is for Lewis much richer than we normally think. There exist dog-headed "men", and chimeras and unicorns—but not in our world. Fortunately, most of our language is relativized to our world, the *actual* world, which for Lewis is set apart from other worlds only indexically: the actual world is nothing but the mereological sum of all things that are spatio-temporally related *to us*. When I say, speaking ordinarily, that there are no unicorns, I mean that no unicorns *are actual*, that the *actual* world does not contain unicorns, i.e., according to Lewis that no unicorn is spatiotemporally related to us. In ordinary non-modal language, quantifiers are restricted to the speaker's world.

But of course, and this is the point of the theory, we can also speak with unrestricted quantifiers. Thus, we

can translate the assertion “Unicorns are possible” to say “There is a world w such that unicorns exist in w ”, involving ourselves in a quantification over all worlds. These sorts of quantifications give sense to modal language. Moreover, Lewis believes his theory of possible worlds makes it possible to give an account of various other philosophical notions. Thus, a proposition is a set of possible worlds—those worlds that it is true at—and a property is a set of individuals, with the set being allowed to extend beyond one world if desired.

This theory is elegant, solves many problems and appears coherent. But why should we think it *true*? Why should we think that reality is so much richer in material objects than we had thought? Lewis’s answer follows in the footsteps of Leibniz’s answer to Lady Masham’s worry about Leibniz’s system. Lady Masham wrote:

But it appears not yet to me that [your system] is more than a Hypothesis; for as Gods ways are not limited by our conceptions; the unintelligibleness or inconceivableness by us of any way but one, does not methinks, much induce a Beleeve of that, being the way which God has chosen to make use of. [\[19\]](#)

Leibniz replied *inter alia* with the following methodological observation:

Pour y joindre mes remarques, suivant vos ordres, je diray (1) qu’il semble que c’est quelque chose de considerable qu’une hypothese paroisse *possible*, quand toutes les autres ne le paroissent point, et (2) qu’il est extremement *probable* qu’une telle hypothese est la veritable. Aussi at-on toujours reconnu dans l’Astronomie et dans la Physique, que les hypotheses les plus intelligibles se sont trouvées veritables enfin: comme par exemple celle du mouvement de la terre, pour sauver les apparences des Astres ... [\[20\]](#)

The very fact that a theory gives a coherent account of difficult problems, where other theories have failed, is evidence for its truth, in philosophy as in science. Thus, Lewis thinks, we should believe his theory because it is elegant, solves many problems and, Lewis thinks, appears coherent.

It should not come as a surprise that such a drastic revision of the account of what we think there is as Lewis provides carries with it counterintuitive consequences. For instance, as we shall learn in Part III, were we to believe Lewis, we would have to become inductive sceptics and revise basic moral notions. This price would be too high. For just as the fact that a theory gives a coherent account of one thing provides evidence for the theory’s truth, likewise the fact that the theory gives rise to seemingly absurd consequences elsewhere, e.g., in issues of induction or morality, gives evidence against it. The theoretical benefits of Lewisian possible worlds theories will be critically considered in Section 5. If another theory can be found that has all or most of the benefits that survive this critical examination, but lacks the demerits of contradicting induction or morality, then that other theory is to be preferred. In Part VI it shall be argued that there is such a theory.

One final theorist who uses the same strategy as Lewis for grounding modality in a larger totality of existent things should be mentioned. This is Aristotle. The inclusion of Aristotle here may seem surprising, but in fact in Aristotle we find two different threads of thinking about modality. One of these threads involves a modal logic based on time. A proposition is *necessary* if and only if it holds at all times and possible if it holds at *some* time. Observe that in order to avoid the absurd consequence that all true propositions are necessary, we need to take an analysis of indexical sentences which makes a sentence like “It is now noon” express *one and the same* proposition at different times, which one proposition is true at noon but false at other times. (This is not the only way one could analyze indexical sentences. We might take “It is now noon” to express a different proposition at each different time during which it is uttered, and there is support in ordinary language for this. [\[21\]](#) But if we did this, then any proposition, however seemingly contingent, will always have the same truth-value at all times, and hence on Aristotle’s theory will either be necessary or impossible, which is absurd.)

This Aristotelian theory is actually very similar to Lewis’s. Just as in Lewis’s theory, modalities are analyzed in terms of quantifications over things that have the same kind of reality as the things we meet. The dog that will be born in ten years has the same kind of reality as the dogs that we meet today, just as the dog in another Lewisian world does. It is also true on Aristotle’s theory that the difference between mere possibility and actuality

is indexical. The actual is what is *now*, just as for Lewis the actual was what is *here*, i.e., in *our* world. There is, however, a difference here. Aristotle does not see, as far as I can tell, the full ontological parity between other times and the present that Lewis sees between other worlds and ours.

And there is a single objection that can be made both against Aristotle's theory and against Lewis's. Both theories share a crucial feature with the accounts of Parmenides, Spinoza, Leslie and Rescher. The *whole* of reality could not be different than it is. In the case of Lewis's theory, the whole of reality, i.e., the mereological sum of all universes, is fixed. There are no semantic resources in his theory for making it possible for this sum to be different. For were it possible for this sum to be different, it would *be* different at some world, whereas all worlds are parts of the *same* total reality. In the case of Aristotle's theory, contingency is only possible in propositions that change in truth-value. Hence, a proposition that reports in a timeless way the sum total of what happens over time would be, for Aristotle, necessary—though Aristotle is apparently unaware of such propositions.

But surely the sum total of reality could have been different. I am not making the controversial claim that it could have been radically different, i.e., that it could have lacked *all* the things it has ^[22], but only the more modest claim that it could have been different in some respect. If Lewis is right, there are infinitely many universes. ^[23] But it seems quite coherent to suppose that there was in fact only one. This coherence Lewis must reject as merely apparent. Similarly, it is quite coherent to imagine the possibility that in fact there never was any change, a possibility Aristotle must, and does (in *Metaphysics* L.6), reject as merely apparent. Both Lewis and Aristotle thus go against common sense modal notions. Lewis, however, can argue that the theoretical benefits of his theory are worth it. But if an alternate theory were found which had the same benefits and fewer paradoxical conclusions, then Lewis will be the first to prefer it (cf. Lewis, 1986a, p. 5).

4.3 The linguistic view

Possible worlds have much theoretical value. Thus, it would be nice to have them without paying the price of Lewis's extravagant ontology. One suggestion that has been made in many forms (see, e.g., Roper, 1982 and Jeffrey, 1983) is that possible worlds can be taken to be maximal sets of compossible sentences. A proposition holds at a world if it is entailed by the propositions expressed by the sentences that the world consists of. The actual world is the world all the sentences of which are true. (This does not mean that all the sentences *uttered* "in that world" are true. We need to distinguish between two senses of a sentence *s* being in a world: first, in virtue of *s*'s being a member of the set of sentences that are in that world, and, second, in virtue of there being a sentence that is "in" that world in the *first* sense which claims that *s* is uttered.) In doing this, we ontologically commit ourselves to sentences, which we already probably believe in, and sets, which have various theoretical benefits and which Lewis, too, needs in order to reap all the benefits of his possible worlds theory. The price is low.

Of course one needs to be more precise about what one means by sentences. It will not do to limit ourselves to actually uttered sentences. That would lead to the absurd conclusion that were there in fact no speakers, nothing would be possible. But if we speak of possible sentences, then our account of possibilities becomes circular, since we were supposed to be clarifying the ontological status of possible individuals. Fortunately, there is a simple solution to this dilemma. By "sentences" we mean *types* of sentences. Now, there is no great ontological extravagance in positing such types. Human-type languages can, to a good approximation, be reduced to sequences of discrete symbols, and types of sequences of symbols can easily be modeled set-theoretically. So this account in fact needs nothing more in the ontology beyond set theory. Since we may well want set theory for independent reasons, this is a very cheap price.

But, Lewis has argued, you get what you pay for, and we shall see in Section 2 of Part IV that we don't get enough. If we use an *actual* language, we have the problem of alien properties: basic properties for which our language has no words, but which are instantiated at other possible worlds. But if we use a non-actual language,

then we need to have some way of specifying what that language is, and that is impossible for us unless by specifying the language we create it, thereby contradicting the fact that it was supposed to be non-actual. Moreover, unlike Lewis's account, this account does nothing to *illuminate* the meaning of modal propositions, because it presupposes modality in the requirement that we talk of maximal *consistent* sets of sentences, whereas of course a set of sentences is consistent if and only if the conjunction of the sentences is possible. I shall also press against the linguistic view an arbitrariness objection. There are many actual and infinitely more non-actual languages. Which language is to be the privileged one? In Section 2.3 of Part IV, I shall argue that no solution to this problem is satisfactory.

4.4 The propositional primitive modality view

An approach that escapes the concern about the arbitrariness and limited expressiveness of linguistic representations is through abstract propositions. For various theoretical purposes, it is useful to introduce entities known as *abstract propositions* which are what our sentences express. Two sentences are synonymous if and only if they express the same proposition. Moreover, it is the proposition which is the carrier of truth, because truth does not belong to a sentence, which is language relative, but to the language-invariant proposition expressed by the sentence.

Now that we have introduced propositions by ostension as entities which sentences express, we can speak of the whole collection of propositions. Not all members of this collection are expressed by some sentence actually uttered. Nor even are all the members actually *expressible* by some sentence of a human language. This is so because propositions are something that is invariant not just between actual languages, but between actual and possible languages. So whatever can be expressed by any possible language is a proposition. And, conversely, any possible proposition p can arguably be expressed by some language, e.g., by the language that uses the sequence "All mimsy were the borogoves" to mean p . Note, however, that we escape the circularity objection because we are not actually *defining* propositions in terms of possible languages. We are defining them by ostension in terms of our language. But then we realize, on theoretical grounds, that propositions have a life of their own going beyond our actual language, rather as the electrons we posit on theoretical grounds to explain some actual phenomena have a life of their own and possess dispositional properties not exhausted by the actual circumstances of this world. Admittedly, the kind of explanatory role the two serve is different: electrons play a causal role while Platonic entities such as propositions do not. But nonetheless the propositions do explain various facts about sentences and propositional attitudes.

The above should have given us a grasp of the notion of a proposition. Propositions, moreover, enter into logical relations. This is so because logical relations between sentences are supposed to be invariant under paraphrase, and hence must be reducible to logical relations between propositions. We can therefore talk of propositions being consistent or not. Now we can define a possible world: it is simply a maximal consistent collection of propositions, assuming there are such maximal collections (I shall argue in Section 3 of Part IV that this assumption is in fact justified). Or, alternately, we can define a possible world as a class of logically equivalent maximally strong propositions, where a proposition is "maximally strong" if it entails every proposition compatible with it. These approaches have been championed most notably by Robert M. Adams and Alvin Plantinga.

Of course both of these definitions presuppose modality whether in the notion of consistency or in that of entailment, and so we will not get a reductive Lewis-type analysis of modality. But possible worlds may still be a useful construct to have, even if they do not give such an analysis. In Section 3 of Part IV, I shall argue that approaches to modality along these lines fail to answer the Parmenidean objection to modality. Presumably, on this account, the truthmaker of the proposition that it is possible that there are unicorns is the having of some property by the proposition that there are unicorns. But how does the having of some abstract property by some abstract proposition relate to the possibility of there being unicorns? I shall argue that this is an insoluble problem if one

limits oneself to the resources of this theory.

Moreover, there is a mystery as to the ontological status of propositions. What are ideas that no one is thinking of? Are they substances? What sorts of substances are they? Someone who is not enamored of Platonism will shrink from propositions.

But on the positive side, propositional and linguistic approaches both avoid the paradoxes that plague Lewis's theory.

4.5 Aristotle again and branching

While one of Aristotle's notions of modality was seen to be unsatisfactory, there is also another implicit in his work to choose from. Parmenides was worried that change involved something's coming to be out of nothing. For when *A* comes to exist, then earlier *A* did not exist. To answer this concern, Aristotle developed his tripartite account of change. There is a substance, a form and a privation. In the case of generation, the substance goes from having a privation of a form to having that form. Thus, a man may go from having a privation of beardedness to having a beard. But the beard does not come from nowhere. Rather, the man at the beginning of the process of change was *potentially* bearded, though *actually* clean-shaven. The privation that he had was a potentiality for beardedness.

On this account, there is something in the substance which can be identified as a potentiality for the alternate states of the substance. If we further accept Aristotle's general thesis that potentiality is grounded in actuality, we have to say that there is something *actual* in the substance in virtue of which that substance can change. But this account not only helps to solve the Parmenidean puzzle about change—it may also help with the Parmenidean puzzle about modality. Even if I never grow a beard, it is true to say it is possible for me to grow a beard because there is in me and in the environment around me something in virtue of which the growing of a beard is possible, i.e., a power (of course further scientifically analyzable) in the hair-follicles on my chin to produce hairs together with the capability, not restricted by the environment, for refraining from shaving. The truthmaker, on an Aristotelian account, of the proposition that it is possible for me to have a beard is to be found in these worldly actual powers and capabilities.

As an account of modality in general, this is insufficient. For one, at first sight it only applies to local *de re* modalities. This approach will not give us possible worlds in any obvious way. Moreover, the account is not reductive, since it accounts for modality in terms of ability, and ability is a modal term. However, while ultimately not reductive, the account is illuminating. For in ordinary language, the notion of ability is arguably more basic (cf. Place, 1997), and from it general notions of possibility are obtained by extension. We have personal knowledge of ability, e.g., in the way Kant outlines in the second *Critique* through recognizing ourselves as morally responsible for an evil act and thus as having been capable of doing otherwise. There is also less mystery about capability than there is about modality in general since capabilities are actual properties of actually existing things, and so the account is indeed helpful. And at the very least, if this approach worked, it would reduce modal talk in general to a particular subset of it.

Aristotle's account when generalized in a moral global way may lead to branching theories of modality (see, e.g., Mackie, 1998). When a substance has more than one alternative before it, these alternatives can be thought of as presenting a world-branch—though unless we want to make Lewis's move of making all worlds concretely existent, we should not think of there concretely existing worlds corresponding to all branches. If we look at all modality as induced by such branchings, so that we see a proposition as possible if it is true somewhere on the full tree and necessary if true everywhere, then we will in fact be generalizing Aristotle's account of the change in a single substance.

There is, however, still a problem: in fact, the same problem as was the decisive consideration against Aristotle's temporalized approach. It is, it seems, possible for all events in time to have been different—while there might be necessary entities such as mathematical, there surely are no necessary temporal events. This intuition

will be a metatheoretical constraint on theories of possibility, and one that it seems difficult to accommodate in a branching theory—for where should the branching point corresponding to this possibility be put?

4.6 Leibniz

Leibniz gave life to the notion of possible worlds. On his view, God necessarily exists, and possible worlds are maximal self-consistent ideas or concepts in God's mind. One could also talk of these worlds as maximal self-consistent thoughts entertained by the divine mind, and this would for all practical purposes be equivalent since any such thought corresponds to the concept of a truthmaker of that thought, whereas any maximal self-consistent concept corresponds to the thought of this concept's being instantiated.

Leibniz had in fact given an argument for the existence of God from the existence of necessary truths, and hence from the existence of modal truths (since assertions of necessity and possibility are necessary truths by S5, which Leibniz accepts). Necessary truths, Leibniz argues, must be grounded in some reality, and the only reality Leibniz can see as capable of this is a necessarily existent mind. Of course, the argument leaves much for discussion. Why can't the necessary truths be grounded in the thoughts of different minds in different worlds? Why can't they be self-subsistent in a Platonic way?

Positing divine ideas as possible worlds gives one the benefits of the linguistic and propositional theories of possible worlds. Like the linguistic theory, this approach allows something that we all have some ordinary pre-theoretical understanding of, the ideas of a mind, to constitute the collection of possible worlds. Admittedly, mystery is introduced by the fact that these are the ideas of a *divine* mind and by the maximality involved in them. But at least we do not have the dark mystery of the Platonic propositions, whose ontological status is almost completely opaque—for what, indeed, does it mean to be “abstract”? And like the propositional theory, the possible worlds have a representational power that a merely human-like language will not.

Of course, the theory does carry the ontological commitments of theism. But it is not revisionary of our ethical and epistemological notions in the way that Lewis's theory will be seen to be. Moreover, there is independent evidence for these commitments, namely all the evidence for theism. [\[24\]](#)

Unfortunately, Leibniz's account fails to answer the Parmenidean worry. Granted, the proposition that such and such a world is possible is an idea in the mind of a necessarily existent God. But what makes this idea *true* if the world in question is non-actual? Where is the idea's truthmaker? [\[25\]](#) And, in virtue of what is the idea that is identical with a possible world self-consistent?

In Part VI, I shall argue that the limitations of Leibniz's approach can be precisely supplied by the merits of the Aristotelian approach in the previous section. The resulting theory will be the most satisfactory account of the nature of possible worlds and the meaning of modal propositions.

But first we need to consider impossible worlds. Then, why possible worlds are useful at all. Then, we will consider Lewis's approach. Next, in turn, we will consider linguistic and propositional approaches. Next, I will reply to arguments of Rescher against the very existence of possible worlds. And finally I will argue for and sketch the new Aristotelian-Leibnizian theistic view. The exposition of this view will, necessarily, be an outline needing further elaboration.

Section 5 Impossibilia

But before turning to the detailed consideration of possible worlds, something needs to be said about impossible ones. It has been argued (see, e.g., Lycan, 1991), that philosophical and explanatory utility arguments made for possible worlds also apply to impossible worlds. However, as follows from the analysis in Part II, none of the applications of possible worlds that survive critical scrutiny are ones that would give one reason to posit impossible worlds. The application most cited in this connection is that of using possible worlds to analyze beliefs

and propositions, since after all people often believe impossible things. Indeed, I know that the conjunction of all my beliefs is logically inconsistent because I *believe* that the conjunction of all my beliefs is false, which belief is logically inconsistent with that conjunction!^[26] However, the application of possible worlds talk to beliefs and propositions will be seen to fail, and hence will give no reason to posit impossibilia.

The two surviving applications are the analysis of modality, for which possible worlds suffice, and the analysis of counterfactuals. One *might* try to make an argument for impossible worlds on the grounds that they allow for an analysis of *per impossibile* counterfactuals. However, *per impossibile* counterfactuals are so contextually bound that it is by no means clear that a unified account of them is possible or even desirable. One might think that a sufficient condition for a genuine subjunctive conditional to be true is that the antecedent entails the consequent, and *per impossibile* counterfactuals fail this criterion, since an impossible proposition entails all propositions.

But other than saying that impossible worlds are not needed, is there any argument for why there in fact are not any such worlds? David Lewis has argued that in his setting there is a simple such argument. For Lewis, the “at” in “*p* is true at *w*” only restricts the scope of quantifiers in *p* and ensures that expressions referring to “the actual world” get evaluated indexically with *w* in place of “the actual world”, and similarly for cognates. Consequently,

(9) (not-*p*) is true at *w*

entails:

(10) not-(*p* is true at *w*).

But it is plain, then, that if there is a world *w* at which *p* and not-*p* are both true, then it is also true that (*p* is true at *w*) and not-(*p* is true at *w*) which is a contradiction not relativized to any world and hence unacceptable by anybody’s account. One might think that this works only for *logical* impossibility, but since I have argued that there is no substantial distinction between logical and metaphysical necessity, by the same token there are no impossible worlds.

Of course, the Lewisian account only works because his worlds are concreta like the actual world. If our worlds are books in a heavenly library, then there is no contradiction at all in the fact that one book might contain a sentence expressing *p* and another sentence expressing not-*p*. So there is no logical contradiction in supposing impossible worlds.

There is, however, a serious disanalogy between possible worlds and impossible worlds. A good account of possible worlds will closely relate possible worlds to the ontological ground of alethic possibility, i.e., it will be closely related to an answer to the Parmenidean problem of what we are talking about when we make objectively true assertions about possibilities. But there is no parallel claim to be made for impossible worlds because arguably we are not talking or thinking *of* anything when we talk or think about impossibilities. If one thinks that a thought counts as an of-*X* conception if and only if were *X* to exist, the conception would be a conception of the *X*,^[27] then the notion of a conception of something contradictory is not viable.^[28]

This does not deny that one can *describe* impossible worlds, and one can produce a logic for handling the contradictions within in them in a way that prevents *every* description from being true of those impossible worlds. But not every description corresponds to a concept that has intentionality. I would submit that when we appear to be talking of impossible worlds, we are either talking about possible worlds in disguise (e.g., “A world with non-mammalian horses is an impossible world” can be paraphrased as “No possible world has non-mammalian horses”) or else we are talking about words, with it being completely up to us how we assign truth values to our assertions. There are presumably many possible ways of individuating impossible worlds, and many different logical systems for handling contradictions in a controlled way, each yielding different truth value assignments.

A parallel statement can be made for *possible* worlds as well, of course, so this argument is not yet complete. There evidently are many systems of modal logic and many ways of individuating possible worlds.

However, statements about possibility and necessity matter to our ordinary life and our ordinary concepts in a way in which I submit statements about impossible worlds, except when paraphraseable into statements about possible worlds, do not matter. For instance, it matters ethically speaking whether there is a possible world in which a child who was actually conceived through rape was not conceived through rape, since knowing the fact of the matter about this *should* affect, perhaps though only to a very small degree, the child's attitude towards her mother's rapist (see Section 4.2.1.b of Part III for more discussion of how modality matters). We would like to know what is the objectively true fact about many issues dealing with possible worlds.

I would conjecture, and arguing for this conjecture in detail would go beyond the scope of this thesis, that any true assertion about impossible worlds can either find a truthmaker within the totality of *possible* worlds or is an assertion whose terms do not have the requisite connection with aspects of our ordinary language where we have reason or need to suppose there is objectivity, and hence is an assertion whose truth value is to be assigned by convention. If this conjecture is correct, then in an ontological study of what underpins the truth of statements about possible worlds, we have no need to worry about impossible worlds. Though it *is* worth noting in passing that even if the conjecture is false, there is still some hope of making sense of impossible worlds in the Aristotelian-Leibnizian setting I will sketch in Part VI, though there is none at all in Lewis's.

Part II. Applications and pseudo-applications

Section 1 Modality

1.1 Box and diamond

The most obvious application of the theory of possible worlds is, of course, to furthering the understanding of modal claims. This can be either as a useful logical device to make it easier to grasp complex modal assertions and to express assertions that cannot be otherwise expressed, or one may more ambitiously see possible worlds theories as giving an *analysis* of all modal claims. Whether one can take the more ambitious approach or not depends on whether one's construction of possible worlds presupposes modality or not. If it does, then obviously possible worlds cannot provide an analysis of all of modality.

The only serious account of possible worlds that does not appear to presuppose modality and hence that supports the more ambitious use of possible worlds is that of David Lewis. Unfortunately, I shall argue in Part III that (a) it, too, must presuppose some modality (see Section 4.2.1.b of Part III), and (b) it leads to too many paradoxes for it to be at all acceptable. I am not claiming that having counterintuitive consequences is enough to refute a view; but the sheer number and weight of these in the case of Lewis's system *is* enough. Just as Lewis's case for his account is a cumulative one based on the multiplicity applications, my case against his account is a cumulative one based on the multiplicity of serious paradoxes.

The general way in which modal claims are expressed in terms of possible worlds is by quantifying over all worlds: for instance, $\Box p$ holds if and only if " w (p is true at w)", while $\Diamond p$ holds if and only if $\$w$ (p is true at w). But the expressive power of possible worlds goes beyond box and diamond operators as Lewis (1986a, Section 1.2) claims and Melia (1992) proves.

1.2 The global nature of modal claims

Moreover, as mentioned in Section 1 of Part I, the notion of a possible world is correlated with our intuition that even the box and diamond modalities have a global component. To tell whether some proposition is possible, one has to have some idea about whether it could be made to fit into a story of a whole world. Of course, if one has Humean intuitions that a world is made up out of parts such that any part is compatible with any other part, considerations of shape, space and time permitting, then this is not so important, since to tell whether a proposition is possible it is then only enough to examine the putative local state of affairs that would make the proposition true, and to decide whether this is possible.

However, there is good reason to reject a view of possibility that does not have the resources for discussing global possibilities of some sort. Many ordinary language modal claims are of an apparently local nature and for disambiguation require globalization. If I say "I might have been a physicist", I am apparently making a local claim: this proposition, that I might have been a physicist, is logically possible. But this claim is not the one I mean to express. I do not mean, for instance, that there might have been a world radically different from ours in which every person has innate knowledge of the laws of physics from conception. *That* world is not a part of the truthmaker of the claim I mean to express with "I might have been a physicist." Rather, I mean to say that much of this world might have been as it is, with me having roughly the mental capacities I do, and yet with me being a physicist. It is this claim that is non-trivial and interesting. But it is a claim that requires one to talk of worlds *as a whole* rather than piecemeal of the possibility or necessity of an isolated proposition. For one is asserting that there is a possible world that matches the actual world in such-and-such respects, but in which I am a physicist. And to disambiguate the claim, I will have to point out, perhaps contextually, what those respects of world-match are, and this is just to specify the set of possible worlds that I am quantifying over in the claim.

Nor will it do to avoid possible worlds by just taking a proposition that describes the way the possible

worlds I am talking about are supposed to be (contain me, have such-and-such laws of nature, have such-and-such a history prior to my conception, etc.), and say that all I am claiming is the compossibility of this proposition with the proposition that I am a physicist. For the possible worlds that I am quantifying over when claiming “I might have been a philosopher” cannot be described in a finitary way, for to do so would be to describe most of the history of the actual world, since all the history prior to my conception is arguably supposed to be fixed by the “might have been”. Of course, one might believe in a rich store of propositions, among which there is an infinitary proposition that describes all those features of the actual world that I wish to keep fixed. But if one has such a belief, then presumably likewise there will be an infinitary proposition that describes *all* the features of the actual world. And since the actual world should not be taken to be exceptionally fortunate vis-à-vis the expressive power of propositions and since propositions have necessary existence, likewise for any possible world there will be an infinitary proposition that describes all of it. But possible worlds under the name “consistent infinitary propositions describing all of a world” smell as sweet, or as ontologically heady, as those under the name “possible worlds”: so one hasn’t avoided possible worlds.

Moreover, there is good reason to believe essentialism is true (some of the reasons will be discussed in 4.2.1.b of Part III), or at least good reason to have a theory of possibility that at least has the expressive capability for making sense of essentialist claims—one does not want to rule them out from the outset. If Kripke is right, then, for instance, the claim that horses are possible involves a claim about a possibility of certain laws of nature, which is a global claim—a whole *universe* satisfying such-and-such laws is claimed to be possible. Or, even without essentialism, we can find a theological case. It is surely possible that there be a God, where “God” is defined as an abbreviation for the definite descriptor “the unique perfectly benevolent, all powerful and all good creator of all other concrete beings.” But then the proposition that it is possible that God and some evil exists could well be a global claim about a possible world: it is a claim that there is some world which *on the whole* has properties that justify God in allowing this evil.

Both in the essentialist case and in the theological case, it is very natural, then, to consider modal claims as bound up with quantifications over whole worlds (or large parts of them, but that would be no gain, since a world itself is, trivially, a part of itself).

1.3 Supervenience

As another example, we can define the notion of supervenience using possible worlds: *A*-type states of affairs supervene on *B*-type states of affairs (the *locus classicus* being the claim that goodness supervenes on natural facts—see Hare [1964, p. 80ff]) if and only if any two worlds which are indistinguishable in respect of *B* are indistinguishable in respect of *A*. David Lewis (1986a, Section 1.2) has argued that in fact such claims cannot be expressed with ordinary box and diamond operators. If so, then possible worlds are indeed a useful tool.

Of course one could also do the same thing with quantifications over “aspects” and occurrent states of affairs:

- (11) $\delta("a"b((a \text{ is an } A\text{-type state of affairs and } a \text{ obtains and } b \text{ is the } B\text{-aspect of the actual world}) \dot{E} \delta(a \text{ does not obtain } \dot{E} b \text{ does not obtain})))$.

However, if we take (11) to be an analysis of the claim that *A*-type states of affairs supervene on *B*-type states of affairs, we have not gained anything over a possible worlds analysis. For we have admitted to our ontology complete aspects of worlds, and after all there is a one-to-one correspondence between a world and the collection of all of its aspects.

1.4 Transworld comparison

One might wish to define the notion of, say, *x*’s being an entity than which no greater is possible or of a picture being such that no picture can be uglier or the like. I shall confine myself to the more hallowed Anselmian

case. As Lewis (1970) has demonstrated, the notion of maximal greatness is *prima facie* ambiguous. At the least, one could reasonably understand it as claiming one of the following:

(12) " w " y ((y exists in w and x exists in w) $\dot{\equiv}$ (y is not greater in w than x is in w))

(13) " w " y ((y exists in w) $\dot{\equiv}$ (x exists in w and y is not greater in w than x is in w))

(14) " w " y ((y exists in w) $\dot{\equiv}$ (y is not greater in w than x is in the actual world)).

And it is in fact (14) that is the best interpretation in an Anselmian context. Of course if one allows oneself quantification over greatnesses, then one can do without possible worlds even in (14), just as if one allows quantification over aspects, one can do without possible worlds in analyzing supervenience. Thus, (12)–(14) are respectively logically equivalent to:

(15) $\delta(x$ exists $\dot{\equiv}$ " y (y is not greater than x is))

(16) δ (" y (x exists and y is not greater than x is))

(17) " g (g is the greatness of x $\dot{\equiv}$ δ (" y (y does not have greatness exceeding g))).

Note, however, that introducing quantifications over greatnesses or aspects is moving to a second order logic. Given possible worlds on the ground level, one can do all this in first order logic. Moreover, one may plausibly argue that (14) is not only easier to understand than (17) but is closer to what is *meant* by the assertion that nothing greater than x is conceivable. For, it seems more natural to say that we are comparing not greatnesses but individuals-in-respect-of-greatness.

Section 2 Counterfactuals and causality

Perhaps the bigger feather in the possible worlds theorist's cap is the Lewisian analysis of counterfactuals:

A counterfactual "If it were that A , then it would be that C " is (non-vacuously) true if and only if some (accessible) world where both A and C are true is more similar to our actual world, overall, than is any world

where A is true but C false. [\[29\]](#)

How one measures similarity of worlds may depend on the context, though Lewis does have a preferred measurement method. Thus, I, though not Lewis, will allow that in some context one might weight similarity in the past more strongly than similarity in the future: When I say "Were I to eat this piece of rotten bread, I would be sick", the worlds where the past differs from the past of my world are too relevantly dissimilar from the actual world to count, just as the worlds in which the laws of nature are different are too dissimilar. (Lewis thinks he can do without something like temporal weighting, but it shall be seen in Section 3 that this is not so.)

One might think that the amount of arbitrariness possible in defining similarity of worlds makes this definition unacceptable. On different accounts of similarity, different counterfactuals will come out true. But, surely, there is an objective context-free matter of fact whether some counterfactual is true or not.

However, there is no such context-free matter of fact in general. Take the joke: "Were Queen Victoria alive today, what would she be doing? Clawing at the inside of her coffin." Consider two worlds, both of which have laws of nature more or less like ours, except that in each a miracle occurs: in one, Queen Victoria today comes back to life, and in another she never died. If we weight similarity in the past heavily, then the world in which Queen Victoria tomorrow comes back to life is closer to ours. And in some contexts we *do* need to weight the past more heavily: specifically, when we are making a rational decision between actions and considering counterfactuals of the form "Were I to f , A would result". On the other hand, if we weight similarity more in terms of closeness of laws of nature, then it is arguably a lesser departure from the actually holding laws of nature to suppose Queen Victoria had never died than to suppose her coming back from the dead after her body has been rotting for many years.

The only reason I am aware of for thinking that counterfactuals should be context-free would be a conditional principle of bivalence that claims that for any p that is possible and any q , either *were p true, then q would be true* or *were p true, then q would be false*. But this conditional principle of bivalence is false. It is neither true that were the moon made out of cheese then it would be made out of *blue* cheese nor that were it made

of cheese then it would not be made of *blue* cheese. For further discussion of conditional bivalence, see Section 3.1 of Part IV.

The counterfactual account provided by Lewis is thus not invalidated by the contextuality of the measure of similarity. However, Lewis's account of causality fares worse as we shall see in Section 3. For if we think that causality is an objectively existing relation in the world with explanatory oomph, then we will be much less inclined to accept a largely context-dependent analysis of causality. (Though of course it is open to say that there are distinct but related senses of the word "cause" in the way that, say, Aristotle talked of "four causes".)

Section 3 The direction of time

David Lewis in his 1979 paper "Counterfactual Dependence and Time's Arrow" (Lewis, 1979a) has argued that according to his possible worlds analysis of counterfactuals, "backtracking" counterfactuals of the form "If event A were to happen at t_A , then event B would happen at t_B where t_B precedes t_A ", are usually false if B does not *actually* happen at t_B . On the other hand, there are plenty of such counterfactuals true with t_B following t_A (such as: "Were I to drop the glass now, it would hit the ground at some point in the future"). This time-asymmetry, Lewis claims, follows from his possible worlds analysis of counterfactuals despite the fact that this analysis of counterfactuals is entirely time symmetric. The asymmetry is, however, a contingent fact about the arrangement of this universe. Lewis argues, further, that this asymmetry gives meaning to the common notion of the future as "open" and the past as "closed"—even if determinism both of the future by the past and of the past by the future were true, which for the purposes of the analysis he assumes and which assumption I shall accept (only) for the purposes of the argument.

Much of the argument of Lewis's paper is a reply to an objection against his analysis of counterfactuals. I shall argue that this reply succeeds in some interesting special cases but fails in others. There are many common events that do not exhibit the kind of asymmetry Lewis argues for—indeed, enough of them to ensure that Lewis's analysis of what it is that constitutes the difference between the openness of the future and the closedness of the past fails.

But most seriously, I shall argue that any asymmetry Lewis finds, if there is one, is actually grounded in the preselection in the kinds of events that tend to figure as antecedents of ordinary language counterfactuals. This preselection I shall suggest is based on the common-sensical notion that it is past events that are the causes of future ones, with it almost [\[30\]](#) never being the other way around. Hence, the asymmetry that Lewis finds through his analysis is parasitic on people's time-asymmetric intuitions. Therefore, Lewis's analysis fails to give independent objective grounding for the counterfactual arrow of time.

3.1 Lewis's argument

Recall Lewis's account of counterfactuals:

A counterfactual "If it were that A , then it would be that C " is (non-vacuously) true if and only if some (accessible) world where both A and C are true is more similar to our actual world, overall, than is any world [\[31\]](#) where A is true but C false.

It does not matter for the purposes of Lewis's account of time asymmetry whether worlds are concretely existing physical objects (as Lewis of course thinks they are), or whether they are to be understood in some "ersatz" way as maximal compossible sets of propositions or in the Leibnizian way as ideas in the mind of a God.

Lewis's notion of counterfactuals is, of course, practically useless without a measure of similarity of worlds. Moreover, it gives rise to the following objection stated by Kit Fine and also supported by a number of other people [\[32\]](#):

The counterfactual “If Nixon had pressed the button there would have been a nuclear holocaust” is true or can be imagined to be so. Now suppose that there never will be a nuclear holocaust. Then that counterfactual is, on Lewis’s analysis, very likely false. For given any world in which the antecedent and consequent are both true it will be easy to imagine a closer world in which the antecedent is true and the consequent false. For we need only imagine a change that prevents the holocaust but that does not require such a great divergence from reality. [\[33\]](#)

To get out of Fine’s objection, Lewis proposes a measure of similarity of worlds that has four factors ranked as follows:

- (1) It is of the first importance to avoid big, widespread, diverse violations of [physical] law.
- (2) It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
- (3) It is of the third importance to avoid even small, localized, simple violations of law.
- (4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly. [\[34\]](#)

These factors are rigged to make sure that Lewis gets the right answer to Fine’s objection. One might of course have serious objections to these four factors and/or to their mutual ordering. [\[35\]](#) But even if they are implausible, it would be very impressive if Lewis could derive a time-asymmetry from them and from his definition of counterfactuals, since (1)–(4) are clearly time-reversal symmetric as is the definition of the counterfactuals.

Lewis’s argument against Fine is then as follows. We need to evaluate Fine’s counterfactual that if Nixon had pressed the button, the world would have been blown-up. Lewis exhibits four different kinds of possible worlds where Nixon pressed the button at t :

- (i) In worlds of the kind of w_1 (and I shall sometimes for short talk of just the world w_1 rather than of kinds) everything happens as in the actual world w_0 until shortly before time t , but then the worlds begin to diverge.

The deterministic laws of w_0 are violated at w_1 in some simple, localized, inconspicuous way. A tiny miracle takes place. Perhaps a few extra neurons fire in some corner of Nixon’s brain. [\[36\]](#)

And so Nixon presses the button and the nuclear holocaust follows. No further divergences from law happen, but the miracle is necessary given the assumption of determinism if the pasts of w_0 and w_1 are to coincide.

- (ii) In worlds of the kind of w_2 , on the other hand, physical laws are never violated, and Nixon presses the button. However, both the past and the future are different, because of the assumption of bi-directional determinism. At no time is w_2 the same as w_1 .
- (iii) Then, in worlds of the kind of w_3 , two small miracles, i.e. violations of the natural laws of the actual world, happen. First the same kind of miracle as in w_1 happens. But then a second miracle prevents the nuclear holocaust from stopping. However, the world is already very different. Nixon will write different memoirs, the wire has heated up, the movement of the finger has changed the gravitational gradient in China, etc. Indeed, it is plausible that nowhere in the whole of the future light cone with apex at the first miracle will the universe be exactly the same.
- (iv) And, finally, in w_4 -type worlds two miracles also happen. The first of these is the same as in w_3 , but the second is a lot more impressive than it was in w_3 . Not only does the nuclear holocaust not happen, but all the traces of the button pressing are removed, and so after the second miracle, w_4 looks just like w_0 .

Which of these four kinds of worlds is closest to ours? To have an answer to Fine, Lewis must argue that it is w_1 , since it is only there that the nuclear holocaust happens. Now, w_1 is definitely closer to our world than w_3 by Lewis’s criteria, because the only advantage of w_3 is that it lacks a nuclear holocaust in its future and hence there is more *approximate* future agreement between w_3 and w_0 than there is between w_1 and w_0 . And indeed this

agreement is *merely* approximate in the future light cone with apex at the event that caused the pressing of the button. However, w_3 has an extra small miracle, i.e. deviation from physical law. Avoiding small miracles is Lewis's third most important similarity factor. Therefore, w_3 is better in terms of the fourth most important factor, and w_1 in terms of the third important factor, and so w_1 is to be preferred to w_3 as a candidate for a closest world.

What of w_4 ? It is true that w_4 matches w_0 in a very large spatio-temporal region: all of the future of the second miracle and all of the past of the first. This is Lewis's second factor. However, w_4 must have a rather large miracle. The gravitational gradient has to be corrected throughout a large region of space. The particles shifted around (admittedly by a tiny distance) in China by the change in gravitational gradient caused by Nixon's hand-movement towards the button have to be shifted back. Nixon's apparent memories have to be altered. The wire has to be cooled. The vibrations from the click of the button have to be stopped from propagating. This would violate the *first* criterion for closeness of worlds in a way that w_1 does not. Hence, w_1 is closer than w_4 .

On the other hand w_3 is just about nowhere in space-time identical with our world. The lack of exact match anywhere in w_3 means that w_3 violates the *second* criterion of closeness, whereas w_1 only violated the *third* by having a small miracle. Hence, indeed, w_1 is the closest of the worlds (or, more precisely, types of worlds) in which the button is pressed—at least if no other worlds are candidates which for the nonce I shall grant Lewis, though in Section 3.4, below, I shall argue that there is an important candidate that Lewis has passed over. And since the nuclear annihilation of humankind does happen in w_1 , it follows that Fine's counterfactual "If Nixon had pressed the button, there would have been a nuclear holocaust" is indeed true on Lewis's account.

Moreover, the analysis does display a past-future asymmetry. For, given that the closest world is w_1 , it follows that counterfactuals of the form "If Nixon had pressed the button then C would happen at t_C ", where in the *actual* world C does not happen at t_C , can only be true if t_C is *after* to the time of the "small miracle" in w_1 , which time is slightly before pressing the button. So, as Lewis admits [\[37\]](#), there may be a modest amount of backtracking in the counterfactual—but only back to the time of the miracle.

What grounds the above analysis is the fact that an event like the pressing of a button has a lot of disparate effects—but a fairly localized cause. It is this temporal disanalogy that, on Lewis's account, grounds the counterfactual arrow of time that gives meaning to our intuitions about the openness of the future and the closedness of the past. In Sections 3.3 and 3.4, however, we shall see that this analysis is hopelessly flawed because of the failure to consider a fifth class of worlds. But first consider a different counterexample.

3.2 The pulled plug

- (C) Suppose that in our world, in fact Nixon has pressed the button. However, Captain Smith working in a computer room saw that the doomsday device was being activated, and he pulled out the plug of the computer that controls the doomsday device, thereby saving the world from certain destruction. Moreover, whereas Nixon's decision to press the button was a highly conflicted one such that there was a single neuron which happened to fire, and had it not fired the decision would not have been made, Captain Smith was a man of high moral caliber who had many times thought to himself what he would do if the doomsday device was activated, and had gone over mental scenarios of pulling the plug, so that his decision to pull the plug was highly overdetermined. Captain Smith had many reasons to pull the plug, and many neurons fired in his brain, any one of which would have been sufficient to make him pull the plug. Moreover, many different muscles fired up simultaneously, any one of which would have been enough to pull the plug (he pulled with both hands while kicking at it with a leg), and since he had a time window of several minutes, if that move had not succeeded, he could be expected to have tried again and again.

Now, consider the obviously true counterfactual: "Were Captain Smith not to have pulled the plug, the world would have been blown up." But on Lewis's account, to affirm the counterfactual on the grounds on which Lewis affirms Fine's counterfactual would require that the closest world to the actual one is a world where a miracle happens in Captain Smith's brain, which miracle prevents him from pulling the plug. But given the structure of Captain Smith's brain, the miracle would actually be a pretty big one. Not only would it have to prevent him from pulling the plug at

the time he did, but also at all other times in the time window. Such a thing would require a fairly large scale reorganization of Captain Smith's brain. Call a world where this happens w_1' and denote the world that (C) describes as "our world" by w_0' .

But now consider the alternate world w_5 . This is the world in which a small miracle happens in Nixon's brain. This small miracle prevents from firing that one neuron whose firing was necessary for the pressing of the button. In w_5' , the urge to press the button might come over Nixon, but quickly passes. It was never a very *decisive* urge anyway. And not surprisingly, in w_5' , Captain Smith doesn't pull the plug—because he has no reason to do so!

But now it can be argued that w_5' is closer to w_0' than w_1' is. For, w_1' involves a much greater miracle than w_5' does, namely a complete reorganization of Captain Smith's brain. It is true that w_1' matches w_0' exactly spatio-temporally for a while longer than w_5' does. However, this is only for a while—more precisely, for the amount of time between Nixon's pressing of the button and the time when Captain Smith's brain gets reorganized in w_1' . And Lewis, we know, is willing to allow worlds to mismatch for a short while in order to ensure that the miracle that happens is a much smaller one. In the original case of Fine, after all, the miracle might have happened a fair amount of time prior to the pressing of the button. For, maybe, Nixon had in the actual world never even gone *near* the button, so in order to avoid the large miracle of transporting him bodily to the button, a modification of his brain when he stood a distance away from the button, in order to get him to walk towards it would have to happen in w_1 , so that w_1 and w_0 would have been mismatched for a significant amount of time prior to the pressing of the button. But if Captain Smith's reflexes are quick enough, then the plug might have been pulled quite soon after Nixon's deciding to press the button and further one can imagine that Nixon was hovering indecisively over the button, so the miracle needed in his brain could have been *very* soon before the button's pressing. So the amount of time of mismatch between worlds brought about by having the miracle happen in Nixon's rather than Captain Smith's brain could be small. And given how much smaller the miracle would be in Nixon's brain than the miracle that would be needed in Captain Smith's brain, w_5' will be closer to w_0' than w_1' will be.

But if so, then Lewis's counterfactual analysis, assuming (following Lewis's discussion of Fine's case) that nothing closer than w_1' and w_5' can be found, implies that the counterfactual "Were Captain Smith not to have pulled the plug, the world would have been blown up" is false, since its consequent is false in w_5' . Thus Lewis's analysis fails here.

However, note that there is still a temporal asymmetry. For, although the counterfactual backtracks to Nixon's decision, which is surely in this case *too* far, it does not backtrack *very* far back, whereas it forwardtracks arbitrarily far: in the future there would always be different consequences of Nixon's not having pressed the button (e.g., history would look at Nixon very differently) than of Nixon's having pressed it but Captain Smith's having saved the world.

It might be objected that "Captain Smith not pulling the plug" is an illicit event description since it is the *complement* of an event. However, Lewis intends his account of counterfactuals to give a counterfactual relation between *propositions*, not events, and hence the objection is one that he cannot make. Moreover, Lewis must allow propositions reporting the complement of an event as antecedents of counterfactuals, because after all he proposes to analyze the sentence "event *A* caused event *B*" as "*A* and *B* occurred, and *were A not to have occurred, then B would not have occurred*", where the italicized phrase is precisely a counterfactual with a negative antecedent. Thus, the objection under consideration is one that Lewis cannot possibly make.

3.3 The button on the laser

Consider another simple example. For a long time, an activated laser has been sitting pointed into the sky. In the actual world, at time t_0 the moon happens to be where the laser is pointed, and a short while later, at t_1 , a place on the moon is illuminated with the light emitted by the laser at t_0 . The laser is equipped with a switch which has the property that when it is quickly depressed and released, the laser beam gets turned off.

Our counterfactual is: “Were the button to quickly move to the depressed state and then to the released state at t_0 , the spot on the moon would not be illuminated at t_1 .” I shall abbreviate “to quickly move to the depressed state and then to the released state” as “to be pressed”, but with it being understood that no inference from “the button is pressed” to “someone pressed the button” is thereby warranted. It may well be that the button gets pressed without a person doing it (e.g., by a sudden rise in air pressure over it). Indeed, I shall suppose that persons are far away from the laser at t_0 so any pressing in a counterfactual world that does not involve a *big* miracle of bringing a person will be a non-personal pressing.

Let w_0 be the actual world, and let w_0^* be its time-reverse. Let t_0^* and t_1^* be the points in the reversed time of w_0^* corresponding to t_0 and t_1 respectively, so that t_1^* is strictly earlier than t_0^* . In w_0^* we have various light rays converging on our spot on the moon at t_1^* (corresponding to light rays that were scattered off that spot at t_1 in w_0), and then being focussed precisely into a coherent beam arriving at the laser at t_0^* . Likewise, heat radiation precisely converges on the laser at that time. And then after t_0^* the button on the laserbeam is pressed in and released. A possible mechanism by which this button moves could be that various influences such as the sound waves, mechanical vibrations and heat in w_0^* corresponding to what was given off in the pressing of the button in w_0 combine to reverse the motion. Such a story *can* be given because we’re working *ex hypothesi* with time-reversible laws of physics.

But now consider a world w_1^* which matches w_0^* precisely up to shortly before t_0^* . Then a miracle happens: of itself, the button depresses and releases at t_0^* (note that the time reversal of the button being pressed is the button being pressed, since “being pressed” was defined as consisting of a depression followed by a release). And after that the laws of nature continue their usual course. Thus, in w_1^* as in w_0 various light-rays converge on a spot on the moon at t_1^* , and then are focussed into a coherent beam pointed at the laser. This coherent beam arrives at the laser at t_0^* , which happens to be shortly after the button was miraculously pressed. [\[38\]](#) Likewise, heat radiation converges on the laser at that time, as do the various sound waves and mechanical vibrations mentioned in the description of w_0^* . What happens when all these things meet the laser at t_0^* is difficult to predict. Perhaps the light converging on the laser beam scatters off somewhere from the turned-off laser. But it does not really matter for our purposes what exactly happens.

Now, let w_1^{**} be the time reverse of world w_1^* . What exactly happens in w_1^{**} prior to t_0 we cannot really say. It will, however, be different from what happens in w_0 . But what matters is that in w_1^{**} the spot on the moon is still going to be illuminated at t_1 even though the button was depressed at t_0 . How will this spot be illuminated despite the laser being turned off at the relevant time? Presumably the mechanism will be roughly as follows. In w_1^* , we had a beam of light converging from the moon to the laser, and then scattering off the deactivated laser all around. Thus, in w_1^{**} , we will first have light from all around the laser converging on the laser at just the needed angles so as to be reflected off the deactivated laser into a coherent moon-pointed beam. And shortly after t_0 , w_1^{**} matches the actual world.

Thus, w_1^{**} perfectly matches the actual world throughout all of the future of t_1 but has a small miracle in it which presses the switch at t_0 . Prior to this miracle, there is a mismatch between w_0 and w_1^{**} .

Lewis’s analysis of Fine’s counterfactual, as applied here, would produce a world w_1 which is like w_0 until shortly prior to t_0 , but then the button is miraculously depressed, and the laws of physics then again come into effect, ensuring that the spot on the moon is not illuminated at t_1 . In order to uphold the truth of the counterfactual that “were the switch pressed at t_0 , the spot on the moon would not be illuminated”, Lewis has to argue that w_1 is closer to w_0 than w_1^{**} is, since in w_1^{**} the counterfactual’s antecedent is verified while its consequent is not. But there is no reason to think w_1 is closer than w_1^{**} to w_0 . Both worlds contain a small miracle. World w_1 matches ours exactly in the past, but not in the future, whereas w_1^{**} matches ours exactly in the future, but not in the past. The only way Lewis could claim that w_1 is closer than w_1^{**} would be if he could argue that the past is longer than the future. But such an assumption is highly dubious as we shall see in the next section where we return to the

original case of Fine.

3.4 The general case

We have seen in the previous two sections that there are rather natural cases where the Lewisian analysis (a) does not yield a correct evaluation of the truth values of these counterfactuals, and (b) does not disallow backtracking in the way Lewis would like to do so.

A major worry is that perhaps Lewis's analysis will in fact fail even for the original case of Fine. After all, there, too, one can form a world w_1^{**} in which Nixon presses the button, and which exactly matches the actual world from a time shortly after the button press forever after. Shortly after the time t_0 of the button press, there is a small miracle in w_1^{**} which ensures that if we deterministically trace back the conditions from the time of the miracle to t_0 , the button is indeed pressed by Nixon at t_0 . The way one is to find such a little miracle is to imagine time running backwards, and then try to figure out how to change things slightly prior to the time corresponding in our reversed sequence to t_0 so that at t_0 Nixon's finger is releasing the button (a release of the button in the time-reversed world corresponds to a press of it without time-reversal). And then we can deterministically backtrack to define all the history of the world prior to t_0 . The major and inexplicable short-coming of Lewis's analysis is the failure to consider w_1^{**} .

In order for Lewis's analysis of "Were Nixon to press the button, the world would blow up" to work, the world w_1 , which, recall, matches our world until shortly before the button press, then has a little miracle causing the button press and then continues deterministically forever, must be closer to the actual world w_0 than w_1^{**} is. Recall, Lewis's four criteria, which together produce a lexicographic ordering, for closeness of a world to a ours world: (1) no large-scale deviations from laws of nature, (2) extent of spatio-temporal region of exact match, (3) size of small-scale deviations from laws of nature, and (4) extent of spatio-temporal region of approximate match.

If one were to dogmatically assume that the future is shorter than the past, then since w_1 matches w_0 exactly in the past and w_1^{**} matches w_0 exactly in the future, by Lewis's criterion (2), w_1 would trump w_1^{**} . However, the assumption is unjustified. It is false on those theories according to which time has always existed and will always exist. Likewise, it is false on those theories according to which time had started at some point, say the Big Bang, and will continue forever, and it is false on those theories according to which time started at the Big Bang and will continue to the Big Crunch, because on such theories we are presumably less than half-way to the Big Crunch from the Big Bang in light of the fact that the universe is *still* expanding. There are *some* eschatological religious views on which the assumption of a future shorter than the past is true, but not even all eschatological views claim this since many religious people believe that the believers will live forever *in time*.

Moreover, if the future is longer than the past, then Lewis has a major problem on his hands. For then w_1^{**} will automatically count as closer to w_0 than w_1 does, *unless* it could be argued that the miracle in w_1^{**} is a *large-scale* departure from the laws of nature, so that by criterion (1), w_1 will win. However, it is not clear that w_1^{**} need be a large-scale departure. If we imagine time running backwards, it would seem reasonable to imagine that with clever engineering a fairly small miracle shortly prior to the time-reversed version of t_0 will ensure that at t_0 Nixon is releasing the button (and hence shortly before this, he was made by this miracle to depress it). We may have a hard time imagining how to engineer such a miracle, but this may be because we're not used to engineering things in time-reversed universes—human engineering is always concerned about bringing about effects in our future. The miracle in question may be bigger than that in w_1 , but there is little reason to suppose it will have to be one of the large-scale miracles that criterion (1) of world-closeness is talking about it.

Lewis has three options at this point for trying to save his analysis. He can argue that in w_1^{**} there must indeed be a large-scale miracle. This would require a careful physical argument, and since we don't know about the engineering of backwards-running brains, it is a difficult task. Alternately, he can try to demote criterion (2), extent of perfect spatio-temporal match, so that it is less important than criterion (3), size of small miracles. But he cannot afford to do this, because then world w_2 where Nixon does press the button and where neither the past nor the future

matches our world but which world is close to ours in terms of (3) as there are *no* violations of laws of nature, will turn out to be closer to our world than w_1 . This will not do, since the temporal asymmetry Lewis is looking for will then disappear, since w_2 is different in the past, even the distant past, from the actual world, and so there will be plenty of backtracking counterfactuals that will hold if w_2 is the closest world in which Nixon presses the button.

The remaining option for Lewis is to insist that the future is not longer than the past. Without invoking apocalyptic scenarios, the only feasible way for Lewis to do this would be to insist that time is infinite in the past and infinite in the future. I shall assume that this is done. Lewis, however, is still not home-free. He cannot argue on the grounds of criterion (2) that w_1 is to be preferred to w_1^{**} as a match for the actual world. It is dubious whether he can argue it on the grounds of (4), since w_1 will differ quite radically from the actual world in the future—there is a nuclear holocaust in w_1 . What remain are (1) and (3), and these come to the same thing here: Lewis must argue that the miracle in w_1 is smaller than that in w_1^{**} .

There is some plausibility on his side. After all, we have a fairly good idea of what kind of an event could produce the miracle in w_1 , namely something happening in Nixon's mind—thus, a fairly local event—which determines him to press the button. However, we cannot say what kind of a miracle is needed in w_1^{**} . The reason for this is that the miracle in w_1^{**} must be one put in *after* the button press which has the property that if we backtrack deterministically from the time of the miracle to t_0 , Nixon ends up pressing the button at t_0 .

Our rational activity constantly requires of us that we calculate what kind of a present state will forwardtrack according to the laws of nature to determine a desired event in the future, and this is the kind of calculation that is involved in figuring out what the miracle in w_1 is to be. However, we rarely strive to figure out what kind of a present state backtracks according to the laws of nature to determine a desired event in the past, and so we are hard pressed to find a small future miracle which will determine the button press in its past. We must thus beware of concluding from the fact that we cannot figure out what miracle to put in w_1^{**} that that miracle must in fact be greater than that in w_1 .

But there actually *is* a reason to think that it is likely that the miracle in w_1^{**} is greater. To see this, note that we can divide up non-actual events into four non-disjoint classes. Class A consists of all non-actual events E which could have been produced by a modification of the actual world that keeps the past fixed up to shortly before the time of E and then inserts a relatively small miracle shortly prior to the time of E which miracle deterministically forwardtracks according to the laws of nature to yield E . Class B consists of all non-actual events E which could be produced by a modification of the actual world that keeps the future fixed from shortly after the time of E , and inserts a relatively small miracle shortly after the time of E which miracle deterministically backtracks according to the laws of nature to yield E . Class C is the intersection of classes A and B, and class D consists of all non-actual events outside of A and B. It is important here that the notion of “a relatively small miracle” be kept constant between the definitions of these classes.

Now, intuitively if E is a non-actual event chosen at random, it is highly probable that E lies in D. To see this, suppose for example that I am now at position x in the universe, and consider all counterfactual events of the form “Pruss being at y now”, where $y \neq x$. For most choices of y , a relatively large miracle would be required, whether in the future (with the past shortly before and back fixed) or in the past (with the future shortly after onwards fixed), to produce this. Most choices of y would place me very far away from where I am right now, indeed even outside this galaxy, and certainly large miracles would be required to quickly transport me there. Intuitively, there are a lot fewer non-actual events that are *close* to a state of the actual world than there are ones that depart wildly from the actual world, and it is the ones that are close to a state of the actual world that are much more likely to lie in classes A or B, so the vast majority of non-actual events will be in D. Of course, we cannot be considering these likelihoods merely in terms of the cardinalities of the classes, since after all A, B and D are all uncountably infinite. However, we know that cardinalities are not the right way to consider probabilities: a random point in the United States is unlikely to be in Pittsburgh, even though Pittsburgh contains the same number of points—continuum many—as the rest of the United States.

That an event falls in the union of classes A and B already says, thus, that the event is atypical. This is enough to damage Lewis's position *in general*. For, if E is a non-actual event neither in A nor in B, then any counterfactual world containing E and having either its past or its future match the actual world will have to contain a massive departure from the laws of nature, and so such a world will be by Lewis's criterion (1) further away from the actual world than a world containing E , satisfying the laws of nature *always*, but differing from ours throughout the past and the future. But if such a world is the closest to ours that contains E , then backtracking counterfactuals are licensed, and this Lewis wanted to avoid. Therefore, if Lewis's account is to have any hope, it must work only for a modest subclass of the class of all non-actual events. There is, however, still hope that it will work for events such as Nixon pressing the button.

To see where this hope lies, consider a random event from A that in fact significantly differs (by a measure of distance similar to that involved in saying that the miracles involved in producing A-events are "relatively small") from the state of the actual world (Nixon's pressing of the button is an example, because it would presumably involve his arm, a significantly large macroscopic object, being in a different place from where it in fact was). It is unlikely that this event is in class B as well. To illustrate this, consider an almost frictionless physically-isolated billiard ball system in our world, with billiard balls moving on it according to the laws of physics. Intuitively, most variant configurations of billiard balls would require a relatively large miracle to produce, *either* by a miracle in the past that forwardtracks to the desired configuration or by a miracle in the future that backtracks to that configuration. But now consider a random non-actual configuration E of the billiard balls and their velocities at time t_0 that happens to be in A, but that in fact is significantly different from the *actual* configuration at time t_0 . Configuration E thus can be produced by making a small miraculous modification to the actual world in the recent past of E , say at t_{-1} , and then forwardtracking deterministically to produce E at t_0 . Let B_0 be the actual world, and let B_1 be the world thus produced. Now, because the miracle at t_{-1} was *small*, it follows that *right* after the miracle the state of the counterfactual world is still fairly close to the state of the actual world. But because E significantly differs from the actual configuration at time t_0 , it follows that as we move deterministically from after the miracle to t_0 , we diverge more and more from the actual state of the world.

Now, our physical intuitions are that if two worlds are diverging under the influence of deterministic laws between t_{-1} and t_0 then after t_0 they are likely to continue diverging at least for a while, unless things have been rigged in some special way. Thus, it is likely that after t_0 , worlds B_0 and B_1 will continue to diverge, or at least are not likely to reconverge, unless the choice of the worlds has been rigged to ensure such convergence. Therefore, it is likely that at a time t_1 shortly after t_0 , the configuration E_1 of the billiard balls formed by forwardtracking from E will be significantly different from the actual configuration F_1 at that time. Therefore, E is not likely to be a member of B, since if it were, then it would occur in a world B_1^{**} whose future matches the future of our world after some t_1 which is shortly after t_0 , but which has a small miracle at t_1 that backtracks deterministically to yield E at t_0 . But then it would follow that the configuration of *that* counterfactual world at t_1 would be close to the configuration of the actual world at that time, since only a small miracle at t_1 would change one configuration to the other. But the configuration at t_1 , prior to the miracle, of the billiard balls in B_1^{**} will in fact have to be close to

E_1 , because E_1 is produced from E by forwardtracking deterministically. ^[39] However, we have already argued that E_1 is probably significantly different from the actual configuration at t_1 , and so it cannot be that the miracle in B_1^{**} is relatively small. Therefore, configurations that are significantly different from the actual ones and that are in A, are unlikely to be in B. The same statement will, note, hold with A and B reversed, however, since the laws governing billiard balls on an approximately frictionless isolated table are approximately reversible.

In general the intuition is that it takes *rigging* to make sure that an event significantly different from what is in the actual world should be in A, since presumably most events significantly different from actual ones cannot be produced by a miracle that is relatively small. But there is no reason to think that such rigging would also rig the event so that it should also be in B. Therefore, intuitively, C, the intersection of A and B, is significantly smaller

than A. And likewise, intuitively, C is significantly smaller than B.

Suppose that the above plausibility argument that C is significantly smaller than A is right. Then, Lewis's analysis of counterfactuals may work for Nixon's case. Nixon's pressing of the button does lie in A, even if we set the fixed measure of the size of miracles used in defining A to be quite small. By the plausibility argument, however, there is little reason to think that the pressing of the button is in C. Most members of A aren't. If this is right, then the miracle in w_1^{**} will have to be greater than that in w_1 , since the miracle in w_1 will not exceed the measure used in defining the classes A, B, C and D, but that in w_1^{**} will, if Nixon's pressing of the button is in A but not in C and hence also not in B.

If, however, the plausibility argument fails, then it is very unlikely that Lewis will be able to make any case for the miracle in w_1^{**} being greater than that in w_1 . He might be able to establish this on an *ad hoc* basis for some events, but the prospects for a general argument look grim. Henceforth, I shall assume that the plausibility argument does indeed work, and C is significantly smaller than A. By parity, we would expect C to be significantly smaller than B, since the same intuitions are in play there.

Consider, now, yet another modification of the original situation:

- (D) The button is located in a sealed container which it is beyond the technical power of us human beings to open or to manipulate the insides of. As a matter of fact, the button is never pressed. The counterfactual is now: "If the button had been depressed, then the world would have been blown up."

In (D), we have a harder time picking out what "small" miracle prior to the depressing of the button would let one forwardtrack to a depression of the button. Perhaps the air density over the button would miraculously increase and thereby depress the button. This is not so small a miracle, of course: a lot of air molecules would have to be moved. But most importantly, it is by no means obvious that a miracle located *after* the pressing of the button could not be equally small while *backtracking* to the depression of the button. In fact, it could even be a very similar sort of miracle: just imagine time running backwards starting with the actual future of the world, and add a miraculous increase in the density of the air over the button. This case is one where the pressing of the button actually falls in class C, and Lewis's analysis fails, because the miracles involved in w_1^{**} and w_1 could be of the same, or very similar, magnitude—and so w_1^{**} would be the better match for w_0 given that it matches throughout a future that is significantly longer than the past that w_1 matches w_0 in.

We are now in a position to see what is special about the Nixon case. In the Nixon case, there is an intuitively *canonical* miracle prior to event A. When asked to imagine the scenario of Nixon (contrary to fact) pressing the button, we imagine him thinking differently from the way he actually was, and the miraculous inducing of this act of him thinking is the small miracle we are after. We know roughly where and how in the past of A to locate the miracle. But when asked to imagine the scenario of the button in (D) becoming depressed, our intuition does not provide us with any such easy answer. Maybe the right miracle is an increase in air density. Maybe, if the button and the console are made of metal, a spontaneous magnetization of the button. There is no canonical location for the miracle in our world prior to the event.

Now, if there is such a canonical location for the miracle prior to the event, then the event in the counterfactual's antecedent will be in class A. By the plausibility argument given above, it is likely then that A will *not* be a member of B, and hence Lewis's analysis will work.

Moreover, counterfactuals of this kind are arguably much more common in our reasoning than counterfactuals of form (D). We often use such counterfactuals, and more generally such subjunctive conditionals, in our practical reasoning:

Were I to choose A, C would result; but were I to choose B, D would result; I prefer D to C, so I will choose A.

And in all counterfactuals that deal with practical reasoning, there is a clear canonical place to put the "small miracle"—in the mind of the agent. ^[40] The counterfactuals and subjunctive conditionals in fact already come with a story about where the divergence in worlds can easily start.

It is true that we do sometimes ask hypothetical questions of merely *theoretical* interest like: “What would have happened had that button become depressed in case (D)?” Of course, we ask such counterfactual questions more rarely than we do in the cases of practical reasoning. But I take it that even when we do ask such theoretically-oriented counterfactual questions, we have in mind some story as to how the antecedent of the counterfactual “could have come about”. In cases where the antecedent is a human action, the story typically involves the person’s having made a different act of will. But even in other cases, we tend to have a story in our minds as to how the antecedent “could have come about”. Otherwise, we would (a) be liable to think the counterfactual question to be ill-defined, and (b) have little reason to *ask* the question in the first place. For, to ask the question, we have to think of the antecedent in the conditional as having been a real possibility. And thinking of something as a real possibility involves having some story about how it “could come about”. Moreover, without such a story the relevant truth value of the counterfactual might be impossible to determine. In the case of (D), if our story about how the button is most easily pressed involves a massive earthquake that shakes up the box, it might not follow from the pressing that the world blows up—for the earthquake could also destroy the wire connecting the box to the doomsday machine.

If we ask counterfactual questions where the listener has no story in mind as to how the antecedent could have come about, we are liable to get puzzlement. “Were Queen Victoria to be alive today, what would she be doing?” The joke answer is: “Scratching at the inside of her coffin.” But actually, the meaning of the question is not clear unless it is specified what miracle story about Queen Victoria’s survival is the relevant one for consideration. E.g., is it that she became rejuvenated in the coffin, or is it that she just never died but lived an extra long time? The answer will be very different in both stories. And an intelligent questioner knows this. That is why he is unlikely to ask a counterfactual question without specifying some process by which the antecedent could come about, unless of course the context clearly singles out a process in virtue of that process being particularly simple—as in the case of the process of the button being pressed by Nixon through the firing of the neurons.

It follows that there are two kinds of intelligent counterfactual questions that might come up in everyday discourse. The first will specify in the antecedent what process for making the antecedent true is meant. The second will be one where there is an “obvious” simplest process, which simplest process is presumably the one that requires the least difference from this world. Actually, the two kinds of counterfactuals are not disjoint. The process in the first kind will almost always be specified *incompletely*, with the gaps being left to be filled in some “obvious” simplest possible way. And so most if not all cases of the first kind will in fact be cases of the second kind, with the partial process description being counted as a conjunct in the antecedent (e.g., “If *Queen Victoria never died and Queen Victoria was alive today*, what would she be doing?”). Moreover, both kinds of counterfactuals will have a common feature. The process, whether implicit or explicitly stated, will be one that is in the past of the event that is the primary concern of the antecedent. If from looking at the antecedent one cannot get an *obvious* process—necessarily in the past, since it is only at the past that our intuitions look for an efficient cause—that could have effected the antecedent, then the intelligent questioner, in every-day cases, is unlikely to ask the counterfactual.

Hence, those counterfactuals about which we are most likely to ask are such that there is an “obvious” process in the past of the antecedent (or perhaps partially overlapping with temporal parts of the antecedent—though even then, the process is likely to be set back in time) which process would bring about the antecedent *A*. The “obviousness” of the process means that the process cannot deviate *too* wildly from physical law. Hence, a Lewisian miracle bringing about this process would not be too big, and so w_1 will be close to w_0 . Thus, our event will lie in class A, and since the intersection of A and B is much smaller than A if the plausibility argument given above works, likely it will not be in B, and hence the Lewisian analysis will work for this event.

This shows that Lewis’s analysis probably works for most “everyday” counterfactuals in virtue of these counterfactuals’ antecedents, for practical reasons, having been *preselected* so as to lie in class A, the class of all non-actual events that can be produced by a relatively small miracle prior to the event.

Without such preselection, Lewis's account fails because it is by no means guaranteed that w_1 is a better match for w_0 than w_1^* . Given an event A without the described preselection, it seems *prima facie* just as likely that a small miracle in A 's future (or, more precisely, in the future of the point in time where one would like A to occur) would backtrack to A as it is that a small miracle in A 's past would forwardtrack to A . So there is no temporal asymmetry, or if there is one, then for events happening in our time it points the wrong way because the fact that the future of the universe is *longer* than the current past gives one a strong consideration in favor of w_1^* as a closer match to w_1 , as already discussed. We see this in case (D), and we can also see it in the case of event N discussed in Section 3.3.

We also see a partial lack of such preselection, though in a somewhat different way, in the example discussed in case (C). There, given how great a miracle would be required to override Captain Smith's resolute plans, the Lewisian will be at a bit of a loss as to where the miracle should be placed—in Captain Smith's brain or in Nixon's—and will end up having to countenance *some* counterintuitive backtracking to Nixon's own decision. Had the counterfactual in (C) been posed, however, in a more natural way as

(E) Had Captain Smith not pulled out the plug *after Nixon's pressing of the button*, then the world would have been blown up

there would have been no problem with a Lewisian analysis. There would then have been an obvious and unambiguous place where the miracle should be located—in Captain Smith's brain, prior to the time where in the actual world he pulled out the plug. And then the counterfactual would not exhibit the extended backtracking that case (C) gave. However, counterfactual (E) is preselected in the way described above—it is formulated in such a way as to force the process that would make the antecedent true to involve Captain Smith having a different mental state.

3.5 Another counterexample

It might be argued that examples (A) and (B) are rather contrived and do not represent most counterfactuals. For a more natural case, consider the following more natural example. It is highly likely that relevantly similar cases have in fact occurred.

(F) In the actual world, for several months prior to t_0 a one kilogram steel ball has been hanging from a strong plastic cable at a height of one meter above a highly elastic rubber pad. At t_0 , the cable is severed in a highly overdetermined way (e.g., several oxyacetylene torches are turned on it), and the ball falls; moreover, things are set up so that it would have been very difficult for the cable to be cut miraculously prior to t_0 (e.g., the only cutting/burning implements were not available prior to t_0), and the whole experiment was quite overdetermined. The pad is so elastic that the ball will rebound to a height of 0.9 meters. For simplicity of calculation, suppose all this is done in a vacuum.

Assuming an acceleration of gravity of 9.8 meters / second², the ball will take about 0.45 seconds to fall. Now, let t_1 be a certain specific time which I will specify more precisely later, but about which I will now only say that it is a time between t_0 and the ball's impact on the pad. Let A be the (temporally extended) event of the ball being suspended at a height of 0.8 meters at t_1 and of not having been above height 0.8 meters at any time in the two seconds preceding t_1 .

Observe that whatever the time t_1 between t_0 and impact is, event A is a rather difficult one to produce by a miracle in its past. One way to produce the event would have been to cut the cable prior to t_0 and then to arrest the ball at height 0.8 meters and suspend it miraculously there for two seconds. However, it was assumed that it was hard to produce a miraculous severing of the cable prior to t_0 . Alternately, one might backtrack several months, tie the ball at a lower height of 0.8 meters, and modify the whole set-up so the cable is cut at t_1 instead of at t_0 . However, that would involve a miracle quite long before t_0 and might require a large one if the whole set-up was overdetermined.

On the other hand, if t_1 is chosen appropriately, it is easy to find a small miracle in the future of A that backtracks to A . Let t_2 be the time of maximum impact of the ball on the pad in the actual world, that is the time at

which the pad is maximally compressed. Since the ball will rebound to 90% of its original height, the compressed pad at t_2 stores about 90% of the gravitational potential energy the steel ball (assuming the steel ball does not itself become compressed) had at its release. The rest of the energy went into such things as the vibrations of the laboratory floor and heat. Consider the time-reverse, v_0 , of the actual world w_0 . In v_0 , we have heat energy and vibrations of the laboratory floor which together with energy stored in a compressed pad conspire to propel a steel ball that at t_2^* is lying on the compressed pad up to a height of one meter at a later time t_0^* . Suppose that in v_0 we added a small localized miracle prior to t_2^* that decreased the energy stored in the compression of the rubber pad by 22%, or redirected this amount of energy away from the steel ball (e.g., by directing it into the ground). Call the modified world v_1 ; this world matches v_0 in its whole past up to shortly before t_2^* . Then, the heat energy and vibrations of the laboratory floor which contain 10% of the energy needed to lift the steel ball to height one meter are still going to transmit energy to the ball, but the pad will now only transmit $(0.78)(90\%) \gg 70\%$ of the energy the steel ball would need to be raised to that height. Hence, the ball in v_1 will be raised merely to a height of 70%+10% of one meter, namely to a height of 0.8 meters. It will achieve this height at some time t_1^* shortly prior to t_0^* , and it will then presumably start falling after t_1^* so that it will *never* be at a height other than 0.8 meters.

Now, let w_1^* be the time reverse of v_1 . In this world, at a time t_1 shortly after t_0 the ball will have zero velocity and be at height 0.8 meters (note that this construction defines what t_1 is). It will then fall on the rubber pad. Moreover, in the two (or even more!) seconds prior to t_1 , the ball will not have been above height 0.8 meters. Hence, A occurs in w_1^* . There is a small miracle after t_2 in w_1^* which backtracks to the occurrence of A , and after that miracle w_1^* matches the actual world exactly. The miracle in question is a small one: it is just the decrease of energy in a compressed rubber pad (note how the compression of the rubber pad makes the miracle even more localized). So, A can be produced by a small backtracking miracle but it does not appear likely it can be produced by a small forward-tracking miracle. Therefore, Lewis's framework for counterfactuals will produce many non-trivial true backtracking counterfactuals with antecedent A , since the past of w_1^* presumably is never exactly like the past of w_1 . This is not only absurd, but exhibits the opposite of the arrow of time Lewis tried to exhibit.

It might be objected that there were some idealizations in this account. For instance, I assume the experiment was done in a vacuum and that even if we decreased the energy of the rubber pad in the time-reverse world v_0 , still the energy in the vibrations of the floor and in the heat would go to propel the steel ball. These assumptions, however, are not essential to the example. Suppose that these idealizations are false. Nonetheless, miraculously decreasing the energy of the rubber pad in v_0 by 22% around t_2^* and then forward-tracking deterministically will have to (by energy conservation) produce a world where the steel ball will achieve a maximum height lower than one meter at a time t_1^* in the future of t_2^* . Let the height that will be achieved be h . Perhaps h is not exactly 0.8 meters. But whatever it is, we can then define the event A as being the event of the ball being suspended at a height h at t_1 and of not having been above height h meters at any time in the two seconds preceding t_1 , where t_1 is the time corresponding to t_1^* in the non-time-reversed world, and the analysis continues to go through. All that matters is that in the time-reversed world if we decrease the energy in the compressed pad, which is a small localized miracle, we will ensure the ball will achieve a lower height.

3.6 Conclusions

3.6.1 The problem with Lewis's approach

The most serious defect that Lewis's analysis suffers from is its complete neglect of consideration of worlds of the form w_1^{**} which match ours in respect of the future shortly after the event of Nixon's pressing of the button, but which include a miracle shortly after that pressing from which one can backtrack to the pressing itself. While this neglect *might* not affect his analysis of the case Fine challenged him with, it does affect the analysis of other cases.

Lewis (1979a, pp. 473–475) thought that the temporal asymmetry that he had found was based on the

contingent fact that overdetermination of the past by the future was more common than overdetermination of the future by the past. This he found by comparing world w_1 with world w_4 which included *two* miracles, one of which was designed to erase the future effects of the first, and noting that this second miracle would have to be comparatively very large. Price (1997, p. 151) has suggested that at the microscopic level the asymmetry may disappear. In fact, we can see that Lewis's intuitions are vitiated by the choice of worlds to look at. For the most relevant worlds to compare to the actual world in a Lewisian setting are not w_1 and w_4 , but w_1 and w_1^* , and so what one requires to get the temporal asymmetry is a condition to the effect that a lesser miracle is needed if we put the miracle in the past of the event described in the antecedent of the counterfactual, from which miracle we can deterministically forwardtrack to the antecedent's event, than if we put the miracle in the future of the antecedent's event and backtrack.

Lewis brings in here an overdetermination asymmetry according to which a single actual event is overdetermined by a number of future events but not by a number of past events.

Whatever goes on leaves widespread and varied traces at future times. Most of these traces are so minute or so dispersed or so complicated that no human detective could ever read them; but no matter, so long as they exist. It is plausible that very many simultaneous disjoint combinations of traces of any present fact are determinants thereof; there is no lawful way for the combination to have come about in the absence of the fact. (Lewis, 1986a, p. 50.)

As an example, Lewis alludes to the phenomenon of the spreading of spherical waves spreading from a point: "Countless tiny samples of the wave each determine what happens at the space-time point where the wave is emitted or absorbed" (*ibid.*) If Lewis was right, then a miracle that keeps the future constant would have to cut each of the nomic connections between a past event and its future overdetermining events, and that would involve a large-scale miracle. However, Lewis must be wrong, since we have already seen in case (B) that a single small miracle in the future of the button-press can backtrack to a button-press.

But one can do better than just saying "Lewis must be wrong". In fact, one can show that the overdetermination does not occur even in the paradigmatic case of the spreading of a spherical wave. Suppose that some event happens at t_0 in the actual world such that an amount E of energy is released from which a spherical wave starts spreading, and for simplicity suppose that in the region in question this is the only relevant source of energy. Lewis's claim is that the release of the energy is overdetermined by various disjoint samples of the spherical wave at some time t_1 after t_0 . But this is false due to energy considerations. For consider the allegedly overdetermining disjoint parts S_1, S_2, \dots, S_n of the spherical wave at t_1 . The S_i are events occurring in disjoint areas of space-time, and their energy comes from the originating event at t_0 . According to Lewis, none of these samples can be present without the originating event. Now, consider the system as a whole at t_1 . By conservation of energy, the total energy of the system is E . Each of the S_i carries some non-zero portion E_i of the energy released by the originating event.

If Lewis is right about his overdetermination claim, then it is nomically impossible that S_1 occur at t_1 without the originating event occurring at t_0 , and in particular without amount E of energy being released at t_0 . However, consider a world whose state at t_1 is just like the state of the actual world, except that S_n is replaced by an event T_n of strictly smaller energy. Then by energy conservation, what happens at t_1 cannot backtrack nomically to the release of amount E of energy at t_0 , since we have ensured that the amount of energy at t_1 is less than E .

We can be more explicit about the construction of the world in question. Take the time-reverse of the actual world, letting t_0^* and t_1^* be the analogues in the time-reversed world of times t_0 and t_1 , respectively, so that $t_1^* < t_0^*$. At t_1^* we have a number of events $S_1^*, S_2^*, \dots, S_n^*$ which are the time-reverses of S_1, S_2, \dots, S_n . At t_0^* an amount E of energy is absorbed. Now suppose that by a miracle occurring shortly before t_1^* we replace S_n^* by an event T_n^* of lower energy, and then we evolve the system until t_0^* . Since all the energy in the events S_i comes from the originating event at t_0 , likewise all the energy in the events S_i^* is absorbed in the event at t_0^* . If we replace S_n^* by a lower energy event, then the amount of energy available to be absorbed at t_0^* will be lower. Hence, if we replace S_n^* by T_n^* , then at t_0^* there will be an amount $E_1 < E$ of energy absorbed. Now take the

world, w^* , where this replacement happens, and reverse time once again to get a new world w . Then, in w , at t_0 there will be a release only of an amount E_1 of energy, and events S_1, S_2, \dots, S_{n-1} will occur but S_n will be replaced by a lower energy event T_n . Moreover, all the right nomic connections hold in the interval between times t_0 and t_1 since in w^* all the right nomic connections held between times t_1^* and t_0^* . Therefore, it is not the case that the occurrence of S_1 nomically necessitates the occurrence of a release of amount E of energy at t_0 . Indeed, not even the occurrence of all of the events S_1, S_2 through S_{n-1} necessitates that, simply because these events do not carry enough energy to necessitate this.

We have argued that in fact the asymmetry in Lewis's analysis of everyday counterfactuals, if and insofar as there is any asymmetry, comes from the fact that everyday counterfactuals are usually *preselected* for having antecedents which can be effected by easily imaginable processes acting in the antecedent's past. But this does not reveal an objective asymmetry that would be independent of people's intuitions that it is past processes that cause future events rather than future processes that cause past events. For those very intuitions are the psychological ground of the preselection of everyday counterfactuals, which preselection yields the asymmetry in the counterfactuals if we analyze the counterfactuals in a Lewisian way. If Lewis's argument could be taken to be an analysis of our everyday intuitions, this would not be so bad. However, it cannot be thought of as such an analysis, because it crucially depends on the fact that any localized event normally has many, often tiny, effects throughout a large spatio-temporal region (e.g., Nixon's button press, even if no nuclear holocaust happens, affects the gravitational field in China), whereas our everyday intuitions about counterfactuals do not depend on this. [\[41\]](#)

And once we depart from the realm of everyday counterfactuals, we can find a number of cases where Lewis's analysis breaks down. If the antecedent of the counterfactual is something like a neuron firing (the case in Section 3.3), or a button depressing in a locked box, the Lewisian analysis fails—or at least may well fail—to yield a past-future asymmetry grounding our view of the future as open and of the past as closed.

None of this contradicts the facts (a) that there is an asymmetry in our counterfactuals, even though Lewis's analysis has failed to give a proper objective grounding to it, (b) that this asymmetry may well be responsible for our view of the future as open and the past as closed, and (c) that there may be an asymmetry in causal overdetermination, even though Lewis's analysis in the end fails to connect this up with the asymmetry in (a). The asymmetry in counterfactuals remains unexplained—unless one could reduce it to the asymmetry in causation where past events are causes of future events but not the other way around. If one is satisfied with taking the asymmetry in causation as basic—as opposed to wanting to make sense of it *in terms of* the counterfactual asymmetry of time—then perhaps the fact of there being such preselection rules for antecedents of counterfactuals as ordinary usage puts in place *is* adequately grounded in the objective reality of world, since as we have seen it *is* grounded in our intuition of the causal asymmetry. But a person satisfied with this would be running a project that is the opposite of Lewis's here.

3.6.2 A fix

Fortunately, we can make Lewis's account yield the right values for counterfactuals by simply building the arrow of time explicitly into the definition of the counterfactual by demanding that any worlds invoked in the analysis should closely match the actual world in the past (cf. Davis, 1979), or the past light cone, perhaps, of the event described in the antecedent except perhaps for the very recent past. [\[42\]](#) But if we do this, then we no longer have a *derivation* of an arrow of time.

More precisely, one possible modification of the counterfactual is to build into the similarity relation between possible worlds a requirement that, in the context of evaluating a counterfactual whose antecedent is the report of an event at t , the worlds that one compares the actual world to should match shortly up to time t . Alternately, and perhaps preferably, one can say that two worlds that match from their beginnings until t_1 are *always* counted as closer together than two worlds that match from their beginnings until a time $t_0 < t_1$. This has the

[43]

advantage that we do not need to specify the time of the antecedent event.

There is, however, one possible pitfall with this approach. If the order of time is dependent on the order of causality, as in fact I will argue in Section 2.4 of Part VI, then we can no longer analyze causality in terms of counterfactuals and hence in terms of possible worlds. I take it, however, that counterfactuals and possible worlds are useful whether we can *analyze* causality in terms of them or not. It is worth noting at this point that possible worlds themselves will in the final analysis be elucidated in terms of causality (see Part VI, below). I do not do not at all mind explanation or causation being ontologically basic.

3.6.3 McCall's approach

Storrs McCall (1984) has suggested a different approach to counterfactuals, also in response to shortcomings in Lewis's approach and also building an asymmetry into the definition of a counterfactual. Moreover, McCall hopes to make the notion of the similarity of worlds simpler and more precise than Lewis does.

Say that a physically possible world w *branches* relative to the actual world at t_0 if it matches the actual world throughout space and time prior to t_0 and if there are times arbitrarily close to t_0 (obviously, they will have to be after t_0) at which w does not match the actual world. To evaluate a counterfactual, McCall suggests that we consider those physically possible worlds at which the antecedent holds which branch relative to the actual world and which have the property that no physically possible world in which the antecedent holds branches relative to the actual world any later than they do. If the consequent holds in these worlds, then the counterfactual is true.

Of course, there is a technical difficulty in the account caused by the fact that there may be a sequence of worlds in which the antecedent holds and which branch relative to the actual world at later and later times that asymptotically approach some time t_0 with the property that no world that branches relative to the actual world at t_0 or later satisfies the antecedent of the counterfactual. But this is easily handled: if there is such a sequence with the property that the consequent fails to hold at infinitely many of the worlds in an asymptotic sequence with this property, then the counterfactual is false; else, it is true.

More seriously, the sheer amount of indeterminism in the world destroys McCall's account entirely. McCall's example of how his account works is the counterfactual: "If Napoleon had won the battle of Waterloo, he would not have died on St. Helena" (p. 466). Admittedly, McCall says, there is a physically possible world where Napoleon wins at Waterloo and dies on St. Helena, because St. Helena had earlier become a popular vacation spot for French officers. But the branching of that world relative to our world is earlier than that of a world that branches during the battle. However, for all physical worlds w which branch during the battle and which have Napoleon dying off St. Helena, with the exception of those worlds in which Napoleon receives a fatal wound during the battle (and there is no reason to think that those worlds would be the ones where victory would be snatched by a later branching), we can find a branching relative to w that has Napoleon dying on St. Helena. After all, it is physically possible that Napoleon might on a whim go to St. Helena and have a fatal accident there, or that someone might on a whim persuasively suggest it to him as a vacation destination and then murder him there, or even that due to quantum randomness he would be randomly teleported to St. Helena and promptly killed there. Worlds like this need not branch any earlier from the actual world than the ones where Napoleon is victorious and never goes to St. Helena.

Lewis himself will not be bothered by such worlds because these worlds exhibit a *qualitative* dissimilarity from the actual world—e.g., Napoleon acquiring a different taste in vacation destinations than in fact he does, or a weird quantum phenomenon occurring whereas no such phenomena occur in the actual world. This shows a superiority in Lewis's similarity-based account, though as I've argued the similarity needs to be weighted temporally or in terms of causal antecedents.

For another counterexample to McCall, consider the following counterfactual: "Were Napoleon to have won at Waterloo, his victory would have come about out of something very weird." This is false, albeit vague. Were Napoleon to have won at Waterloo, presumably it would have been because his soldiers fought a little more

bravely than they did, or the English faltered a little more. This kind of a variation is not a particularly weird event. But any world where victory is snatched by such non-weird means will have to branch from the actual world at some time significantly prior to the end of the actual battle. However, there are physically possible worlds which branch *very* close to the end of the actual battle and where Napoleon wins a victory. These are worlds where the victory is won because of an extremely weird event. For instance, the quantum phenomena in the hearts of all the English ensure that all but one of them faint at almost the end of the battle. [44] Since these worlds branch later, they are the ones that on McCall's account need to be considered when evaluating the counterfactual. Presumably there are no physically possible not-very-weird victories that could be snatched seconds before the actual defeat, and so indeed it will be true on McCall's account that were there a victory of Napoleon, it would have come about in a very weird fashion.

Section 4 Propositions

4.1 Unstructured propositions

On Lewis's first take, a proposition is a set of possible worlds (Lewis, 1986a, Section 1.4). We say a proposition p is true at w if w is a member of p . This won't quite work, because as we shall see later (Part III.7.2), on no reasonable account of possible worlds is there a set of all possible worlds. However, it is open to us to say that a proposition is a class or collection of possible worlds, and for all intents and purposes this will be just as good.

Providing we know what collections are and have an account of possible worlds, we thus have an account of propositions. Unfortunately, as Lewis certainly realizes, this account does not distinguish between propositions that are logically equivalent. But the standard criticism of Lewis here, made forcefully by people like Plantinga, is that for many purposes such a distinction is necessary. For instance, if one thinks that propositions are both the bearers of truth and what sentences express, then one may be uncomfortable with saying that all the necessary truths are one and the same proposition which is identical with the collection W of all worlds. In particular, all mathematical truths are the same. Matters are even worse if we accept essentialist claims that genus-species relations are necessary: the necessary *a posteriori* proposition that horses are mammals will turn out to be the same as the proposition that spiders are invertebrates, and both will be identical with the *a priori* necessary truth that Fermat's Last Theorem is.

If one thinks that knowledge is of propositions, one then wonders what we have learned when we learned that Fermat's Last Theorem was true that went beyond the proposition we already knew that $1=1$. One might want to say that we just learned something about our language, namely that when English speakers use the words

$$(18) \quad \text{"There are no positive integers } a, b, c \text{ and } n \text{ such that } n > 2 \text{ and } a^n + b^n = c^n\text{"}$$

they express a necessarily true proposition, namely the proposition W . But this answer won't do on either of the two reasonable interpretations of the word "English." Either "English" is a rigid designator here of the language which I am now using or it is a definite sociological description of a language spoken by a group of people who are qualitatively described. In the first case, the proposition that when English speakers use the words in (18) they express a necessarily true proposition is itself a necessary truth, and hence on the above account of propositions saying that we have learned *this* proposition isn't saying anything more than that we have learned that $1=1$. For if "English" is a rigid designator of our language, then what (18) means in English is an essential property of (18). On the other hand, consider the case where "English" is a definite sociological description. Then there is a possible world w_1 where "English" does uniquely pick out a language but where (18) means that water is H_2O or, if we don't like Kripkeanism, that $1234 \cdot 4321 = 5332114$. Evidently then what we have learned in discovering Fermat's Last Theorem is different from what the English speakers in w_1 have learned upon discovering that the words in (18) express a necessary truth. But if what we learned was that (18) expresses a necessary truth in English, then we have

indeed learned nothing other than those people have. And this is absurd.

Nor are propositions that are logically equivalent the same proposition, as this Lewisian account would make it out. We can see this when we observe that explanation is a relation between propositions. Now let w be a forwards- and backwards-deterministic world. Let L be a proposition reporting the laws of nature of w . Let S_t be a proposition reporting the complete physical state of w at time t . Then, two-way determinism ensures that S_t and L jointly entail S_u for all times t and u , so that the conjunctions $(S_t$ and $L)$ and $(S_u$ and $L)$ are always logically equivalent. Now, the conjunction of S_0 and L evidently explains S_1 in a deductive nomological way. Hence, if propositions that are logically equivalent are to be identified, likewise $(S_1$ and $L)$ explains S_1 , since $(S_0$ and $L)$ and $(S_1$ and $L)$ are equivalent. But this is absurd, because the claim that “ $(S_1$ and $L)$ explains S_1 ” could only possibly make sense if S_1 were a self-explanatory proposition ^[45], which in this case it is not. We can also apply similar reasoning to conclude the absurdity that $(S_1$ and $L)$ explains S_0 since the logically equivalent proposition $(S_{-1}$ and $L)$ does.

4.2 Structured propositions

Realizing the need for an account of propositions that allows for differences between logically equivalent propositions, Lewis calls those propositions that he identified with collections of possible worlds “unstructured propositions”, and suggests that we also define “structured propositions” as set theoretic constructions out of the unstructured ones.

For instance we could associate the modifier ‘not’ with the unstructured relation [collection of all pairs of individuals in all possible worlds that fall under the relation] N that holds between any unstructured proposition and its negation, that being the set of all worlds where the original proposition does not hold. Then a negative structured proposition could take the form $\langle N, P \rangle$, where P is a (structured or unstructured) proposition. (Lewis, 1986a, p. 57)

This process can be continued with other connectives. If A is the relation that holds of a triple $\langle p, q, r \rangle$ of unstructured propositions if and only if r is the conjunction of p and q , then in the context of the previous section we can say that $\langle A, \langle S_0, L \rangle \rangle$ explains S_1 but $\langle A, \langle S_1, L \rangle \rangle$ does not, even though the unstructured propositions corresponding to $\langle A, \langle S_0, L \rangle \rangle$ and $\langle A, \langle S_1, L \rangle \rangle$ are identical. Explanation is a relation where structure matters. Similarly, this approach does let one say what it is that one learns when one learns that Fermat’s Last Theorem is true: one learns that a certain complicated structured proposition is true.

However, there are many set-theoretic constructions that will work equally well or equally badly for these purposes. Now, why should we call $\langle N, P \rangle$ “the proposition which is the negation of P ” instead of bestowing that title on the pair $\langle P, N \rangle$? Moreover, even if we choose an order for the ordered pair, there are multiple set-theoretic constructions for ordered pairs. For instance, we can use $\{ N, \{N, P\} \}$ to set-theoretically represent the ordered pair $\langle N, P \rangle$ or we can use $\{ \{ \emptyset, N \}, \{ \{ \emptyset \}, P \} \}$ or even $\{ N, \{N\}, P \}$. It is up to us—all do the job. So which constructions should we choose?

One might wonder: What is the fuss about since if all the constructions do the job equally well, then can we not just choose whichever one we want? Concerning a somewhat related issue with properties, Lewis writes:

It’s not as if we have fixed once and for all, in some perfectly definitive and unequivocal way, on the things we call ‘the properties’ ... (1986, p. 55).

But to dismiss the fuss over which construction is the right one is to forget what job propositions were supposed to do. Propositions are theoretical entities introduced as that which we mean by our language. They are supposed to provide criteria for synonymy within and between languages: two sentences (in the same or different languages) are synonymous if and only if they express the same propositions. It is essential that propositions be language independent—otherwise, for most intents and purposes we could just define propositions as sentences of some fixed language, say Latin. The multiplicity of set theoretic constructions that “do the job” mirrors the multiplicity of

languages, and hence Lewisian set-theoretically structured propositions are no great improvement over the situation we have when we just stick to languages, except for the advantages of formalization and the availability—for many semantic purposes unnecessary—of alien properties that our languages have no terms for.

There are two possible responses here. The first is that we should choose a particular set theoretic approach, and call the resulting constructions “the propositions”. These propositions will do the job that propositions are supposed to do. However, if this is done, then it becomes mysterious how it is that propositions are supposed to be what we *mean* by language. Suppose I affirm the negation of an unstructured proposition. Let us grant that I affirmed $\neg N, P\tilde{n}$. But how could we possibly find out that when I affirmed this, my meaning in fact used the one privileged set theoretic construction for ordered pairs rather than another? Indeed, it is not even plausible to suppose that it *did* use one construction rather than another. What kind of a queer fact would it be about our language that when we affirm negations our meanings are one kind of set theoretic construction of ordered pairs rather than another?

It is tempting to say that this is a misunderstanding of the role that the set theoretic constructions are supposed to play. Take a particular axiomatic rendition of the general theory of relativity. This theory *models* our space-time as some kind of a Riemannian manifold. But there are, of course, many set theoretic ways of expressing Riemannian manifolds, just as there are many set theoretic ways of expressing real numbers (one can express them as pairs of lower and upper Dedekind cuts, or just as lower Dedekind cuts, or just as upper Dedekind cuts, or as equivalence classes of Cauchy sequences, and so on). Which one of these constructions is “the right one”? Well, surely, the question does not matter—the same physical reality is modeled. And likewise, it might be suggested, it does not matter which set theoretic construction is used to model propositions. But this suggestion forgets that the construction of propositions was not supposed to give a *model* of propositions—it was supposed to give us the propositions themselves, the things we mean by our assertoric sentences. If Lewis were only giving a model of propositions, then that would mean that there would *be* propositions out there, of which the set theoretic constructions are mere models, and which propositions presumably would not *be* the constructions, just as, on some interpretations of relativity, there really *are* entities in the world that are points in space-time. But if this were so, then Lewis’s account of propositions would no longer provide an ontological reduction of propositions to set-theoretic constructions out of possible worlds. One could just as well consider the objectively existent propositions themselves, instead of his model. The model might be theoretically helpful, but it would not remove the ontological puzzlement that Lewis himself admits to feeling (Lewis, 1986a, Section 3.4) about what propositions are.

The other possible approach is to go in the opposite direction. A proposition now is not just *some* set-theoretic construction, but *all* of them in some sense. Suppose that C_1 and C_2 are systems of set-theoretic methods for modeling propositions, and p_1 and p_2 are variables that range over the set-theoretically constructed entities within these respective systems that model propositions. Then, there is an equivalence relation that holds between the pair C_1 and p_1 and the pair C_2 and p_2 if and only if p_1 and p_2 play the same role in their respective systems (e.g., if C_1 is the system where the negation of P is represented by $p_1 = \neg N, P\tilde{n}$ and C_2 the system where it is represented by $p_2 = \neg P, N\tilde{n}$). Perhaps, the suggestion goes, *the* propositions are equivalence classes of pairs under this relation. And perhaps there is some underlying intuitive and natural notion of a “relation” and “equivalence class of pairs” that does not require a specific set-theoretic construction (given that these are the only two notions we are dealing with, this is more plausible than the suggestion that there are always such notions for all the set-theoretic constructions that would be involved within a single system C for constructing structured propositions)—since otherwise we haven’t gotten rid of the arbitrariness in the definition.

But the difficulty here is in the equivalence relation. If we can help ourselves to “the relation” that holds between two system-entity pairs when the entities “play the same role”, then by the same token we could help ourselves to “the relation” that holds between two language-sentence pairs when the sentences are synonymous. Moving from languages to quasi-linguistic set-theoretic constructions does not gain one much here. There is little

hope that we could specify the relation in either case adverting to the notion of a proposition and saying, in the linguistic case, that we are talking of the relation that holds when the sentences express the same proposition, and, in the set-theoretic case, of the relation that holds when the two constructions model the same proposition.

Even more serious than the mere fact of arbitrariness in the construction is that different constructions give different truth values to semantic claims. On the account of conjunctive propositions that was introduced above, the sentences “The sky is blue and grass is green” and “The grass is green and the sky is blue” are not synonymous, because the former expresses the proposition $\langle A, \text{that the sky is blue, that grass is green} \rangle$ and the second expresses the proposition $\langle A, \text{that grass is green, that the sky is blue} \rangle$. But one might also give an alternate construction for the structured conjunction of p and q : one might define this conjunction as $\langle A, \{p, q\} \rangle$. On this construction, the two sentences will end up expressing numerically the same proposition. The choice between $\langle A, p, q \rangle$ and $\langle A, \{p, q\} \rangle$ cannot thus be arbitrary, because it has substantive philosophical consequences.

My own intuition is that the conjunction of two sentences expresses the same proposition regardless of the order of conjoining ^[46]; I also think the parallel point is clear in the case of properties (surely there is no difference between an electron’s spinning and being charged and an electron’s being charged and spinning). But whether this is so should not be legislated by constructional fiat. If as realists we think there actually is a fact of the matter as to whether the conjunction of two propositions is the same regardless of the order of the conjuncts, then this fact of the matter will be grounded in some reality independent of our construction of “structured propositions”. But if so, then Lewis’s theory of structured propositions, if it be correct, actually presupposes as primitive a fact about whether conjunction depends on the order of conjuncts, and hence is no longer a *reduction* of propositions to collections of worlds and to things built up out of worlds.

One might not want to be a realist about synonymity or identity of propositions, of course and hold that there is no fact of the matter that determines whether two sentences are synonymous or two propositions identical. For various theoretical purposes one might choose various accounts of “propositions”. This may even be the parallel of the point that Lewis is making above for properties. But it is significant to note that this is a point that Lewis in the large structure of *On the Plurality of Worlds* cannot make. Lewis is offering an argument for why we should believe that there are possible worlds by arguing that they have great philosophical utility, and offers his accounts of propositions and properties as examples. But in order for this argument to provide justification for *realism* about possible worlds it is surely necessary that the possible worlds should figure in an *explanation* of the philosophical phenomena in question. That possible worlds are useful for producing an account of propositions is only going to give one reason to be a realist about worlds if one is a realist about propositions or at least about the sort of talk that propositions are used to clarify. If the propositions are arbitrary constructions useful contextually, the worlds can be such, too. Admittedly, no one ersatz construction of worlds works equally well for all cases, but we can use different ones for different contexts, just as we have different constructions of propositions for different purposes.

There is another serious difficulty with the arbitrariness involved in the construction of structured propositions. The purpose of propositions is to be language-independent analogues of assertoric sentences, one per equivalence class of synonymous sentences. But Lewisian structured propositions are basically linguistic entities, albeit ones constructed in a language whose symbols are various sets, classes or collections, instead of marks on pages or vibrations in the air. Indeed, the fact that they are linguistic entities is seen from the fact that there are *different* languages that all provide constructions of propositions, e.g., a language that specifies that the grammatically correct form is $\langle p, q, A \rangle$ instead of $\langle A, p, q \rangle$.

But if moving from sentences to propositions is just moving from sentences in a natural language to ones in a more rarified language whose symbols are abstracta, then the move fails to provide the language-independence that the notion of propositions was supposed to yield. And, indeed, it threatens a classical third-man regress. If we introduce propositions to be that whose expression synonymous sentences have in common, and if propositions are

themselves sentences, albeit ones in a rarified language whose terms are Platonic abstracta, then either:

- (a) there should be a yet third class of sentences to explain the synonymy between the sentences of natural language and the propositions which on this account are sentences, or
- (b) we have no principled reason to introduce propositions, at least for this purpose, since if (a) is false then at least sometimes synonymy between sentences can be explained without reference to any entities other than the sentences themselves (at least in the case where one of the sentences is a proposition) and hence the possibility is opened up of such an account being given of synonymy in general.

The first option leads to a vicious regress while the second removes a good part of the epistemic ground for supposing there to be propositions—and hence removes much of the warrant that the existence of Lewis’s account of propositions gives to the real existence possible worlds. Of course, one might have *other* reasons for positing propositions. But if propositions are really sentences, albeit in a rarified language, then one doubts whether there will *ever* be a principled argument for why one needs something other than the “more ordinary” sentences.

It might be argued that there is a gain. Lewisian structured propositions have as their ultimate constituents, along with set-theoretic constructions, the *unstructured* propositions and properties, which are arguably relevantly different in kind from ordinary language sentences and predicate expressions. Hence, there is a significant conceptual difference between a Lewisian structured propositions and properties and parts of natural language—the structured propositions, for instance, include as parts collections of those possible worlds that satisfy those parts, while no such thing can be said about ordinary language expressions. That is true, but it does not help Lewis much. For if one of the main theoretical goals of an account of propositions is to provide an account of synonymy, then the part played by the unstructured propositions is to account for logical equivalence. But that two sentences are logically equivalent (i.e., they cannot differ in truth value) is only a very small and least troublesome part of their being synonymous. The structuring of propositions that Lewis engages in is supposed to allow them to play the more difficult role. But this structuring is essentially linguistic and hence is no gain—else the third man regress arises.

Finally, observe that Lewis’s account of structured propositions presupposes that there are *simple* or *atomic* propositions that cannot be analyzed as connected combinations of other propositions and that out of these simple propositions all other propositions are composed—call this thesis *general atomism*. For the Lewisian structured propositions are built up out of unstructured propositions, and the unstructured propositions cannot be analyzed into other propositions, since they are just sets of worlds. ^[47] If general atomism is false, some proposition *p* is analyzable into a combination of simpler propositions, some of which in turn are analyzable into a combination of yet simpler propositions, and so on without termination. To model *p* in his system, Lewis would have to cut off the infinite complexity of *p* at some point, making some level of analysis consist entirely of unstructured propositions, thereby destroying the finer structure of *p* and hence failing to model it. If we think that the existence of propositions of such infinite propositional complexity is a genuine possibility, then we will have reason to be dubious of Lewis’s theory of propositions prejudging against this (observe how Lewis, 1986b, himself worries about the epistemic possibility of infinite complexity and uses it as an argument against Armstrong’s view of structural universals).

If Lewis’s account were to give one an intelligible ontology for propositions in terms of possible worlds, this would be a big asset for those possible-worlds that do not construct possible worlds out of propositions. But, alas, Lewis’s account has proved unsatisfactory, whether in its unstructured or structured incarnation. However, the unstructured incarnation may have some limited theoretical uses. For some intents and purposes, the differences between logically equivalent propositions can probably ignored. For these purposes, Lewis’s account may be a useful tool. But since the “unstructured propositions” are not really propositions, the anti-realist about possible worlds can accept unstructured propositions just as much, considering them to be mere useful fictions.

We can observe that this attack on Lewis’s account of propositions takes much of the wind out of Lewis’s

sails when he assails those accounts that construct possible worlds out of propositions for their failure to give an account of what propositions are. For, Lewis, too, surely needs propositions, albeit for different purposes, and he cannot give a satisfactory account of them either.

Section 5 Properties

Lewis offers an account of properties that is similar to his account of propositions. Thus, a property may be seen in an unstructured way as the collection of individuals in all possible worlds that has that property. And, similarly to the way structured propositions, if we wish to distinguish properties that are necessarily co-extensive, we can use set-theoretic constructions to produce “structured properties”. Lewisian propositions can then be seen to be properties of worlds.

The difficulty with the unstructured properties is the same as the one in the case of unstructured properties. However, there is a serious problem with structured properties as well. Granted, one can in a quasi-linguistic way distinguish a number of co-extensive properties. For instance, if \dot{U} is the ternary relation between unstructured properties which holds if and only if the third relatum is the intersection of the first two, then we can distinguish the conjunctive property of being a horse *and* a mammal from the necessarily co-extensive property of being a horse by using $\acute{a} \dot{U}, \acute{a} H, M \tilde{\tilde{n}} \tilde{\tilde{n}}$ for the first property and H for the second, where H and M are the properties of being a horse and a mammal respectively. The main objection against structured propositions was the arbitrariness of the set-theoretic construction. One might level a similar objection against structured properties now. If one thinks of the properties of an object as those entities in virtue of possessing which various predicates are predicated of the object, then one might wonder why a horse should possess $\acute{a} \dot{U}, \acute{a} H, M \tilde{\tilde{n}} \tilde{\tilde{n}}$ as opposed to, say, $\acute{a} \acute{a} H, M \tilde{\tilde{n}}, \dot{U} \tilde{\tilde{n}}$. Properties are those aspects of reality that are pointed out by an attribution of a predicate to a subject. These should thus be uniquely defined, just as propositions should.

There may be thought to be another objection against Lewisian structured properties. Let us suppose the structured account is correct. Now, a structured property that a thing has is built up set-theoretically out of other properties. At the pain of a regress, we must eventually come to *basic* properties which are not themselves set-theoretic constructions out of other properties. [48] These basic properties will be unstructured. But now here is the problem: Is it not possible that there be two basic properties that are necessarily co-extensive? If there are such properties, then Lewis’s account fails. For *these* properties in light of their basicness cannot be distinguished as different set-theoretic constructions, since they are not set-theoretic constructions at all.

But can we distinguish necessarily co-extensive basic properties by some set-theoretic construction that involves, say, attaching numbers to them? Thus, supposing F and G are necessarily co-extensive basic properties that are distinct, we could let S be the (unique) unstructured property corresponding to them and then label F as $\acute{a} 1, S \tilde{\tilde{n}}$ and G as $\acute{a} 2, S \tilde{\tilde{n}}$. But there is no reason why we should not label them the other way. *Ex hypothesi*, they are different properties, and it will not do to say that we might as well treat “ $\acute{a} 1, S \tilde{\tilde{n}}$ ” as F and “ $\acute{a} 2, S \tilde{\tilde{n}}$ ” as G —one might as well say that it doesn’t matter whether we treat F as F and G as G rather than treating F as G and G as F . We should not be just modeling—for then we could just use that reality we are modeling instead of the Lewisian properties—but are supposed to be identifying the properties, surely, perhaps *pace* Lewis’s intent.

Lewis thus needs to deny the existence of extensionally equivalent distinct basic properties. But that should not be a great burden, since it is difficult to point to any. [49] The arbitrariness of construction objection remains, but it is only serious if one thinks, *pace* Lewis, that there is an objective fact of the matter as to what *the* properties are.

The problem of infinite complexity reappears in a particularly clear way. There is a plausibility to the logical possibility of a property having *infinite* complexity, i.e., of it not being capable of analysis into basic properties that cannot themselves be further analyzed. On one account of properties, when we learned about

molecular motion, we learned that the property of being hot is nothing but the property of having molecules that move about rapidly. When we learned that molecules are built up out of atoms, we learned that the property of having molecules is the property of having atoms that are bonded together. When we learned that atoms are built up out of electrons, protons and neutrons, we learned that the property of being an atom is the property of having electrons, protons and neutrons appropriately arranged. When we learned about quantum mechanics, we learned that “being appropriately arranged” is a certain mathematical property of wavefunctions (as opposed to, as we had wrongly thought, being a mathematical property of sharp spatial positions). Is it absurd to suppose that this process could continue without an end, without there *being* an ultimate analysis of the property of being hot as a combination of properties that themselves on conceptual grounds cannot be analyzed further? It is traditional to call this scenario the “possibility of infinite complexity”, though the term is misleading: the property of “having F_1 or having F_2 or having F_3 or ...” is not “infinitely complex” in this sense if the F_i are basic non-analyzable properties because there is an ultimate, albeit infinitary, analysis into properties that are not themselves further analyzable.

If infinite complexity is possible, then Lewis’s account of properties is wrong. For Lewisian unstructured propositions cannot be further analyzed, whereas his structured propositions can all be analyzed into unstructured ones. Hence, the Lewisian account is at base an atomistic one, and therefore incompatible with the possibility of infinite complexity. Now perhaps Lewis could say that the fact that the best account of properties, namely *his* account, leaves no room for infinite complexity gives one good reason to think that infinite complexity is impossible. But this would only work if Lewis’s account was *the* correct account of propositions and properties understood in a realistic fashion. If it is simply a somewhat arbitrary account of one particular way of modeling properties, then its inability to model infinite complexity should not reflect on the logical possibility of infinite complexity—it might simply reflect on its inadequacy. If the account were a realistic one that gave an account of what *the* properties *really* are, then by inference to best explanation we would have reason to believe the account to be realistically true and hence to believe that what is incompatible with it is false.

One might also object to the Kripkeanism in the above account. I am not merely claiming that those things that are molecules are made of atoms. I am claiming that *to be* a molecule is *to be* made of atoms arranged in a certain way (and I allow here the possibility that we may not be able to specify reductively what this way is other than as “that way of arrangement of atoms that things of *this* kind have”, with the “this” pointing to atoms), which entails the Kripkean claim that, *necessarily*, whatever is a molecule is made of atoms. It would, however, go beyond the scope of this paper to argue for this claim, beyond pointing out that the naturalness of telling the story I did. It is quite natural to say that we have learned something about the *property* of heat when we learned that heat is the rapid movement of molecules, and not just something about the hot things that exist.

Of course, one might come up with an argument against the possibility of infinite complexity. ^[50] But without such an argument, the Lewisian construction of properties should not prejudice the issue.

Section 6 Overall assessment

Possible worlds let one formulate in a uniform way various modal notions that seem to intrinsically involve consideration of and comparison between more than one world: e.g., supervenience, transworld comparison of individuals, and counterfactuals. Moreover, ordinary modal claims like “I might have been a physicist” or “Hitler might never have been born” are naturally disambiguated against a background of possible worlds, with the context determining which worlds we are quantifying over. Since there is good reason to think that all these modal claims make sense, and since a very natural way to make sense of them is possible worlds, this gives us good reason to believe there are possible worlds.

But of course there is more than one theory of possible worlds. If we should find that only one extant theory is a serious option that withstands all criticism, then the fact that there is good reason to believe there are possible worlds will provide us with good reason to believe that this theory is true. I shall argue that of the theories

under consideration, only one has the hope of being satisfactory: an Aristotelian modification of Leibniz's theistic theory. Until a better theory should be found, the fact that this theory fits the theoretical data while the others do not gives us some reason to think it true (and, by implication, to think that God exists). But first we must consider the alternatives.

Part III. The Lewisian ontology of extreme modal realism

Section 1 The Lewisian account of possible worlds

For purposes of discussion of Lewis, I will call a maximal mereological sum of objects that are spatio-temporally related to one another a “universe”. I shall assume that being spatio-temporally-related-to is a transitive and symmetric relation. ^[51] By definition, if there are two distinct universes, there are no spatio-temporal relations between them. David Lewis’s extreme modal realism (EMR) then claims that each possible world is an existing universe ontologically on par with the universe we in fact inhabit and every possible way for a universe to be is a way that some concretely existing world is. In a Lewisian context, thus, “possible world” and “universe” are interchangeable. Not so, of course, in other contexts: if possible worlds are maximal sets of compossible propositions, say, then a possible world is not a universe, since there are no spatio-temporal relations between propositions or sets of propositions.

A proposition, then, is true *at* a world providing it truly describes a state of affairs obtaining in that world. Quantifiers in many ordinary language propositions are restricted to the world at which we are concerned with their truth value. The tokening of a proposition is truthful if and only if the proposition is true at the world at which it is tokened. It is true at w that there exists a unicorn if and only if in the concretely existing universe that w is, there is an individual that is a unicorn. My ordinary utterance of “There exists a unicorn” is truthful if and only if in *our* universe, i.e., in the maximal aggregate of things that are spatio-temporally related and that includes me, there is an individual that is a unicorn. It is irrelevant whether there are such individuals elsewhere. Thus, it is false that there exists a unicorn. However, context may indicate a wider scope for quantifiers, allowing them to range over individuals in multiple worlds. Thus, speaking as a Lewisian philosopher, I could say “There exists a unicorn”, provided the context makes clear that I allow the quantifiers to extend over the domains of all worlds.

For Lewis, ontologically all worlds are on par. Our world and universe is just one among infinitely many. We say it is *actual*, but this only says it is *home*—being actual is not an absolute non-relational property of our world, any more than something’s being *home* (as opposed to being *a* home) is of a house. Like something’s being home, a world’s being actual is nothing but an indexical claim. The actuality of the actual world consists in nothing but its being the maximal spatio-temporally connected aggregate that contains *us*. Moreover,

at any world w , the name “the actual world” denotes or names w ; the predicate “is actual” designates or is true of w and whatever exists in w ; the operator “actually” is true of propositions true at w , and so on for cognate terms of other categories. ^[52]

In the standard Lewisian model each individual exists in only one possible world. For if some individual x is a member of worlds w_1 and w_2 , then all items in w_1 are spatio-temporally related to x and likewise all items in w_2 . But since being *spatio-temporally-related-to* is symmetric and transitive, it follows that all items in w_1 and w_2 are spatio-temporally related to one another, and since the worlds are supposed to be maximal, w_1 and w_2 are identical.

But if every possible way for a world to be is a way that some world is, then surely there is a world at which ARP has dyed his hair green. Since it is false at the actual world that he has done so, one might think that it follows that ARP exists in some other world, a world where he has green hair. However, on Lewis’s account, it is not *ARP* himself who exists at that other world, but someone very much like him, a *counterpart* to him, albeit with green hair. Lewis can then enrich the semantics of “... is true at ...” by saying that a proposition about some particular individual x is true at some other world (where x does not exist by the above argument) if and only if a *counterpart* to x exists at that world and what the proposition had asserted of x is true of the counterpart. On this semantic move, one can say that indeed there is a world at which ARP has green hair. That is a world at which ARP’s *counterpart* has green hair.

To say that one individual is a counterpart of another is to affirm that there is a certain contextually-determined similarity between them. If we consider *being-a-counterpart-of* as a relation that, for any given individual x in a world, picks out the unique individual, if there is one, in another world who most closely resembles x , assuming this resemblance is “close enough”, then *being-a-counterpart-of* will be a non-symmetric non-transitive relation for Lewis. The non-transitivity will be discussed below, in Section 4.2.1.b. The non-symmetry can be seen as follows. Suppose that I have an almost identical twin in the actual world. The only significant difference between him and me is that he has a big birthmark on his arm which I do not have. Now, there is a world, w , where there is a perfect copy of this twin of mine, but where there is nobody else even remotely like me or my twin. The copy of my twin is then a counterpart of mine and makes it true to say that I have a big birthmark on my arm at w . But I am not the counterpart of that copy of my twin—my twin is. For my twin resembles him more than I do, since it is he who has the birthmark.

Actually, for the interpretation of many propositions Lewis (1979b) thinks we will need to speak not just of truth at a world but truth at a world and at an individual in that world. This is how Lewis makes sense of the fact that there can be a world w containing me with a big birthmark I don't in fact have *and* containing another person just like me but who does not have that birthmark. If we imagine such a universe, my counterpart will be the fellow *without* the birthmark since he is more similar to me. [53] Hence, that universe is *not* one at which I have a birthmark. Here, we need to interpret the statement that at w it is true that I have a birthmark and there is someone just like me but without the birthmark by assigning it a truth value at the ordered pair (w , the fellow in w just like me with the birthmark) and not just at w . This consideration shows that a Lewisian needs to allow that a given person has more than one counterpart at a world. The fellow without the birthmark can be a counterpart of mine as the fellow with the birthmark. The context constrains who *could* count as a counterpart of whom. It is certainly not true, except in a really strange context, that I have a counterpart with a birthmark at a world at which no human beings have any birthmarks but all donkeys do.

Since Lewis wants to reap the benefits for modality of the available of possible worlds, he will define possibility as truth at some world and necessity as truth at all worlds. To take care of the sorts of issues mentioned in the previous paragraph, one might have to sometimes modify this by saying that possibility is truth at some world-individual pair (or even world-individual₁-...-individual _{n} ($n+1$)-tuple), where the individual (or n -tuple of individuals) is a counterpart to the individual (or individuals) that the proposition is ostensibly about. Thus, we can say it is possible that I have a birthmark which actually I don't have but there be someone just like me, but distinct from me, and without the birthmark. This is possible, because it is true at (w , the fellow in w with the birthmark) where w is as in the previous paragraph.

Section 2 Identity vs. counterpart theory

If we are going to believe that all the ways that some world actually is a way that some world really is, which I shall call “basic EMR”, we still have a choice whether to accept identity theory or counterpart theory. If we accept identity theory, then the same individual ends up existing in multiple worlds. Lewis refers to this as an “overlap of worlds”, in that different worlds overlap in having the same individuals in them. If we accept counterpart theory, then each individual exists in exactly one world but may have counterparts in some others.

2.1 Arguments for counterpart theory

The only positive argument for counterpart theory that I am aware of is an argument, perhaps implicit in Leibniz's reasoning behind his view that each individual existed in only one world, to the effect that (a) there evidently are such things as essential properties, and (b) there is no intelligible way of drawing a distinction between essential and other properties (cf. Leibniz's “Discourse on Metaphysics”). As a result of (a) and (b), it is concluded that all properties are in fact essential. But “existing in world w ” is a property, and hence an essential property, and

thus no individual that has it can lack it, and so if an individual exists in world w it exists in no other world.

This argument is not the usual reaction to someone's claiming (b). The natural reaction is not to claim that all properties are essential, but that all properties are inessential. However, I shall not take the route of this objection to the argument, because there are very good reasons to believe in essential properties as we shall see in Section 4.2.1.b. Instead, I shall object to the claim that there is no intelligible distinction between essential and non-essential properties. For if basic EMR is true and if identity theory is true, there is a very simple and highly intelligible way of drawing the distinction: a property of an individual is essential if and only if the individual has this property in all worlds. To deny the intelligibility of this is to beg the question against the identity variant of EMR and not to give an argument against it. Moreover, even if we have a theory of possible worlds other than basic EMR, then the observation that there are clear cases of essential properties, e.g., being a point in space-time, and there are clear cases of inessential properties, e.g., having hair that is dyed green, outweighs the worry that there is no distinction. The fact that we do not always *know* how to draw the distinction does not militate against the existence of a distinction.

Lewis has given a more formidable argument. We normally distinguish between relational properties, such as *being a father*, and non-relational or innate properties, such as *being square*. But if the identity variant of EMR is true, then something (e.g., a sponge or an amoeba) may be square in one world and round in another (e.g., a sponge), and so being square is a property we must relativize to a world. Otherwise, we violate the law of non-contradiction. But, generalizing, it follows that *every* non-essential property is relative to a world.

For similar reasons, Lewis rejects the theory of numerical identity over time, preferring a theory of temporal stages that make up an individual considered as a space-time worm. However, it is certainly open for an identity theorist to insist on the older Aristotelian understanding of the law of non-contradiction according to which things are barred from having contradictory properties *at the same time*, but are allowed to have them at different times. By the same token, perhaps, the same being can have contradictory properties in different worlds.

But what does it *mean* to say that something is square relative to one world and circular relative to another? What is it *really* like? (One cannot give the answer that what things *really* are is their individual essences, whatever those are. For the essence is not the answer to the question of what things are *like*, but what they are *simpliciter*.)

The mystery about what properties like shape are, given that they turn out to be relational, does not even disappear if we say that space itself is relational. For in a given extended object, there will presumably be internal spatial relations as well as external ones. The internal ones will differ depending on the shape of the object. Thus, a circular disc has the *internal* property of containing a point such that all peripheral points are equidistant from that point. But since the disc could have been square-shaped, it follows, assuming basic EMR and identity theory, that these internal distances in the disc are not merely binary relational properties between points in the disc, but are ternary relations between the two points and the world (or maybe other items in the world). This seems to be an unpleasant complication.

But a complication is not a contradiction. It makes a theory more expensive, but does not knock it down. Maybe we are wrong about distances being binary relations. Maybe we are wrong about electric charges being innate properties. Perhaps everything is more holistically relational.

But now here is a more serious problem. What are the innate properties relational *to*? First to show the problem more clearly, suppose our identity theorist accepts a non-relational view of space-time. She will then presumably accept identity theory for space-time as well, since the same arguments as she advances for identity

theory in other cases apply [\[54\]](#), so that a given point will exist in the space-time of several different worlds, and different worlds might have the same space-time. But if she does *this*, then she cannot say what might be the most natural solution to the problem, namely that shapes are relational properties with the other relatum being space-time. For if something exists in more than one world, it might be that both worlds have numerically the same

space-time. So when we say that something is square relative to one space-time and circular relative to the other, we are affirming something contradictory—since there is only one space-time in the two universes, and nothing can be square and circular relative to the same space-time, at least not in the same way.

If we think space-time is relational, the same problem reappears. For unless we think that “the world” is something over and beyond the mereological sum of its parts, something that could itself stand in relations, we will want to say, e.g., that this amoeba is square in relation to the other individuals in the world, while in another world it is circular in relation to the individuals of that world. But it is conceivable that the two worlds have the same individuals, the only difference between them being the shape of this amoeba. Then, we have said that the amoeba is square in relation to the same individuals in relation to which it is circular. And this is absurd.

The remaining solution for the advocate of the identity variant of EMR is to insist that “the world”, which basic EMR insists is nothing else than “our universe”, is something that things can stand in relation to. Thus, the amoeba is something that is square in relation to one world and circular in relation to another. But now another problem appears. If the world is itself a concrete individual, as for Lewis it is, then the same kind of counterfactual reasoning that would make one think that *this* very amoeba which is actually circular could be square apply to the *world* at large. *This* very world, which in fact is populated by people, might have been populated by mere hydrogen gas. *This* very world or universe, which in fact is populated by our circular amoeba, might have contained this same amoeba which, though, was square. But then our amoeba is circular and square in relation to the very same world. The identity theorist who accepts basic EMR thus needs to exempt worlds from his identity theory. But this is not satisfactory, given that basic EMR equates worlds with universes, and hence with concrete individuals, and identity theory should surely hold either for all or for no individuals.

Another EMR-based argument against identity theory is that based on Lewis’s analysis of actuality. If for a world to be actual is for it to be the world in which I exist, and if I exist in more than one world, then more than one world is actual, which is absurd.

However, without EMR, we have no good argument against identity theory, this time construed as the assertion that when we are saying that something could have had that property we are not saying this in virtue of any individual other than the one we’re talking about, except the dubious one that proceeds by denial of the difference between the essential and the accidental.

2.2 Arguments for identity theory

2.2.1 General arguments

As has often been argued, someone very much like me becoming a spelunker in another world cannot be a truthmaker of the proposition that it is possible that I become a spelunker. How indeed, it is asked, is it at all even relevant to the proposition? This way of putting the issue is not sharp enough since that someone in *this* world who is very much like me becomes a spelunker is good *evidence* for the proposition that I could become one, too. But it is not *because* someone like me becomes a spelunker in this world that it is possible for me to become a spelunker. Indeed, the explanation goes the other way. Because it is possible for me to become a spelunker, it is possible for someone like me to become one, and those capabilities that ground that possibility in that other person also enter into the explanation of his actually becoming a spelunker. Why, then, should things be different when that fellow who is like me in fact dwells in another world? Why should he *then* become involved in the truthmaker? His example may inspire my imagination, thinking about him might lead me to regret me not having taken his exciting road in life, but his example plays an essentially different role than the role of *making* it true that I could have become a spelunker.

An analogy might help. Suppose we have a system which emits a light when a button is pressed. We certainly would not want to say that the fact that light was emitted at t_0 , t_1 , t_3 , and t_4 when the button was pressed, with the situation these times being closely analogous to t_2 in terms of setup, is what *makes it true* to say that at t_2

the system also had the dispositional property that it would respond with light were the button pressed. Certainly, the facts about the activation of a like dispositional property t_0 , t_1 , t_3 and t_4 are evidence, perhaps even conclusive evidence, that the property was had at t_2 . But one must not confuse the evidence with the truthmaker. In this case, such a confusion would lead one to buy into a Humean reduction of dispositional properties (and laws, too) to occurrent states. Lewis appears to be guilty of a similar reduction, albeit the occurrent states in his account are in other worlds. At the same time, the Humean comparison shows that a respected philosopher *can* hold a view of this sort, and so more of an argument is needed than just pointing out the kind of reduction that is going on. [\[55\]](#)

2.2.2 Attributions of ability

A person is only free to do something if she *can* do it. A Lewisian analysis of the modal claim that someone can do something will involve statements about what some of the person's counterparts in other worlds in fact do. For instance, if one's notion of ability is of the sort incompatibilists bring in, then a necessary condition for my now being able to do something is, in Lewisian terms, that some counterpart of me who shares a copy of my past up to now in fact does it.

However, here the objection discussed in Section 2.2.1 comes in: That I *can* do something is surely a statement reporting a fact specifically about me, not about another person such as a counterpart of me. After all, my being able or unable to do something has normative import for *me*: if I am unable to do something, then I was not free to do it, and hence I cannot be held responsible for not having done it. The objection, however, as given appears to merely beg the question. After all, the statement that my counterpart does something is, on Lewis's theory, a statement about *my* modal properties, and hence in a straightforward sense of "about" a statement about *me*.

If the objection is not to beg the question, a more precise sense of "about" must be brought in. The notion of a truthmaker helps. We can say that a true proposition is *about* some existent thing, if that thing is one of the items *involved* in the proposition's truthmaker [\[56\]](#), where the notion of being "involved" in the kind of thing which is the truthmaker of a proposition is taken as primitive and illustrated by propositions such as:

1. The butler, the master, the stabbing and the knife all are *involved* in the butler's having stabbed his master to death with a knife.
2. Socrates is *involved* in the truthmaker of the proposition that Socrates' existed.
3. Clinton is *involved* in the truthmaker of the proposition that there was a 42nd President of the United States.
4. If [\[57\]](#) pain is (reductive identity) nothing but a firing of one's C421, then my C421 is *involved* in my being in pain. [\[58\]](#)

The objection now is that surely I must be involved in the truthmaker of the proposition that I am now able to, say, jump. But the truthmaker of that proposition on Lewis's account, it seems, is simply the jumping of relevant counterparts of mine, and this involves merely the actions of certain *counterparts* of mine and not me.

This objection, however, fails because it commits a *de dicto* / *de re* modal fallacy. Suppose x , y and z are rigid designators of those counterparts of mine whose actions are involved in making it true that I can jump. Then, it is not only x 's, y 's and z 's jumping that makes it true that I can jump, but also their being relevant counterparts of mine. It is true that for Lewis that I can jump is made true by the jumping of certain counterparts of mine, but here "counterparts of mine" must be read *de dicto*: one cannot substitute the rigid designators " x , y and z " for "certain counterparts of mine". But I am involved in the truthmaker of the proposition that x , y and z are relevant counterparts (or even just counterparts) of me, since the truthmaker of that proposition is the relevant similarity between me and x , y and z , so that likewise I am involved in the truthmaker of the proposition that I can jump.

But despite this analysis, one may have a feeling that x , y and z are interlopers in the truthmaker of the

proposition that I can jump. Why should their actions be *at all* relevant to the truth of this proposition? The worry now is not that I am uninvolved in the truthmaker, but that some people who have no business being involved are. One may think that in some sense the proposition that I can jump should *only* be about me, my intrinsic properties and my immediate surroundings (including my relations to my immediate surroundings). For why should the actions of other people causally isolated from me have any bearing on the normative claims that follow from my being able or unable to jump?

There is, however, a way in which Lewis can accommodate some of the background of the intuition that I, my intrinsic properties and my surroundings are the only things that are involved into the truthmaker of the proposition that I can jump. Lewis can say that I, my properties and my surroundings are the only *actual* things that enter into this truthmaker. But this would only help if one thought, as Lewis vehemently denies, that to be actual and to exist are the same thing. Given Lewis's distinction between the actual and the existent, the mystery as to why some non-actual but nonetheless existent persons should be involved in the truthmaker of the proposition that I can jump remains.

Moreover, the answer that the other persons involved in the truthmaker are non-actual will not do on Lewisian principles. For we can imagine a case where many of those persons whose actions make it true that I can jump are in fact actual. Imagine I am in an almost deterministic world in which at the beginning of every century one new human being has come into existence, lived for a hundred years, and then disappeared. Then a new human being, qualitatively indiscernible from the previous, appeared, and the cycle went on. This would of physical necessity go on forever in the future, except that there is one and only one bit of freedom and contingency that each human being is allowed. At the beginning of the 27th year of his life he can jump or refrain from jumping. If he jumps, everything will go on as before, a new human being appearing at the end of this one's life, and so on. But if he refrains from jumping, the laws of nature and initial conditions are thus constituted that after he dies, no more humans will live, and there will no longer be any contingency allowed by the laws of nature. Suppose now I am one of these human beings, and I refrain from jumping. Up to my time, every century was repeated. After me, everything is different and the laws of nature allow no more contingency.

Consider the true proposition that I could have jumped at the beginning of my 27th year. This is made true by the jumping of my counterparts in worlds with the same causal structure. First assume that there are no qualitatively identical but distinct worlds. Then in fact, there are only two relevantly similar different worlds with the same causal structure as the actual world. One of these worlds, w_1 , is a world of endless recurrence where in fact every human being from $t = -\infty$ to $t = +\infty$ jumps at the beginning of his 27th year. The other world, w_2 , is one where an infinite number of human beings jump, but then one doesn't, and then everything is different. The proposition that I could jump at the beginning of my 27th year is made true by my counterparts in these two worlds thus jumping (i.e., by the jumping of all the people in w_1 , and of all the people in w_2 prior to the first who did not jump). But now observe that w_2 is in fact qualitatively identical with the actual world, and hence is in fact the actual world since we have assumed that qualitative identity implies identity for worlds. Hence, among those counterparts whose jumping makes it true that I could have jumped, are counterparts living in my world, i.e., *actual* persons. The same conclusion holds even if qualitative identity does not imply identity of worlds, because nonetheless the actual world will be among those worlds that are relevantly similar to the actual world, and my worldmates in the past will still be good counterparts for me. Hence in any case, some of the people involved in the truthmaker of the claim that I could have jumped are *actual* people who existed prior to me in the actual world—and *these* are interlopers by any account. These people's existence and their jumping may *illustrate* my being able to have jumped, but surely my ability does not even in part *consist* in their having jumped.

Observe that there is a simple argument that shows that Lewis cannot possibly give a possible worlds analysis of "I can now jump" if, as is plausible, the truthmaker of this proposition only involves me, my properties and my causally relevant surroundings. For if the truthmaker only involves these entities, then the proposition would be intelligible in the absence of any worlds other than this one—since its truthmaker is wholly this-worldly.

But if the proposition would be intelligible in the absence of worlds other than this one, then other worlds cannot enter into its analysis.

The best strategy for Lewis here would be the *tu quoque*. On competing views of possible worlds, entities other than I, my properties and my surroundings are also involved in the truthmaker of the proposition that I can jump. For instance, views which construct possible worlds out of abstract propositions will have abstract propositions involved in the truthmaker. For instance, the truthmaker of the proposition that I can jump might be the-proposition-that-I-jump's being compossible with circumstances similar to those in which I in fact find myself, which clearly involves *the-proposition-that-I-jump*, an entity not part of me, my properties or my surroundings, and in no way causally relevant to anything concrete. Lewis can ask: Why should the intrinsic properties of this abstract entity be involved in *my* concretely being able to jump? Thus, perhaps, whether we be Lewisians or not, we simply must accept that my being able to jump involves things other than me, my properties and my surroundings.

Lewis's opponents here can argue that the entities *they* involve in my being able to jump are *abstract*, and hence one should not complain of their presence here. But why not? Similar intuitions to those that say that the truthmaker of the proposition that I can jump should only involve me, my properties and my surroundings will also say that the truthmaker should be something entirely concrete—which on Lewis's view it is.

However, *tu quoque* answers always have the weakness that they become useless when a theory comes into view that does not share the bad features of the old theories. In Part VI, I shall argue for a theory on which the truthmaker of the proposition that I can jump involves only me, my qualities and my surroundings, and is in a relevant sense a *concrete* part of the world. But this is a distance in the future. At present, however, we can observe that there is a theory in view that *almost* does this: Lewisian multiple worlds with an identity theory of transworld identity. For, then, although other-worldly properties of me are involved in my being able to jump on this theory, the only *substances* involved are ones that exist in this world: I and those substances that exist in my surroundings. For, that I can jump, on this theory, just says that *I* jump—in some world with relevantly similar surroundings. Hence it is more understandable why *I* could be held responsible for not jumping if I do not.

Further ethical considerations will be brought to bear on this issue in Section 8.

2.3 Conclusions about identity and counterpart versions of basic EMR

Neither the identity nor the counterpart versions of basic EMR are completely satisfactory. However, the identity theory only requires that we revise our beliefs about what properties are relational and what are intrinsic, while allowing an exception for *worlds*, all of whose properties need to be taken to be essential, since otherwise EMR evidently collapses—multiple worlds become one, as it were. Such an exception is an uncomfortable thing given that basic EMR takes worlds to be themselves individuals, but it is not absurd. The counterpart version of basic EMR does, however, involve one in serious problems as to what the truthmaker of possibility claims is. On the whole, the identity version may be preferable.

The fact that neither version is quite satisfactory provides a dilemma argument against basic EMR. Either the identity or counterpart version is true. In each case problems ensue.

Section 3 Indiscernible worlds?

Some of my arguments, notably those in Section 8.3, Section 9 and Section 10, against EMR will presuppose the thesis that there are no indiscernible worlds, i.e., that no two worlds are copies of each other. The arguments in Sections 8.3 and Section 10 can be made to work in the case where there are only finitely many copies

of each world. ^[59] The real problem for those two arguments, then, is with the cases where there are infinitely many worlds. The argument in Section 9 will work if each world has an *equal* number of copies, finite or infinite.

Lewis (1986a, p. 84) himself is not committed to either the view that there are indiscernible worlds or to the view that there are not.

First of all, however, there is a simple argument based on the apparent absurdity of the following claim:

(19) It is possible for everything to be as it is while the actual world is not actual.

This claim is true if there are multiple copies of this world. For it is possible that the actual world not be actual, which claim simply says that there is contingency, and this possibility would be verified by another of the copies of this world in which copy everything would be as it is. If one thinks (19) to be absurd, then one will not accept multiple copies of worlds. But Lewis can always answer that we just lack proper intuitions about things like (19) since such claims are not made in ordinary life. Thus, other arguments are needed.

A quick argument against any indiscernible worlds could be based on the Principle of Identity of Indiscernibles which states that no two distinct objects are indiscernible, i.e., share all the properties describable in purely general terms. But one can also argue against indiscernible worlds on weaker assumptions about individuation, using an argument of Donald Turner's (forthcoming). The argument claims that indiscernible objects can be individuated only if they are in a common space and/or time and thus capable of spatial and/or temporal separation. If one takes this view, then distinct Lewisian worlds could not be indiscernible, because there are no spatio-temporal relations between Lewisian worlds.

One difficulty with this argument is that current physics appears to provide a counterexample in the form of elementary particles that are bosons (French, 1988), e.g., photons. Such particles can have the same wave function. Consequently, they are not spatio-temporally distinguished, and moreover they can be indiscernible in all other respects. The defender of Turner's argument against indiscernible worlds can proceed in two ways. The first is to adopt a controversial interpretation of quantum mechanics according to which while normally one describes the universe as containing n photons with k degrees of freedom each, it would be more correct to speak of a single "n-fold Photon", existing in up to n places at once, and having nk degrees of freedom. [\[60\]](#)

The second retort is to weaken the principle that indiscernibles can be distinguished only by actual spatio-temporal relations to a much more reasonable principle that says that individuals can be distinguished only by actual *or potential* spatio-temporal relations or actual *or potential* differences in properties. Multiple photons that in fact share a wave-function can be distinguished by virtue of being *potentially* spatio-temporally separated—namely, by virtue of it being possible that they are thus separated. But even this very weak principle of indiscernibles does not allow one to distinguish indiscernible worlds. For on Lewis's account there are no alternate possibilities for *pairs* of worlds, but only for individual worlds. So it is impossible to speak of a joint possibility for the two worlds where they are distinguished, e.g., by having different qualitative properties.

Another argument against indiscernible worlds might be from Ockham's razor. Why posit multiple worlds where a single will do? But, a plausible reply goes, a single world *will not* do, since the absurdities resulting from EMR that we shall observe in Section 8.3 and Section 10 to ensue on the assumption that there are no indiscernible worlds show this.

However, there is a plausibility argument based on arbitrariness considerations. If there are multiple copies of some world, one asks: How many? If 227, why not 327? One answer might be that the number is proportional to the objective probability of that world being actual. This might neatly solve the problem of induction that will be raised in Section 9: the worlds where induction hold maybe *are* more numerous because there are a lot more copies of them. The problem with this probabilistic answer is that surely probabilities of various worlds being actual surely can be irrational numbers (certainly objective probabilities of events can be irrational, if one thinks that there are non-deterministic events of the sort that non-deterministic interpretations of quantum mechanics talk about, while this approach would make them only rational numbers, i.e., ratios of numbers of copies worlds. So this will probably not do.

Multiple finite numbers of worlds thus introduce a very unfortunate element of arbitrariness into EMR. And we could surely just as well conceive of *different* numbers of copies of worlds. However, this conceptual possibility is one that EMR has no conceptual resources to analyze. For suppose there are 167 copies of this world

(counting this world as a trivial copy of itself), and we wish to say that there could have been 168 copies of that very world. How could we express this possibility? First of all, EMR as it stands has no conceptual resources for possibilities for multiple worlds at a time. But maybe we can extend EMR. We can say some proposition about multiple worlds is possible if there is some set of multiple worlds which is aptly described by that proposition. Thus, the claim that there could have been 168 copies of this very world is true if and only if there *is* a set of 168 worlds which makes it true. But these 168 worlds would all have to be copies of this world, thereby contradicting the assumption that there are precisely 167 such worlds. So, no, even an extension of EMR cannot handle this conceptual possibility. But unless there is a non-arbitrary account of the numbers assigned to each world, this possibility is surely there—and this possibility Lewis cannot admit.

Nor will it do to posit *infinite* sets of copies of each world. For the cardinalities of these sets will have the same arbitrariness that the finite numbers would. Why should there be \aleph_{167} instead of \aleph_{168} copies of this world? Is there not then a conceptual possibility of this cardinal number being different, a possibility EMR cannot countenance? This arbitrariness makes the theory of multiple copies of Lewisian worlds untenable, even if the Principle of Identity of Indiscernibles types of arguments did not.

There is a solution to the arbitrariness problem, and this is to posit that there is a Cantorian Absolute Infinity of copies of each world. Cantor believed that there was an Absolute Infinity, in which all the paradise of sets was contained. John Leslie (personal correspondence, 1999) used a similar solution to solve a problem for his theory according to which, in order that things be as best possible, there are infinitely many indiscernible omniscient deities. How many deities are there? Since there could always be more, the answer is Absolute Infinity. However, the notion of Absolute Infinity is so very murky that to have to escape to it is unacceptable. And of course if one escapes to it then one presumably can no longer even say that the collection of all worlds (or deities in Leslie's case) is a proper class. And this is surely absurd if the worlds (or deities) are concrete things.

Another solution to the arbitrariness problem would be to say that the number of indiscernible copies of each world depends on the number of individuals in it. Suppose that a world w_1 contained two discernible individuals, A and B , and that it would be logically possible for A to occupy the role occupied by B while B occupied the role occupied by A . Then, there would be some plausibility to the idea that there might be two copies of w_1 in order to model the distinct possibilities of A and B occupying the roles they in fact occupy in w_1 and of them occupying switched around roles. If instead of accepting counterpart theory one accepted identity theory, then in fact one could say that in w_1 the very individuals A and B occupy certain roles while in the copy-world the two individuals occupy the same roles switched around. This is indeed quite a plausible story given identity theory, and hence the identity theorist might quite reasonably want to have indiscernible worlds.

And perhaps the counterpart theorist might want indiscernible worlds for a similar reason. Were w_1 actual, she might say that a different world could have been actual, one in which A occupied the role of B and conversely. The most natural Lewisian way of making sense of this claim might well be to posit indiscernible worlds. And the number of copies of a given world w_1 is then equal to the number of ways that individuals can be assigned to the roles of the individuals in w_1 . Note, however, that the number of ways here is in fact plausibly greater than two. For, plausibly (at least for an appropriate choice of w_1), there will be a world w_2 , discernibly different from w_1 , which contains an individual C distinct from A and from B , which individual could fill either of the two roles. Thus, we want more copies of w_1 to account for *this* possibility. In fact, if we want there to be as many copies of w_1 as the number of pairs of possible individuals that could fill the roles of A and of B .

But now we have a problem. For on Lewis's account, we cannot distinguish between a possible individual that *could* fill the role of A from a possible individual that is identical with A , since all transworld identity is just a counterpart relation, and those individuals that *could* fill the role of A are precisely those individuals whose counterpart is A . Thus, in fact, the number of copies of w_1 must be equal to the number of pairs whose counterpart is the pair $\langle A, B \rangle$ in w_1 . This is indeed a non-arbitrary number, but the intuitive reasoning for getting this non-arbitrary number has just been undercut. For that $\langle A, B \rangle$ is a counterpart of $\langle C, D \rangle$, is supposed to be a model

of a transworld identity between $\langle A, B \rangle$ and $\langle C, D \rangle$. The properties $\langle A, B \rangle$ have in w_1 represent from the point of view w_2 the properties that $\langle C, D \rangle$ itself would have were w_1 actual. Hence if one says that $\langle C, D \rangle$'s filling the role that $\langle A, B \rangle$ has in w_1 is an alternate possibility from $\langle A, B \rangle$'s filling that role, one has contradicted the central insight of counterpart theory. Hence, the counterpart theorist is not entitled to the above intuitive account of the number of copies of a world. And even if she were, there would be the problem of individuating such worlds.

And I leave it as an exercise to the reader to check that each of the arguments against Lewis that uses the thesis that there are no indiscernible copies of worlds either is an argument specifically against the counterpart version of Lewis's theory or else is an argument that can be made to work given the above account of indiscernible worlds on assuming identity theory. [\[61\]](#)

Section 4 Lewis's arguments for his ontology

4.1 The analysis of actuality argument

Taking actuality to be nothing but indexical is something that Lewis cannot escape. For suppose on the contrary that actuality is an absolute non-relational property that happens to accrue to one and only one possible world. But "[s]urely it is a contingent matter which world is actual. A contingent matter is one that varies from world to world. At one world, the contingent matter goes one way; at another, another. So at one world, one world is actual; and at another, another. How can this be *absolute* actuality? — The relativity is manifest!" (Lewis, 1986, p. 94). Hence, Lewis cannot allow a non-world relative notion of actuality.

But Lewis does not mind being saddled with the idea of actuality being world-relative and indexical. On the contrary, it can be taken as an argument *for* his theory that all possible worlds concretely exist as universes. For observe that if, on the contrary, only this world was or corresponded to [\[62\]](#) a concretely existing universe [\[63\]](#), then one could define a non-world-relative notion of actuality: a world would be actual if and only if it is or corresponds to a concretely existing universe. Hence, by *modus tollens*, if actuality is merely world-relative and a non-world-relative notion is unavailable, then all possible worlds *do* concretely exist as universes. In other words, Lewisian worlds *must* be ontologically on par. This shows that *pace* Lycan (1994), not only is Extreme Modal Realism not independent of the thesis that actuality is indexical, but most likely EMR is true if and only if actuality is indexical. [\[64\]](#)

An analysis of actuality that would make it indexical thus would provide an argument for Lewis's theory. One might, for instance, be impressed by the fact that each world is actual at itself. This might seem to suggest that being actual is world-relative. But such an inference is fallacious. One might as well conclude that truth is story-relative because each story is true in itself, where we say that a proposition is true in a story if it is implied by the propositions of the story. The qualifier "... in a story" is truth canceling, and the opponent of EMR will insist that so is the qualifier "... at a world".

One might think that Lewis can give an argument for the indexical account of actuality simply on the grounds of its ontological economy. If we adopt this account, then we do not have to worry about ontological questions of what the mysterious nature of actuality is, of what it is that singles out one world amongst others to be actual (cf. Lewis, 1986a, Section 1.9). But the ontological economy here is perhaps illusory. The opponent of EMR can simply say that a world is actual providing the universe that exists at that world, the universe that corresponds to that world, exists ("*simpliciter*" one is tempted to add, but the temptation should be resisted since the opponent of EMR does not recognize another kind of existence), or, equivalently, if the universe that exists corresponds to that world. At first sight, nothing has been gained. The notion of existence is as philosophically mysterious as the notion of actuality. Indeed, to say that a world is actual if the universe that exists corresponds to it is not to go very far in analyzing actuality. But what it does show is that the opponent of EMR can analyze

actuality into a notion that Lewis's ontology *also* has: existence (or, existence *simpliciter*, if one insists). After all, Lewis himself asserts that all of his worlds are *existing* universes. If this assertion is comprehensible, then so is his opponent's insistence that a world is actual if the universe that corresponds to it exists.

This reply is not completely fair to Lewis, since it hides the fact that the anti-Lewisian notion of existence is of an absolute but *contingent* quasi-property. Many of the things that *are* might not have been. This is not so for Lewis's absolute notion of existence, construed as existence in the totality of all worlds. Each world there exists and it makes no sense to talk of a possibility of its not having existed. Thus, it might seem, if we attack the Lewisian account of actuality as above we are committed to a notion of existence that is contrastive in a way that Lewis's is not: existence is to be contrasted with mere possibility. However, this contrast is no more mysterious than one that Lewis himself must admit: existence (in the totality of all worlds) as contrasted with non-existence, i.e., impossibility. Just as Lewis will explicate impossibility in terms of absolute non-existence (square circles are impossible if and only if there is no world at which a square circle exists), the opponent to EMR will insist that non-actuality is to be explicated in terms of absolute non-existence. One is tempted to reply on Lewis's behalf here: "How can you talk of some *thing*'s not existing? Is *that* not mysterious?" But this should be no more mysterious than talking of impossible "things" not existing.

The only difference is that the opponent to EMR owes one an account of how it is we can say a non-existent thing is *possible*, what it is that we are talking about when we talk of such things. But this is a different issue: we are no longer debating the question of relative *versus* absolute actuality, but of whether there is a non-Lewisian account of modal language.

4.2 The cost-benefit argument

Lewis presents a cost-benefit argument for the truth of his theory. On the benefit side, it solves the problem of what modality is and, he thinks, has other useful applications.

4.2.1 What is modality?

4.2.1.a *A solution to the Parmenidean challenge and the mystery of modality*

Recall the Parmenidean challenge. It seems that when we talk about things that are non-actual, we are not talking of anything at all, and hence our sentences are either false or meaningless. Moreover, even if it is granted that there is some way of rendering talk of unactualized possibilities meaningful by providing it with objects, nonetheless it is not clear *what* sorts of things modal language affirms of these objects. EMR handles both concerns. What is a possible but non-actual horse? It is nothing but a horse, albeit in a different world. We know what horses are, and the non-actual horse satisfies the description "horse" in exactly the same way the familiar horses in this world do. The non-actual horse exists, though not here, and there is no problem in talking about it. Neither is there any problem in talking of a Socrates who was never a philosopher or of a unicorn. All these things exist elsewhere. There is no difficulty about their ontological status, because they have the same ontological status as the things we are familiar with. They only differ relationally from these things in that they fail to be spatio-temporally related to us.

Moreover, EMR tells us what we are actually asserting of a situation when, e.g., we say it is possible but not actual. Rather than predicating some mysterious shadowy existence of a shadowy object, we are simply affirming that the object exists and lacks the relational property of being spatio-temporally related to us. EMR allows us to reap in the fullest sense the benefits of a possible worlds rendition of modal language (cf. Section 1). On the official story, there is no primitive modality left on the ontological level: everything modal is analyzed in terms of the relations between concrete and unproblematic entities. The mystery of modality is removed. [\[65\]](#)

It would be a mistake to object that what worlds exist depends on what worlds count as possible, since the only way we can know about what worlds exist is in terms of what worlds are possible. Epistemological priority

does not imply ontological priority, as Aristotle had already noted. According to EMR, while it may well be that modal claims are “more knowable to us”, the plurality of worlds is “more knowable by nature”.

4.2.1.b *Essential properties, counterpart relations and primitive modality revisited*

However, there is an aspect of modality that Lewis’s official account does not provide a satisfactory analysis of. I clearly could have been a bank robber rather than a philosopher. But it is evident that I could not have been a set, the number seven, a point in space-time, the space-time manifold itself, a trope (a particular innate quality of a particular, e.g., *this*-negative-chargedness-of-*this*-electron), the Bank of England, the Swedish monarchy, the law of gravitation, or a particular act of jumping. It is less clear, but plausible, that I could not have been an electron. But it is not at all obvious whether I could have been a woman, or conceived by other parents, or a chimpanzee. Thus, there are some kinds of things that I clearly could have been, some that I clearly could not have been, and some where it is not obvious either way.

Even if it is not obvious either way, the presumption is that there is a fact of the matter as to whether I could have been something. After all, the question *prima facie* appears meaningful. It appears to have normative implications. For instance, in connection with one variant of the problem of evil, a theist might readily agree that it was possible for God to create a world populated with beings that are overall happier than we are—e.g., a world populated by an infinitude of angels that lead deeply meaningful lives—but that nonetheless God cannot on this account be said to be less than perfectly kind *to us* because in that world *we* would not exist whereas our lives are on the whole better than nonexistence. The debate over this kind of problem of evil presupposes the meaningfulness of the question whether *we* could have been the kinds of creatures that exist in the hypothetical world populated

with happier beings. ^[66] Or consider the following case. A girl was conceived as a result of this act of rape.

Because she was conceived through rape, her life was not perfectly happy. Assume, however, that overall her life was worth living. ^[67] Suppose then she sues the rapist. ^[68] Can the rapist’s lawyers claim that she could not have existed except as a result of rape, and hence overall has benefited from the rape since overall her life was worth living?

I am not claiming that if the answer to the modal claim that she could not have existed except as a result of

rape is positive, the rapist owes her nothing. ^[69] But I am claiming that the question is at least *relevant* and meaningful.

Another example where it is relevant what I could have been is the Rawlsian concept of justice. According to this concept, the just society is that which rational self-interested negotiators would come up with while under a veil of ignorance of their actual position and role in the world. Crucial to this account is the question of what exactly should be veiled. If the fact that the negotiators are not mosquitoes is veiled, then the just society will be one where the killing of mosquitoes is forbidden. If, on the other hand, I fail to veil the fact that my hair is dark, then in self-interest I will come up with a society where dark-haired people are especially favored. A natural suggestion as to what is to be veiled is that one should veil everything but the fact that one is a rational being. But that will not do, for then in a “just” society people who are severely mentally retarded for all of their lives will have no rights, which is absurd. It is reasonable to veil all properties that are not purely qualitative but have an indexical or haecceitistic component (if there are such properties). And since the point is to produce societal rules, we need only consider publically observable properties. But there is still the question of which purely qualitative observable properties should be veiled. Perhaps the only reasonable non-arbitrary determination of what is to be veiled is that if *Q* is a purely qualitative observable property, then it should be veiled whether one has *Q* if and only if it is both possible that one have *Q* and possible that one not have *Q*. In other words, we veil all non-essential purely qualitative observable properties. How well this approach works depends on what properties are in fact essential. If our only essential properties are being human and that which that property entails, then the approach

will work as well as a Rawlsian approach can be expected to work. ^[70]

If it is logically possible for me to be a mosquito, then the approach will yield a radical mosquito-rights ethic. It might be argued that if Kripkean intuitions about the necessity of origin are right, then the Rawlsian approach must fail entirely. For instance, if much of my genetic code is an essential property of me, then a society rationally negotiated by me will lopsidedly benefit people like me, it might seem. One way out of this problem would be for the Kripkean Rawlsian to veil all temporally impure properties, i.e., properties whose possession is a fact about more than one moment of time (assuming that we can make sense of this difficult notion), and consider the negotiation as being done by temporal stages. But my *genetic-code-at-t* is perhaps not an essential property, because it would arguably have been possible for a science-fictional genetic treatment to have gradually transformed me, without destroying the continuity of the underlying subject, into having the genetic code of another human. I do not know whether this is indeed so. But the question of whether it is or not, and the general question of what properties are essential or not, is crucial to gauging the success or failure of the Rawlsian analysis, and hence is of significance if that analysis has any plausibility.

The proposition that I could have been an *x* is a meaningful *de re* modal statement. On Lewis's account, the proposition is true if and only if there is a counterpart of me in some world who is an *x*. Thus the counterpart relation encodes modal facts about who and what I could have been. This encoding is a form of "primitive modality", i.e., modality that is further unexplained. Lewis's project is to explicate modality in terms of quantifications over possible worlds and possible individuals. Now, Lewis does explicate the counterpart relation in terms of relations of relevant similarity. However, he is willing to admit that this description underdetermines the counterpart relation, and holds that the counterpart relation is a contextually determined one. What we need to do, it seems, is *ourselves* to specify the contextual details we want in the counterpart relation in any given context when we want to invoke counterpart-based reasoning.

But this will not do when we want to decide a difficult case. Could God have made *me* a member of a different, happier species of creatures? (Note, by the way, that this question does not presuppose the existence of God, but only the *possibility* of a God's existing. If by "God" one only means something modest like "a very powerful and perfectly good supernatural creator of everything else" then the possibility will be readily granted.) Could the child conceived through rape have been conceived in a different way? We can stipulate that our counterpart relation is such that only members of the same species are counterparts, and we can stipulate that counterpart relations are not concerned with the history of the origins of beings. These stipulations would have one conclude that God could not have made me a member of a better species but the child conceived through rape could have been conceived differently. However, these claims would then only be true by stipulation and hence would fail to be relevant to the normative questions that these issues arise in.

Of course, one might claim that further analysis will reveal which counterpart relation is contextually implicit in, say, the question whether I could have been a member of a different species when the problem of evil is the issue that provides the context to the question. But to claim that the normative issue under consideration determines the counterpart relation contextually involved in the associated "Could *x* have been a *y*?" question is to deny the intuition that the answer to the normative issue depends on the absolute context-independent answer to the *de re* modal question.

The only way counterpart theory can make sense of "Could *x* have been a *y*?" questions in a way that is not stipulative or context-relative—and I have argued that we *do* want to make sense of these questions—is to make the counterpart relation primitive. On this variant of EMR, there is a distinguished relation between individuals which we call "*the* counterpart relation". Since this counterpart relation encodes modal facts, it is a form of unanalyzed primitive modality. Following Lewisian criticisms of ersatzist theories we can ask: "What is it that sets *this* relation apart from others and makes it into *the* counterpart relation?" This variant of EMR cannot answer this. The counterpart relation is one relation among many, but in some magical way set apart. While I take it that one cannot avoid primitive modality, *this* form of primitive modality fails to connect up with our intuitive notions of modality closely enough to render the analysis intelligible. We do not have an intuitive grasp of the counterpart relation,

except through an intuitive grasp of “Could x have been a y ?” questions. However it is precisely Lewis’s project to avoid basing modality on our grasp of modal statements, but to analyze these statements in terms of something prior and intelligible. Recall his trenchant way of putting his criticism of the Platonist approach:

What will not do is the pretence that we have somehow reached the unreachable, and got hold of a word for a specific one of the internal relations that lie beyond our reach. (Lewis, 1986a, p. 178)

But the counterpart relation is unreachable in this way.

Nor can Lewis allow “ x could have been a y ” propositions to be primitive in general, so that a counterpart relation could be based on them. For if he did, then much of modality could be reconstructed on this basis, thereby undercutting most of the need for EMR. After all, as Plantinga (1982) has noted, that p is possible can be analyzed into the *de re* proposition that p could have been a true proposition. Lewis could avoid this by only allowing “ x could have been a y ” assertions to be primitive in the cases where x is concrete. But doing this would mean that for many propositions, a primitive-modality analysis will work (“It is possible that Jones mows the lawn” is analyzed into “Jones could have been someone who mows the lawn”), and Lewis’s quantificational possible-individual analysis will only apply in a patchwork of cases. This would not be a satisfactory state of affairs.

What is left for Lewis is to insist on the claim that there is no privileged context-independent objective answer to “Could x have been a y ?” questions. The revisionary nature of this account of course counts as evidence against Lewis’s view. Lewis, however, is not without positive arguments for his revisionary account. He recalls Chisholm’s *sorites*-based argument that bids us to consider a sequence of possible persons, the difference between any two successive ones being small, but where the first person on the list is qualitatively identical with, say, me and the final one with, say, Napoleon. If one accepts the transitivity of the counterpart relation, as Lewis himself does not, then one concludes that I could have been qualitatively identical with Napoleon, and on counterpart theory, this simply means I could have been Napoleon. But the same kinds of *sorites*-based arguments can surely be run between any two physical objects (cf. Lewis, 1986a, Section 4.5).

If one is impressed by the analogy between the true *de dicto* modal axiom

($\hat{a}\hat{a}p$) \hat{P} ($\hat{a}p$)

and the *de re* claim

(x could have been (an entity that could have been y)) \hat{P} (x could have been y)

then one will accept this *de re* claim as a conceptual truth and hence one will take counterpart relations to be transitive. [71]

The Chisholmian arguments may now make one think that if one wants to stick to an objective non-contextual counterpart relation, which presumably will be transitive, one needs to allow that anything can be anything, which is paradoxical and leaves us without objective guidance vis-à-vis the normative applications discussed above. Hence, one should go for contextually-based counterpart relations, where the limited range of applicability of the relations makes it plausible to suppose that the relations are non-transitive.

But this argument fails when we consider the fact that there is a large range of things that I definitely could not have been. First of all, I could not have been a set or the number seven, but Lewis can actually account for this in his counterpart theory. The reason for this is that sets and numbers are necessary entities that Lewis allows to exist at every possible world and be numerically identical across world. The number seven of this world and the number seven of world w_{2481} are identical. Worlds overlap on these necessary entities. Thus, Lewis can analyze “ x could have been y ” where y is one of these necessary entities which lie in the overlap of all worlds into “ x is y ”.

However, recall some other examples I gave. On Lewis’s view, each world has a separate space-time manifold. Thus, points in space-time are not in the overlap of all worlds, nor are space-time manifolds themselves. There is no *sorites*-type of argument that shows that I could have been a point in space-time or a space-time manifold, because these things are of a radically different sort from me. There is no enmattered (a term I use to denote something of which matter is at least a part) entity that is sufficiently close to any entity that is a point in space-time or a manifold to count as its counterpart, and so the *sorites* cannot get off the ground. Of course, my

claim presupposes certain ontological statements. For instance, it presupposes that space-times have points, as opposed to space and time being merely relational.

If one accepts a trope theory on which entities have particular innate qualities such as “the negative unit chargedness of this electron”, in respect of which they may be objectively perfectly similar (this electron and that electron are perfectly similar in respect of charge), then these qualities will neither lie in the intersection of the domains of all worlds, nor be the sorts of things I could have, in any *reasonable* context, been (note that of course distinguishing which contexts for counterpart analysis are reasonable is a modal task). Moreover, I could not have been a socially-constituted entity, such as the Bank of England or the Swedish monarchy. While the Bank of England and the Swedish monarchy may be realized by particular enmattered objects, they are not identical with these objects. One way to see I could not have been a socially constituted entity is to observe that I am a person, and hence I am one of those beings which are prior in *logos* (to use an Aristotelian term) to society, and hence to all socially constituted entities. Of course, here, too, substantial ontological assumptions lurk, but Lewis’s theory should not prejudge things in their disfavor.

Neither could I be the law of gravitation. If one’s view of the laws of nature is that they are merely descriptive and supervene on the actual events of this world, then Lewis will have no problem with the claim that I could not be the law of gravitation. For, then, ontologically the law is nothing but some sort of a descriptive proposition—ontologically its lawlikeness consists in nothing over and above what happens in this world—and this proposition is a necessary being. However, because of the usual objections against the Humean view, I do not take this to be plausible. On the other hand, if one views the laws of nature as actual explanatorily relevant entities, then they are merely contingent entities, and so the Lewisian approach described above for *could-have-been* assertions involving entities in the overlap of all worlds fails here.

Finally, I could not be a particular act of jumping. I am the sort of thing that does the jumping, but not the sort of thing that is an act of jumping.

Now, Lewis could claim that all of these examples can be accommodated by some sort of stipulation on the counterpart relation based on categorial distinctions. Thus, what we would have is not primitive modality *as such* but primitive categoriality, which is less mysterious. However, Lewis’s account still fails to explain in a uniform way what it is that we mean when we say that being a member of some category is an essential property, a property one could not lack. On this stipulative story, being a member of a category is essential solely because we have *stipulated* that the counterpart relation not relate things across categories. But what is it that the modal claim that something is *essentially* a member of a category has in common with other modal claims that Lewis wants to explain non-stipulatively, such as that it is possible for unicorns to exist, and necessary that unicorns be members of a natural kind that has one horn? One would want an account of modality that does not involve piece-meal stipulation that explains some modal claims in terms of categorial difference and others in terms of quantification over possible worlds.

Nor are categorial distinctions the only sort of thing for which it is intelligible to seek a fact of the matter about their modal status. It is plausible that having electric charge is part of the essence of an electron—something that is an electron would not exist if it did not have electric charge. This is not a categorial distinction. It is not something we can legislate stipulatively. Lewis might disagree that electric charge is essential to an electron. But whether it is or not is a meaningful question, one to which one would think there is an objective non-stipulative answer. Nor is it at all obvious that a *sorites* argument that varies an electron between possible worlds is possible. True, one might say that e_1 is a counterpart of e_2 , where e_2 has slightly smaller charge, and continue working in this way until one gets to e_n which has no charge. But one can dispute that an electron-like entity that has some charge can be a counterpart to one that has no charge. Or one can claim that an electron *essentially* has *exactly* the charge it does. For another example, it is plausible that an electron could not have been the mereological sum of two electrons, and *a fortiori* could not have been a horse, since a horse contains many electrons.

Moreover, it is *prima facie* possible for there to be immaterial beings. And *prima facie* a material being

could not have been a immaterial being. This distinction is not amenable to *sorites*-type arguments, and is arguably not a categorial distinction—both material and immaterial beings are *substances*. Furthermore, one might argue controversially *pace* Freud that there is no *sorites*-type of transition between a conscious mental state and an unconscious mental state, and that being conscious is an essential property of a state—nothing unconscious can count as its counterpart.

Thus, Lewis's theory cannot account in a uniform way for essential properties without accepting a primitive modality encoded in the counterpart relation, a modality that on Lewisian grounds should be considered objectionable. The only uniform way to do this on Lewisian grounds would have been to distinguish some relation as *the* counterpart relation, while it would be unintelligible (without reference to some ontological prior modal facts) what it is about *that* relation that makes it *the* counterpart relation.

Before closing, let us, however, consider another argument of Lewis for the claim that the counterpart relation had better be contextual without there being anything that can be called "*the* counterpart relation."

The Great Western Railway ought to have absorbed two other railways early on: the Bristol and the Gloucester, and the Birmingham and Gloucester. But it tried to drive too hard a bargain. In 1845 the line from Bristol to Gloucester to Birmingham fell into rival hands. Therefore, after the grouping of railways in 1923, the post-grouping Great Western lacked a part that it might have had. What we know as the Great Western, without the missing line, was the whole of the Great Western; not, as it so easily might have been, a part of a still Greater Western. [\[72\]](#)

Here is the paradox. Let GWR- be the Great Western as it actual was without the missing line. Let GWR be the Great Western. Let GWR+ be the sum of the GWR- and the missing line. Here is GWR, in other words GWR-; they are identical. But the plural is a nonsense of grammar: 'they' are one thing, and *it* is self-identical. What might have happened to *it*? It is GWR; so it would have been greater, in fact it would have been identical to GWR+. It is GWR-; so it would have been only a part of GWR, not the whole, and hence not identical to GWR, which would instead have been identical to GWR+. Most certainly it, that is GWR-, would not have been identical to GWR+. We contradict ourselves about what would have become of

this one thing which we can refer to in two ways: as GWR or as GWR-. [\[73\]](#)

If we think that the counterpart relation is something objective, then we need to say that had the GWR absorbed the other lines, either *it* would have been identical to GWR+ or to GWR-, since there are no other options: there is nothing else in that counterfactual world which could play the role of what the GWR would have been.

Lewis emphasizes that when we are talking of the GWR and the GWR-, we are talking of *one thing*. However the problem is that they are not one thing. When we are talking of the GWR, we are talking of a *company*. But when we are talking of the GWR-, we are not talking of a company, but of an aggregate of *railway lines*, which is the company's property. To see that Lewis uses "GWR-" to talk of an aggregate of railway lines observe that his claim that the GWR- would have been only a part of the greater GWR+ entails that "GWR-" is not a company, since in that world in which the GWR is greater, there is no *company* whose lines correspond to GWR-.

This equivocation between a company and what it owns destroys the argument. For suppose first that we consistently use both "GWR" and "GWR-" to indicate companies, the former defined as "the Greater Western Railroad company" and the latter as "the Greater Western Railroad company without the extra lines." Then, in the counterfactual world where the GWR is greater, the term "GWR-" lacks reference. There is no paradox then in saying that that entity which the GWR- is in our world, namely that company that is the GWR without the extra lines, is identical with the GWR in the world where the GWR is greater: for, indeed, the company is increased then. There is no competitor for identity that is a company without the extra lines, since there is no such company there.

On the other hand, suppose we consistently use both "GWR" and "GWR-" to indicate aggregates of railway lines, the former defined as "the lines owned by the Great Western Railroad company" and the latter as "the lines owned by the Great Western Railroad company without the extra lines." Then, there is no paradox either. For that entity which the GWR is in this world, namely a certain aggregate of railway lines, is a proper part of what the term "GWR" refers to in the counterfactual world, indeed the same proper part that "GWR-" refers to in *either*

world. There is no paradox here, beyond the uninteresting observation that the term “GWR” can refer to different entities in different worlds, just as many other similar terms like “John’s favorite possession” can.

But perhaps Lewis will insist that there is no ambiguity. A railway company is nothing over and beyond the aggregate or mereological sum of the railway lines, with “line”, of course, not meaning just the physical rails, but also the staff and the rolling stock. We should not multiply entities, and so there is just one thing there. Thus, “GWR–” and “GWR” refer to the same entity in the actual world, but what counterpart this one entity has in another world will depend on one’s point of view: whether we think of it as or as “GWR”. There just is no non-contextual truth of the matter as to which of these is the right way to pick out this entity’s counterparts. And there may be *some* plausibility about such a claim, because in ordinary language we might well talk as if “GWR–” and “GWR” had the same referent.

I submit that the plausibility that this context-relativity claim has derives from the fact that the example is an artifact. In the case of an artifact, we bestow criteria identity on the artifact largely as we please. Whether the ship of Theseus is the same ship after a piece-by-piece replacement or not is up to us and determined by such social facts as how insurance companies conventionally treat cases of gradual replacement of parts. There is no problem with allowing Lewis that counterpart relations can be context-relative in the case of artifacts. However, that will not be so in the case of natural entities.

Consider a similar question for a natural object such as a dog. We can talk of a “Spot–”, which is the sum of the parts that Spot has actually has, without the extra toe that it might have grown were its foot genetically “treated”, and of a “Spot” who is just the dog. Then if we are to insist that there is a single entity that “Spot–” and “Spot” refer to, there will have to be an objective fact of the matter as to what the right counterpart to that entity is in a world where our dog does have an extra toe due to a genetic “treatment” by a mad scientist. The reason there will have to be such a fact is that either one can ontologically reduce dogs to the mereological sums of their parts (e.g., the way one can reduce Zeus’s arrows to lightning) or one cannot. Remember it was assumed that “Spot–” and “Spot” refer to the same entity. Now, if one *can* do such a reduction, then it is true to say with Democritus that there really are no dogs, except by human convention, and so reality itself only determines for the dog the counterpart relations that are had by the sum of the parts. Any counterpart relations that do not follow those of “Spot–” must be simply our projections since in reality Spot is *nothing but* Spot–.

But if one *cannot* reduce the entity that Spot is to the entity that Spot– is, and yet one insists that they are the same entity, then one must either allow them both to be reduced ontologically to some third item or allow Spot– to be reduced to Spot. Otherwise what could possibly be meant by saying there are the same entity? Now, if both can be reduced ontologically to some third item, their *objective* counterpart relations and identity criteria will be defined by that third item, and any other relations will be our subjective projections. And if Spot– can be reduced to Spot, then it is Spot’s counterpart relation and identity criteria that will count.

Going back to the GWR, if we allow that what we are after is not the socially-instituted counterpart relation but rather whatever *objective* counterpart relation is there, then we need to look at the natural objects out of which the GWR is constituted, people, bits of iron, and the like, and ask about the counterpart relations that the aggregate of these natural objects stands in. It is then more plausible to say that, considered as a natural object, the GWR– and GWR are both identical to GWR– in the other world. But considered as socially constituted, they should not be conflated into one object.

It is enough for an argument against Lewis that *some* counterpart relations be objectively determined. It is not necessary that they all be. Thus, socially constituted cases are not a problem.

And in fact there is a positive argument why on broadly Aristotelian grounds one would not think that socially constituted entities like the GWR should be the things about whose counterpart relations we talk, at least if we are going to be talking of objective ones. The reason for this is that for an Aristotelian, identity relations hold in the first instance between substances, and only in the second instance between items in the other categories. Now, when we talk of substances and their identity (or counterpart) relations, one might claim that we should refer to the

substances *intrinsically* or in what Aristotle might call a “*per se*” way of referring. A term x is an *intrinsic* reference if and only if:

- (20) It is possible that x exists, and
- (21) it is necessarily the case that if the proposition that x exists is true then the truthmaker of this proposition involves no entities other than the referent of x itself and the mereological parts of x .

Observe that the requirement of being an intrinsic reference immediately rules out many (though not all ^[74]) non-rigid references. “John’s favorite object” does not refer *intrinsically* because “John’s favorite object exists” is made true not just by John’s favorite object but also by John and his mental states. Intrinsicness rules out references to entities in categories other than of substance. “Green exists” (unless we are talking of a Platonic Form which is basically an “abstract” substance) is made true by *something’s* being green, and this something is distinct from the green that it has. Indeed, one might think that it is a necessary and sufficient condition for something’s being an Aristotelian substance that it be possible to refer to it by a term that is intrinsic and that has the property ^[75] that in no possible world does its referent consist of more than entity (e.g., of an aggregate of two entities). But at least it is clear that this is a *necessary* condition for being an Aristotelian substance.

One might argue on holistic grounds, of course, that the truthmaker of every true proposition, other than one logically equivalent to the conjunction of all true propositions, involves other things in existence, because the assertion that some entity presupposes a *context* within which one is speaking. In that case, the only terms that have intrinsic reference would have to be terms that refers to the actual mereological sum of everything that exists. The exclusion of this holistic claim is the broadly Aristotelian presupposition of this strategy for removing socially constituted entities like the GWR from the present discussion. The Kripkean will agree here. The truthmaker of “Smith exists” is just Smith. Although it is true that our ways of fixing the reference to Smith involve all kinds of things other than Smith (e.g., the person who gave Smith her name), nonetheless these things are not a part of the truthmaker of the proposition that Smith exists in the way in which the social structure necessary for the existence of railroads not only helps fix the reference of “the GWR” but also enters into the truthmaker of the proposition that the GWR exists.

If one accepts this Aristotelian presupposition, then, we might very plausibly then insist that when we are talking of the counterparts of some substance, we should only consider the substance under descriptions that refer to it intrinsically. But “GWR” does not refer intrinsically, because part of the truthmaker of “GWR exists” is the social conditions that define what a company is, what a company’s property is, etc. Were there no society other than at most the GWR, the GWR would not exist either, and the existence of society over and beyond the GWR is part of the truthmaker of “the GWR exists”. ^[76] And if we were to fix the references to be intrinsic, e.g., by giving a name to the aggregate of the parts of the GWR (which aggregate would not be a substance, but we might allow aggregates of substances to have objective counterparts induced by the counterparts of the substances, providing we have referred to the aggregates intrinsically), then there would be no paradox.

4.2.2 Other applications and assessment

We have seen in Section 4 of Part II that the use of possible worlds to explicate propositions and properties is not helpful. However, we have also seen that possible worlds can be used to clarify modal concepts, e.g., supervenience and counterfactuals, which are difficult to clarify without invoking possible worlds. Lewisian possible worlds are eminently suited for this clarificational task. Admittedly, as has been argued, the order of time cannot be analyzed completely into possible worlds, and hence neither can causality. But there are enough applications left that there is good *prima facie* reason to believe Lewis’s theory, unless there are some strong objections. But there in fact are *decisive* objections.

Section 5 Objections to Lewis’s account of actuality

5.1 Common sense

Normally we make no distinction between actuality and existence.^[77] We say interchangeably that no unicorn exists and that unicorns are non-actual. Lewis thus draws a distinction that common sense does not make. But of course, as Aristotle notes, the *hoi polloi* are not very good at making distinctions, and we philosophers must often introduce distinctions where there were none. This in no way counts against Lewis.

However, what does count against Lewis is that the distinction he makes is one that the language of educated English speakers is already capable of marking, and it is in fact *they* who see the possibility of a distinction where Lewis sees none. For, we can speak in English of something “being spatio-temporally related to us” if we are not afraid of latinates. And this is a concept that is different from our concept of actuality. Even if we think that as a matter of fact everything actual is spatio-temporally related to us (and many of us don’t: after all, many English speakers are theists of the sort who believe God is beyond space and time), we do not analyze actuality into such spatio-temporal relatedness. Lewis thus *drops* a distinction which our language has: the distinction between being spatio-temporally related to us and being actual.

It does not count as “introducing a distinction” when a person realigns three concepts *A*, *B* and *C* such that originally *B* and *C* were thought synonymous with *A* different from both whereas *A* and *B* are thought synonymous and *C* different. But it is precisely such a realignment that Lewis introduces between “being spatio-temporally related to us”, “being actual” and “being existent”. To realign the terms in this way is indeed to go against common usage, and is indeed a consideration weighing against EMR.

Lewis says that a “spokesman for common sense” would affirm three theses:

- (22) “Everything is actual.”
- (23) “Actuality consists of everything that is spatiotemporally related to us, and nothing more (give or take some ‘abstract entities’). It is not vastly bigger, or less unified, than we are accustomed to think.”
- (24) “Possibilities are not parts of actuality, they are alternatives to it.”^[78]

Lewis then goes on to say that his critics take (22) to be analytic and (23) to be “up for grabs.” However, Lewis writes:

[T]he two theses, indeed all three, are on an equal footing. Together they fix the meaning of ‘actual’, but they go far beyond just fixing meanings. I don’t see any evidence that the analyticity is concentrated more in some of them than in others.^[79]

Therefore, Lewis does not think he goes against common sense any more by violating (22) than he would by violating (23). Presumably the point here is that if the alternative to Lewis’s view of actuality is one on which the whole “paradise” of EMR with all possible worlds is seen as actual, then (23) would be violated just as unacceptably. But even if this were so, the serious alternatives to Lewis’s account are not in violation of (23). The serious alternatives keep all three theses in place (more or less—most people believe in non-spatial entities like deities and ghosts, and some believe in a God who is not spatio-temporally related to anything), but replace the “paradise” of EMR with some other account.

Moreover, it is worth noting that the evidence of common sense for (23), which Lewis so much wishes to respect, is no greater than the evidence of common sense for:

- (25) *Existence* consists of everything that is spatiotemporally related to us, and nothing more (give or take some ‘abstract entities’). It is not vastly bigger, or less unified, than we are accustomed to think.

And *this* Lewis rejects. Since common sense affirms *both* (23) and (25) with equal force, indeed seeing them as equivalent, and since Lewis must go against (25), his acceptance of (23) is not that much comfort. It would be best if we could keep all four theses, at least to some extent.

Another commonsensical objection to Lewis’s account is that it has the absurd consequence that there are

people with whom we do not disagree, though they believe propositions we disbelieve, and it is logically possible for one to be in the position that there are people with whom a person disagrees even though they believe all the same propositions that she does.

For the first of these observations, observe that if Smith in some other world believes that there actually are unicorns and I believe that there actually are no unicorns, then the proposition Smith believes is the negation of the proposition I believe. (This is particularly clear on Lewis's account of unstructured propositions, since the proposition I believe is the collection of all worlds that do not contain unicorns, and the proposition Smith believes is the complement of that collection of worlds, but my claim does not depend on this account of propositions). But Smith and I do not disagree. For if I disagree with someone, I am claiming that there is something wrong with her doxastic system, or, in different terminology, that this doxastic system has failed to achieve its *telos*, which of course is truth. Moreover, a virtuous person tends to be saddened by disagreements, taking the disagreement as an indicator that either she or the other person is in a doxastic state that is unfortunate. But none of this need apply if I believe that there are no unicorns and Smith believes the negation of this proposition. It is in fact quite possible that both my and Smith's doxastic faculties have achieved their *telé*. It is quite possible that there is nothing unfortunate about either Smith's or my doxastic state, nothing to be sad about. For, Smith's world might contain unicorns while mine does not, in which case neither of us is in a bad doxastic state, while yet one of us believes something the other disbelieves.

For the second observation, suppose that there is some property F which is uniquely instantiated by the actual world and which a person believes to thus be uniquely instantiated. It might in actual fact be that there is no such property, but it is surely logically possible for there to be such a property. After all, it seems logically possible for someone with a particularly advanced mind to live in a particularly simple world and thus know the conjunction of all propositions true in that world. Then, this person, call her Jones, believes that F is uniquely instantiated by the actual world. But now consider another person, Smith, in another world who also believes that F is uniquely instantiated by the actual world. They believe the same proposition (a proposition that on Lewis's account of unstructured propositions is identical with the singleton set consisting of Jones's world). It might even be that Jones and Smith believe all the same propositions. But they *disagree*. For Jones takes *her* world to be the one that uniquely instantiates F and Smith takes *her* world to have that property.

5.2 Indexicality and ordinary language

Richard Gale (1996, chapter 5) has raised some powerful objections to Lewis's view that actuality is indexical. It is natural to think that if actuality is indexical, then "the actual world" and cognate terms can be paraphrased by some explicitly indexical expressions.

5.2.1 "The actual world" means "the world in which I exist"

One attempt at a paraphrase, supported by some Lewisian texts (cf. Lewis, 1986a, Section 1.9), is to define "the actual world" as "the world in which I exist", and to say that something is actual providing it is in the actual world. Observe first with Gale that this approach requires counterpart theory, since otherwise several worlds will be such that I exist in them. And even given counterpart theory, Gale makes short work of the suggestion. After all, it is evidently possible that:

(26) I do not exist in the actual world,

while it is impossible that:

(27) I do not exist in the world in which I exist.

Therefore, (26) and (27) cannot be synonymous.

Nor will it do to define "the actual world" in the context of some sentence as "the world in which this tokening of this sentence occurs". For, evidently, it is possible that no sentences are tokened in the actual world, and yet it is impossible that no sentences are tokened in the world in which this tokening of this sentence occurs.

5.2.2 “The actual world” means “the universe that contains *this*”

Gale also notes that demonstrative renderings of “the actual world” into “the universe that contains *this*” or “*this* world” fail. Demonstratives are rigid, while “the actual world” is non-rigid. First, observe that if we allow ourselves *de re* modalities, the rigidity of demonstratives makes true the *de re* modal claim:

(28) This world (or the universe that contains this) could not have failed to be this world.

But, since there is contingency,

(29) The actual world could have failed to be the actual world.

Given (28) and (29), the demonstrative definitions fail.

If *de re* modalities are not allowed, one can use Gale’s next example. Someone can disbelieve that the actual world is the universe that contains *this* while not at all disbelieving the tautology that the actual world is the actual world.

5.2.3 Gale’s general argument

Gale also gives a general argument for why should think that actuality is not indexical. ^[80] Tokenings of “Now is *t*” at different times express different propositions if *t* is a rigid designator. But tokenings of “The actual world is *w*” express the same proposition whichever world they are tokened in if *w* is a rigid designator. This even follows on Lewis’s own theory of unstructured propositions. The proposition expressed by a tokening of “Now is *t*” at t_0 consists of the set of all possible worlds at which the time that is counterpart to t_0 is the time that is counterpart to *t*. This set will clearly change as the tokening time t_0 changes. Indeed, when t_0 differs from *t*, the set may even

be empty. ^[81] However, the proposition expressed by “The actual world is *w*” is, for Lewis, the singleton set $\{w\}$, whichever world “The actual world is *w*” is tokened in.

The argument goes through even if one is sceptical of rigid designators. Let *F* and *G* be any purely non-indexical predicates. Then, “Now is *F*” expresses different propositions at different times, while “The actual world is *G*” always expresses the same proposition.

We have seen in Section 4.2 of Part I that Aristotle held a view according to which indexical propositions involving “now” express the same proposition at different times, a proposition whose truth value is liable to change. This Aristotelian usage of “now” is analogous to the Lewisian usage of “actual”. But Lewis cannot paper over the damaging disanalogy between actuality and ordinary indexicals in this way, because Aristotle’s analysis of “now” is patently incompatible with Lewis’s view of propositions as collections of possible worlds. For if *p* is a proposition, then *p* is true at t_1 if and only if the actual world is a member of *p* at t_1 . But since for set-like collections (Lewis actually talks of “sets”, but we shall see in Section 7.2 that this is impossible) the membership does not change with time—indeed, set-like collections have their members essentially—the truth value of *p* cannot change with time.

One could, however, fix up Lewis’s theory to make it more Aristotelian and to avoid the fatal disanalogy between “now” and “actual”. Instead of considering propositions to be collections of worlds, Lewis can consider them to be collections of time-slices of worlds, viz., of those time-slices of worlds at which they are true. On this view, both “Now is *F*” and “The actual world is *G*” express the same proposition whatever time-slice of whatever world they are tokened on. So Lewis does appear to have a way out of Gale’s argument.

But does this solution not simply relocate the problem? The point of the observation that “Now is *F*” expresses different propositions at different times was that the words “Now is *F*” are semantically incomplete in the sense that to figure out what truth value to assign to the claim that they express, in addition to knowing English, we need the additional information of what time the worlds were uttered. This kind of semantic incompleteness is essential to indexicality. But “The actual world is *G*” is not semantically incomplete in this way. We can know whether it is true or false without knowing by whom, when, or where it was uttered.

Lewis can, however, reanalyze the semantic incompleteness of “Now is *F*” in the following way. To tell whether some utterance of “Now is *F*” was right *in its context*, we need to know when this was uttered. But this is

quite analogous to the fact that, given EMR, to know whether some utterance of “The actual world is G ” was right *in its context*, we need to know which world this utterance was made in. The only reason we feel as if “The actual world is G ” is semantically complete is that we are used to evaluating the rightness of utterances made in the actual world. If, on the other hand, we wish to tell whether the utterance “The actual world is G ” would be right in *our* present context, then of course there is no incompleteness. But likewise, there is no incompleteness should we wish to tell whether the utterance “Now is F ” would be right in our present context.

So there is no asymmetry if we are willing to abandon the notion of evaluating utterances for truth or falsehood *simpliciter* and settle for evaluating their truth or falsehood relative to the circumstances of utterance, while at the same time taking propositions to vary their truth value with time—itsself a rather dubious position. [\[82\]](#)

Note, too, that this Lewisian reply presupposes EMR, and hence the conclusion that the Lewisian analysis of actuality is revisionary vis-à-vis ordinary language is unshaken as ordinary language does not recognize the concrete existence of other worlds. Moreover, Gale’s disanalogy performs the valuable task of showing that one cannot conclude that actuality is indexical on the basis of an analysis of ordinary language.

5.2.4 Nor can one simply substitute

However, even if there is no obviously indexical paraphrase, perhaps there is another sense in which actuality can be taken to be indexical. Rather than trying to paraphrase “the actual world” by a single phrase, we might simply explain how this term behaves semantically, context by context.

To do this, first paraphrase all modal statements into quantifications over possible worlds. E.g., “It is possible that p ” is paraphrased into “ $\exists w (p \text{ is true at } w)$ ”. It is important here not to leave out any implicitly modal operators unparaphrased (e.g., we must take care to translate “ p entails q ” into “ $\forall w ((p \text{ É } q) \text{ at } w)$ ”). We now proceed inductively, limiting ourselves for simplicity to sentences not involving any propositional-attitude or modal verbs, or any cognates of “the actual world” other than “the actual world” itself. Suppose s is a sentence where all modalities were paraphrased into quantifications over possible worlds as above. Then, those occurrences of “the actual world” that occur outside the scope of the left hand sides of any “... is true at ...” operators are to be replaced with “*this* world”. In the case of the remaining occurrences, replace each “the actual world” by “ w ” where “... is true at w ” is the *innermost* “... is true at ...” operator in the scope of whose left hand side the given occurrence of “the actual world” is found. Thus,

“(((The actual world contains unicorns is true at w) and horses exist in the actual world) is true at v) and z is the actual world”

is to be paraphrased into:

“(((w contains unicorns is true at w) and horses exist in v) is true at v) and z is *this* world.”

If this paraphrase works, then we have sense in which “the actual world” is indexical, namely that “the actual world” is a variable symbol that picks up different objects depending on context.

It is easy to convince oneself that this approach works *salva veritate*. The paraphrase has the unfortunate property, however, that if O is a modal sentential operator, s a sentence, Tr the translation of a sentence r according to the above scheme, then $T(O(Ts))$ and $T(Os)$ can have different truth values. To see this, let s be the sentence “The actual world contains horses”. Then, Ts is “*This* world contains horses”. If O is δ , then in fact $O(Ts)$ is true, because the rigidity of “*this*” ensures that it is true at every world that *this* world contains horses. But then likewise $T(O(Ts))$ is true because T preserves truth value. But, O_s is “Necessarily the actual world contains horses”, which is false, and hence $T(O_s)$ is likewise false.

This illustrates the problem that although there are ways of translating “the actual world” that work at the level of an isolated sentence, one cannot employ the scheme on subsentences *salva veritate*. If one thinks, as one should, that a necessary criterion for T to count as a paraphrase is that it behave well under sentential operators thereby ensuring that $T(O(Ts))$ and $T(Os)$ are paraphrases of one another, it is difficult to see how we can possibly claim that there is a way of translating “the actual world” into indexical-type expressions.

It is for these reasons that one should not characterize Lewis's view as saying that "actuality" is to be translated as an indexical expression. Rather, I characterize Lewis's view as making an ontological claim that I sometimes abbreviate with a phrase like "actuality is nothing but an indexical property". By this I mean that any sentence involving "the actual world" and cognates, while perhaps not paraphrasable into sentences involving just indexical operators, can nonetheless be given *truth conditions* that do not involve any occurrences of "the actual world" or its cognates, but instead either substitute demonstratives or variable letters coming from "... is true at *w*" operators. This situation is not in any obvious way absurd. Richard Gale has argued in conversation that in the theory of time, the B-theorist's best bet is to claim that A-propositions can be given truth conditions expressed in terms of B-propositions, even if there is no possibility of paraphrase. If one further thinks that the B-propositions are ontologically prior, then this might count as an ontological reduction. We can take Lewis to be making a similar claim about "the actual world" and cognates.

More weakly, one could at least say that Lewis is claiming that one could give complete instructions for when and in what world locutions involving "the actual world", with the instructions not involving the term "actual" or any non-relativized cognate such as "actually true". This is parallel to the way that I can use B-language to give instructions to completely specify when the expression of an A-proposition is appropriate. I am most grateful to Robert Brandom for this suggestion.

Section 6 The possibility of non-spatio-temporally related co-actual entities

A consequence of Lewis's view that worlds are universes, i.e., maximal temporo-spatially related aggregates, is that it is necessarily true that all entities are temporo-spatially related. For indeed, any two entities existing at a world are thus related if worlds are universes. Concerning this, Lewis (1986a) writes:

A first, and simplest, objection is that a world might possibly consist of two or more completely disconnected spacetimes ... and one world with two disconnected spacetimes is a counterexample against my proposal. Against this objection, I must simply deny the premise. I would rather not; I admit some inclination to agree with it. But it seems to me that it is no central part of our modal thinking, and not a consequence of any interesting general principle about what is possible. So it is negotiable. (p. 71)

However, it is not clear that this is negotiable if one is to hold that EMR is intelligible. For let us start off from an ordinary pre-EMR understanding of modality. Either two completely disconnected spacetimes will be intelligible to us or not. If it is intelligible, then it will be fair to demand that a theory of possibility should allow this possibility to be accounted for. But if they are unintelligible, then EMR itself will be unintelligible. After all, EMR crucially depends on the idea that distinct worlds are completely disconnected spacetimes. If EMR is intelligible, then such spacetimes are intelligible.

This argument appears to commit the fallacy of equivocation. Lewis distinguishes between actuality and existence, and will say that although it is intelligible that two completely disconnected spacetimes *exist* it is unintelligible that they be both *actual*. However, ordinary modal thought makes no such distinction, and consequently it finds two existent completely disconnected spacetimes intelligible if and only if it finds two actual completely disconnected spacetimes intelligible. Of the two horns in my dilemma argument, Lewis cannot possibly accept the horn that according to ordinary modal thought two completely disconnected spacetimes are unintelligible, for then someone who subscribes to ordinary modal thought would simply fail to understand EMR and it would be irrational for him to become an extreme modal realist. Thus, Lewis must insist that although ordinary modal intuitions definitely *do* allow for the possibility of completely disconnected spacetimes, nonetheless these intuitions are wrong. We know, Lewis can say, that they are wrong because the best theory of modality that we have, namely EMR, is incompatible with them.

While this incompatibility does to some degree count against EMR, it is not as awful as one might think at first sight. For, Lewis need not consider the modal statements of a believer in the possibility of completely disconnected spacetimes to be nonsense. He can reinterpret them within his framework by saying that in the context

of the interpretation of modal claims made by such a person, modal operators are to be understood as quantifications over all possible mereological sum of universes/worlds.^[83] Thus, “It is possible that there be a unicorn and a griffon that are not spatio-temporally related” is made true by the mereological sum of a world containing a unicorn and a distinct world containing a griffon. This may not give one everything one wants, but it helps.

What is more serious, however, is Lewis’s inability to fit into his theory entities that are not spatio-temporal at all. Lewis (1986a) himself discusses the case of spirits and comments that perhaps he does not need “to defend the possibility of spirit tales—after all, people have been known to accept impossible theories, as witness naive set theory” (p. 73). But surely there is a presumption of possibility when dealing with statements in a well-established language game where the statements have not been shown, and do not appear to be in any imminent danger of being shown, to be incoherent. Lewis, however, has a better answer than just denial of the possibility of spirits. After all, the spirits will be temporally related, which is a special case of spatio-temporal relation. It’s not as tight a relation as exists between material objects, but it is good enough for binding the spirits into worldmates.

Presumably the sorts of temporal relations that Lewis is thinking of between ghosts are things like *later than*, *earlier than*, and *simultaneous with*.^[84] But why does it make sense to talk of non-spatial entities as standing in such relations? Is it any less mysterious that two ghosts are temporally related than to say that they are worldmates in some primitive sense? We can easily make sense of what it means for ordinary spatio-temporal objects to be spatio-temporally related. Objects occupying a single spatio-temporal continuum can be brought together in contact with one another, made to interact with one another. Apart from this possibility, it is not at all clear that it is intelligible to suppose that they *are* spatio-temporally related. Of course, likewise, non-spatial objects like ghosts can be conceived of as interacting, too. One ghost might in some way communicate with another, for instance. And if the pattern of causal interaction is tightly enough knit together, this might suffice for the intelligibility of the claim that they inhabit one time.

I claim that in fact it is the capacity for causal interconnection rather than occupation of a single temporal order that is more plausibly considered as rendering two ghosts worldmates. Imagine two non-spatial ghosts between whom there is no (nomically) possible causal interconnection.^[85] What does it mean that some event that happens to one is earlier than some event that happens to another? Is it any less mysterious to say that they occupy one temporal sequence than to say they occupy one world?^[86] If we do not think the idea of a primitive worldmate relation is acceptable, neither should we accept a primitive co-temporality relation.

Unfortunately, the suggestion that the capacity for causal interconnection is what makes two ghosts occupy the same temporal continuum is one that Lewis cannot adopt. For, firstly, it is highly plausible that *unless* there necessarily is a single creator to whom all things are necessarily causally interrelated, something that Lewis does not admit, there *could* be two ghosts that are not capable of causal interrelation. But this is a possibility Lewis cannot admit if the capacity for causal interconnection is what makes two ghosts live in the same world, since given the possibility of there being two ghosts who are not causally interrelated, there is a single possible world at which there exist two ghosts that are not capable of causal interrelation. Moreover, if this suggestion were taken, then possible worlds could not serve as the logically primitive basis of an analysis of modal language. For, the boundaries between possible worlds would depend on causal *capacities*, and capacities are modal. Nor is there any plausibility at all in the suggestion that it is actual causal interrelation that makes two entities worldmates (unless, again, there is a necessarily existing creator). For no one can deny that there could be entities unproblematically existing in one world which *in fact* do not interact, but could.

If Lewis is to allow for non-spatial entities, then it does not appear that mere temporal relatedness gives any benefit over a primitive worldmate relation. If the latter is unacceptable, so is the former. But the latter *is* unacceptable, because we have no way picking out which relation, of all the infinitely many relations between

individuals, the worldmate relation is. Suppose that Casper and the Ghost of Christmas Present (GCP) are two causally interrelated ghosts, who are themselves causally unrelated to anything else. That they are causally related uncontroversially ensures that they are part of one world. But now suppose that there is some other individual, call him Smith, in the world which Casper and the GCP inhabit. What makes Casper and the GCP worldmates of Smith is that the worldmate relation, R , obtains between them and him. But how are we to pick out *which* relation between individuals we mean by “the worldmate relation”? Let Jones be an individual in a world different from Smith’s. Define the relation R^* between individuals as follows. If x and y are both distinct from Casper and the GCP, then xR^*y holds if and only if xRy holds. However, if x is Casper or the GCP and y is distinct from them, then xRy holds if and only if y is R -related to Jones, while yR^*x holds if and only if xR^*y holds. Finally, if both x and y are chosen from the set { Casper, the GCP }, then xR^*y . It is easy to check that R^* is a transitive reflexive symmetric relation if R is. But then what is it about R that makes *it* rather than R^* “the worldmate relation” and thus make Casper and the GCP worldmates of Smith rather than of Jones?

It *does* matter which relation we single out for being “the worldmate relation” since some relations we could single out might give the wrong answers to modal questions. In a Lewisian paradise of all existent individuals, there is a transitive reflexive symmetric relation R^* respecting causal interrelation (so that causally interrelated things, or even things capable of causal interrelation, are related by this relation) which puts all the individuals that are ghosts who cannot be causally related to the material objects in a single world that, as a matter of fact, has no unicorns. But with respect to R^* as the worldmate relation, it is then impossible for there to be a ghost and a unicorn that cannot be causally interrelated, whereas with respect to another relation, this is possible.

Nor can we say that we should fix a worldmate relation that gives the right answers to modal questions. For one of the main points of EMR was to give a realistic *analysis* of modal propositions, and if we fixed a worldmate relation in this way, we would be proceeding circularly.

It does not, thus, appear possible for Lewis satisfactorily to accommodate non-spatial concrete entities in his system. (Abstracta are a different matter: these exist in all possible worlds and are worldmates to everyone.) But such entities are *prima facie* possible, and Lewis has not given any argument against their possibility.

Section 7 Cardinality and the “set” of all possible worlds

Section 7 of Part III is largely taken from Alexander R. Pruss, “The Cardinality Objection to David Lewis’s Modal Realism”, *Philosophical Studies*, 104 (2) (2001) 167-176. For copyright reasons it is not available online.

7.1 Introduction

7.2 There is no set of all possible worlds

7.3 Lewis’s proviso

7.4 What should Lewis do about this?

Section 8 Ethical issues

8.1 The ethical objection

Robert M. Adams (1974, pp. 215–216) has objected to EMR on ethical grounds:

We may be moved by the joys and sorrows of a character known to be fictitious; but we do not really believe it is bad that evils occur in a nonactual possible world, or good that joys occur in a nonactual possible world, though of course it would be bad and good, respectively, for them to be actual. I think that our very strong disapproval of the deliberate actualizing of evils similarly reflects a belief in the absolutely, and not just relatively, special status of the actual as such. Indeed, if we ask, “What is wrong with actualizing evils, since they will occur in some other possible world anyway if they don’t occur in this one?”, I doubt that the indexical theory can provide an answer which will be completely satisfying ethically.

A similar objection was raised by D. C. Williams in regard to why one should strive to get rid of evils, since the evils one has pushed out of this world just reappear in another. [\[87\]](#)

Lewis, could, of course adopt a revisionary view of ethics that accepts Adams' and Williams' allegations. However, instead he claims that far from demolishing ordinary morality: "If my modal realism has any bearing at all on matters of value and morality, it pushes me towards common sense, not away" (Lewis, 1986a, p. 128). I shall argue that while Lewis's arguments do succeed in part in countering Adams' and Williams' charges, and thus allow one to save a *part* of common sense morality, nonetheless there are other important parts of morality that EMR would undermine. If one takes, as I do, these parts of morality to be grounded in objective moral truths, then one must take Lewis's EMR to be false.

8.2 The indexicality of morality

Consider the following case, which I give on behalf of Adams:

- (30) It is possible for me to murder Mr. Smith. Therefore, if I do not murder Mr. Smith, then a counterpart of mine in another possible world will murder the counterpart of Mr. Smith. Having just heard of Lewis's extreme modal realism, I murder Mr. Smith, reasoning that in another possible world—just as *real* as ours—a counterpart of Mr. Smith will die if I do not murder Mr. Smith, and hence by murdering Mr. Smith I do not cause any *overall* decrease in the sum total of good and evil since the life saved in the actual world will be lost in another possible world.

Lewis (1986a, p. 123) agrees that in such a case "[t]here would indeed be the same sum total" of good and evil. He does not deny that the evils in the other worlds are every bit as *real* as those in ours, though they lack the indexical property of being *actual*, i.e., of being in the same world as I. However he says:

If you actualise evils, you will be an evil-doer, a causal source of evil. That is something which, if you are virtuous, you do not want to be. Otherworldly evils are neither here nor there. They aren't your evils. Your virtuous desire to do good and not evil has nothing to do with the sum total of good and evil throughout reality. It has to do with what befalls you and your worldmates, and in particular it has to do with the way in which what befalls yourself and others depends causally on what you do.

For those of us who think of morality in terms of virtue and honour, desert and respect and esteem, loyalties and affections and solidarity, the other-worldly evils should not seem even momentarily relevant to morality. Of course our moral aims are egocentric. And likewise all the more for those who think of morality in terms of rules, rights, and duties; or as obedience to the will of God. (Lewis, 1986a, p. 127)

There seem to be two distinct replies here. The first reply insists on moral aims being "egocentric" and on morality being indexed to the moral agent. Moral obligations are indexical and addressed to the individual: "*Thou* shalt not kill." Indeed, the indexical "you" and its cognates appear ten times in the first paragraph quoted from Lewis. The second reply implicit in the paragraph insists on the importance of virtues. By murdering Mr. Smith, I make myself into a murderer. I neglect the obligation of respect for the life of Mr. Smith. In doing so, I deprive myself of virtue. And this is bad.

However, the two replies are actually not distinct. For the importance of virtues is also an *indexical* importance for Lewis. My primary task is to promote *my* virtue: "our moral aims are egocentric." Lewis has to claim this, since as Richard M. Gale has pointed out in conversation, if I do not murder Mr. Smith and thereby do not engender in myself the vice of murderousness, then my counterpart in another world will murder the counterpart of Mr. Smith, and the vice of murderousness will be engendered in that counterpart and so I could look at my deed of murdering Mr. Smith as a selfless sacrifice of my own virtues for the sake of allowing my counterpart to retain these virtues. Put so pointedly, it seems as if murdering Mr. Smith can be a very good thing to do, since by doing so I am saving another human being—my counterpart—from decay of moral character, at great personal cost, and I am not increasing the total amount of evil in any way since, although now Mr. Smith dies, still his counterpart would have died had I not murdered Mr. Smith. Moreover, Gale has noted that not murdering Mr. Smith violates the

deontic principle that it is impermissible to do anything that entails that something impermissible is done. ^[88] For, not killing Mr. Smith entails that my counterpart does something impermissible!

However, this argument against Lewis fails. First of all, Lewis can reject the deontic principle invoked by Gale. After all, why should we think that the principle holds true? In ordinary life, given free will, *x*'s action ^[89] never entails that *someone else* has done something impermissible. At most it entails that *one* has done something impermissible. So any intuitive support that ordinary life can give to Gale's deontic principle will be just as compatible with the deontic principle that you cannot do anything that entails that *you* do something impermissible. But since your counterpart is not *you* but merely your *counterpart*, this more limited principle is not violated by refraining from murdering Mr. Smith. Moreover, Gale's deontic principle begs the question against Lewis's view by making it too *straightforwardly* obvious that no action will be permissible if the principle is accepted and if Lewisian modal realism is true. For, with the principle understood as quantifying across possible worlds (as it must be for it to have the kind of application that Gale envisions for it), since it is a necessary truth that some world contains an impermissible action, and since every proposition entails every necessary truth, it follows that every doing entails that an impermissible action is done.

As for the fact that the total amount of virtue remains the same if I kill Mr. Smith, this only shows that Lewis's two replies were not independent. Lewis can simply insist that the obligations of morality *are* indexical and it is one's primary duty to nurture *one's own* virtue. Nor is the argument that one does something good and selfless by murdering Mr. Smith because one sacrifices the virtue of one's own self for the sake of the virtue of one's counterpart's virtue successful. For, one may not violate *one's own* moral duties. Nor do we need multiple-world scenarios to get the structure of the case:

- (31) You are imprisoned by a tyrant. The tyrant has a thousand other innocent prisoners. He brings you to a room where another innocent prisoner is tied up. He hands you a syringe with a lethal liquid and tells you that unless you murder the other prisoner, you will be executed, and then the offer made to you will be extended to another prisoner, going through the 999 prisoners (other than yourself and the one tied up), executing those that refuse to murder and stopping only when someone actually kills the prisoner who is tied up. And if no one murders the prisoner, then the tyrant will do it himself. You know that the tyrant means what he says—whatever his vices, he always tells the truth.

The structure of (34) with regard to the promotion of virtue is not very different from that of (33). For, if you do not murder the prisoner, then there is an extremely high chance that one of the other prisoners will. After all, it is highly unlikely that *none* of the 999 prisoners will fall for the temptation. Can you then reason that by killing the prisoner yourself, you are not only ensuring there is less bloodshed overall, but you are also taking the guilt and loss of virtue upon yourself, thereby selflessly saving another prisoner from moral downfall? A deontologist, and many people's intuition, will answer in the negative to this question, too. Moral rules are inexorably indexical. On such a view, in general, in the *first* instance each of us is obliged to obey those moral rules that apply to her, and not

maximize the number of people who obey moral rules or anything like that. ^[90] In the first instance the traditional command is: "*Thou shalt love thy neighbor as thyself*", and only at best secondarily: "*Thou shalt maximize the sum total of love in the world.*"

This deontological answer to (33) is controversial. However, as the completely intraworldly story (34) shows, this controversy has nothing to do with the nature of Lewis's possible worlds story. Lewis's view of (33) fits well with a respectable deontological view and moral intuitions that many people do share. In fact as Turner (forthcoming) notes, a theory like Lewis's EMR is in a somewhat better position than the standard deontological analysis of (34), since EMR does not even conflict with utilitarian intuitions in the way that the standard deontological analysis of (34) might. For if EMR is true, one's fulfillment of one's duties will never result in lesser overall utility, since the sum total of all possible worlds is fixed by EMR.

The challenge for Adams and Williams is now to come up with a clear-cut moral objection to David Lewis's view without relying on consequentialistic insights which Lewis expressly rejects.^[91] Such an objection will have to take on board the force of David Lewis's observations of the essential indexicality of morality, the apportionment of moral duties to each individual *I*. I shall attempt at least a partial fulfillment of this challenge in the next section.

But before doing this, observe that if instead of counterpart theory we accept identity theory, as one may argue (see Section 2, above) should be done, the Lewisian reply crumbles. For then by murdering Mr. Smith in this world I ensure that *I* do not murder him in another, and my not murdering him here entails that *I* do something impermissible (which would violate a rather more reasonable form of Gale's deontic entailment principle). Lewis's indexicality of morality reply is much less compelling in this case. It is still true that "[i]f you actualise evils, you will be an evil-doer, a causal source of evil", but it is also true that if you do *not* actualize evils in this world, *you* will still be an evil-doer, a causal source of evil in another existent world in which *you* exist. Therefore, even if the moral objection in this section of the paper fails against Lewis's considered view, it damages a "transworld identity" variant of his theory, which for the reasons discussed in Section 2.2 is the only tenable version of the theory.

Of course, in reply a "transworld identity Lewisian" might claim that morality is not only indexical on the "I", but also on "now" and "here (in *this* world)". However, even if this might have some plausibility in the deontological case of avoiding murder (e.g., it might be that one has an obligation to refrain from murdering *now* even if one should somehow know^[92] that one will commit it later if and only if one refrains from it now), it has less plausibility for other virtues. Must one strive hard to improve some virtue if one knows that this will result in one not having that virtue at some point in the future or in another world? That would seem a little questionable.

Hence, even though Lewis can handle some deontological type and virtue-centered cases in *his* extreme modal realism, a "transworld identity" variant would have difficulties with them.^[93]

8.3 Problems for counterpart theory

8.3.1 A transworld counterfactual

The objection in Section 8.2 worked for an identity variant of EMR, but failed for counterpart theory. I will now formulate a complex of logically possible ethical counterexamples to counterpart theory. All of these counterexamples will be built out of the following simple counterexample to Lewis's claim that there are no transworld counterfactuals (cf. Lewis, 1986a, p. 78).

Imagine a possible world where the laws of nature and instantial conditions allow for one and only one free *and* indeterministic choice, everything else being deterministic and there being no other choices (deterministic or not) nomically possible. At time *t*, I get to decide whether to freely perform action *A* or whether to freely refrain from *A*. There are two variant worlds involved here. Let w_P be the variant world in which my counterpart does perform *A*, and let w_R be the variant world in which my counterpart refrains from *A*. I will say that one world "matches" another if it shares the same instantial conditions and laws of nature.

Then, the following subjunctive conditional is true at w_P on standard Lewisian semantics:

- (32) Were I to refrain from *A*, my counterpart in the other world that matches the actual world ("other matching world" for short) would perform *A*.^[94]

For, to evaluate (35) at w_P , we look at the closest world to w_P at which the antecedent holds, which is nothing else than w_R . But at w_R it is true that "my counterpart in the other matching world would perform *A*" because it is true at w_R that "the other matching world is w_P ", and my counterpart *does* perform *A* there.^[95]

8.3.2 The first case: saving a life

Now that the basic situation has been set up, we can formulate two specific ethical cases in this setting. The first case is that *A* is the trivial action of e-mailing a petition note to a dictator saying “Free Smith!” which I somehow know would result in Smith, an innocent stranger sentenced to death by a dictator on another continent, being freed. Should my note not be sent, Smith will die; for some reason, no one but I is in a position to send such a note.

Being a Lewisian for the nonce, I deliberate over the decision as follows. Were I to send the note, a dictator’s innocent prisoner would be spared. Were I *not* to send the note, however, my counterpart in the other matching world would send the note, by (35), and hence it would be true that in the other matching world the counterpart of the innocent prisoner would be spared. What should I do? Whatever I do, somebody dies. Admittedly, in one case, a worldmate of mine dies, and in the other case someone in another world dies. However, this worldmate is a stranger to me. I know no more about him than I do about his counterpart. Admittedly, he is spatio-temporally related to me (perhaps, though, very distantly—we can change the case so he be in another galaxy and in the far future), while the counterpart is not, but why should mere spatio-temporal relatedness carry any ethical oomph?

Lewis’s answer, presumably, will be that if I send the letter, *I* am causally responsible for saving Smith’s life. This, however, will not do if we analyze causation in a Lewisian manner in terms of counterfactuals. For it is just as much true that if I do *not* send the letter, I cause it to be the case that the counterpart of the prisoner in the other matching world lives. For suppose that I refrain from sending the letter. Then, the counterfactual “Were I to have performed *A*, my counterpart in the other matching world would have refrained from *A*” is true by the reasoning behind (35). But then it is true that “Were I to have performed *A*, the counterpart of the prisoner in the other matching world would have lived.” And this counterfactual relation is all there is to causation on Lewis’s account. Hence, on Lewis’s account, if I do send the letter, I cause the prisoner in the actual world to live, and if I do not send the letter I cause the counterpart of the prisoner in the other matching world to live. Since it seems to make no difference, I might as well take the easier course, that of not sending the e-mail. But in doing so, I go against ordinary morality, which says that in the situation as set up, if a trivial action of mine can save a person in the actual world—the only concretely existing world ordinary morality knows of—then the action is *prima facie* obligatory. Nor does ordinary morality care that I am working with a case where the choice I make is the only free choice ever made in the history of the world.

One might think that there is a *de dicto* / *de re* modal fallacy lurking in the above account. Suppose that “Smith” is a rigid designator for the prisoner in my world. Then, while not sending the letter would make it the case that “the prisoner’s counterpart in the other matching world lives”, it would not make it the case that “Jones lives”, where “Jones” is a rigid designator of Smith’s counterpart in the other matching world. For, if “Jones” is a rigid designator, then on counterpart theory propositions such as “Jones will be executed in his world” and “Jones will die in his world” are *necessary* propositions, since they are true at all worlds. So, if I send the letter, I will make it the case that Smith lives; if I do not send the letter, I will not make it the case that Jones lives. Thus, I should send the letter.

But this reply fails, because if “Smith” is a rigid designator in exactly the same sense that “Jones” is, then it will be false that if I send the letter, I will make it the case that Smith lives. For that Smith lives or is executed in his world becomes once again a necessary proposition on counterpart theory. So given counterpart theory I cannot base my decision whether to send the letter on whether *x* lives or dies, if “*x*” is a rigid designator in the strong sense of designating the same individual in every possible world.

Observe also the following alternate presentation of the dilemma. Call “*x*” a *quasi-rigid* designator of *y* if “*x*” designates *y* in the actual world, and in every possible world where *y* has a unique counterpart, “*x*” designates that counterpart. “Smith” in “If I send the letter, I will thereby cause it to be the case that Smith lives” should be read as a quasi-rigid designator in this sense. I now claim that if we do this, then it is true that “If I do not send the letter, I will thereby cause it to be the case that Smith’s counterpart in the other world matching Smith’s world lives”,

given Lewis's analysis of causation. For suppose I do not send the letter. The actual world then is w_R . The nearest world in which I fail to refrain from sending the letter is w_P . At that world it is false that "Smith's counterpart in the other world matching Smith's world lives". For, at w_P , "Smith" refers to the prisoner in w_P . But *his* counterpart in the other world matching his is the prisoner in w_R , who dies. Therefore, the counterfactual "Were I not to have refrained from sending the letter, Smith's counterpart in the other world matching Smith's world would not live" is true at w_R , and hence it is true that I cause it to be the case that "Smith's counterpart in the other world matching Smith's world lives." Why then should I prefer a scenario on which I cause it to be the case that Smith's counterpart in the actual world (i.e., Smith himself!) lives over a scenario on which I cause it to be the case that Smith's counterpart in the other world matching Smith's world lives, if all worlds are ontologically on par? If you cut Smith's counterpart, will he not bleed just as much as Smith?

Lewis has two ways out. The first is the unattractive one that our ordinary notions of morality are predicated on the fact that the actual world is *not* one with only one free choice occurring in an otherwise deterministic framework. But why should moral principles depend on whether more choices than one are made? The second way out is to insist that, *pace* Lewis, there is more to causation than the truth of the counterfactuals discussed above, and this "something more" makes it impossible that I cause anything in another world. I do not object to such an account of causation. However, this approach has another weakness: It rests everything on the importance of the fact that by sending the letter I *cause* Smith to live. But ordinary moral thought need not consider this very important. If the people in danger are strangers, I may not care whether *I* save a life or whether someone else saves a life, providing overall a life is saved. It would be nice to save a life, but why shouldn't I let someone else have the pleasure? While there is a moral imperative to take a trivial action if it is true that were that action taken, a life would overall be saved, and were it not taken, a life would not be saved, there is no such moral imperative to ensure that *I* be the one who saves a life.

Of course, sometimes this may matter. It may matter that *I* be the one who saves a friend. But in the case of a complete stranger, this does not matter much, not nearly as much as the fact that a life is saved. While with Lewis I reject consequentialism as a general account of morals, it is true that consequentialistic reasons sometimes have a rightful place in our deliberation. I may save the life of a stranger solely because I may want fewer people to suffer. Nor is there anything morally blameworthy in this reasoning. On the contrary, one might think that there is something selfish in being motivated by wanting to be the one who saves the life. But Lewis's theory leads to the revisionary result that consequentialistic reasons are *never* relevant.

The indexicality of moral *obligation* is irrelevant here. For while I do have the *prima facie* obligation to perform a trivial action if it can save a stranger's life, a defeater for this obligation is that if I take the action, then someone, who otherwise wouldn't die, dies. If I save Smith, Smith's counterpart dies, given Lewis's theory. Likewise, it is a defeater for the *prima facie* obligation to perform a trivial action to save a stranger's life if there is no consequentialistic improvement *on the whole* from one's action.

8.3.3 The second case: self-torture

Suppose now that I am in one of the two worlds which are deterministic and am unfree except for one choice I can make. This choice is whether to perform act *A*, which is to stick my finger in the light socket, thereby causing myself an excruciating bit of pain, but, I know, not actually killing or causing permanent damage to me. I know nobody in this world would benefit from this. It would clearly be irrational, indeed crazy, to perform *A*. Certainly, the act would not be praiseworthy. However, I shall argue that on Lewis's account, this act could be seen as praiseworthy.

When, and only when, I perform *A*, my counterpart in the other matching world refrains from *A*. Thus, if I perform *A*, my counterpart will be spared the pain of the electric shock. Moreover, on Lewis's analysis of causality, if I do perform *A*, then my performing *A* causes my counterpart in the other matching world to be spared the shock. For then, were I not to have performed *A*, my counterpart in the other matching world would have—this fact is just a

variant on (35). But it is a heroic and praiseworthy thing to spare someone else pain at the cost of receiving serious pain oneself. If I and a stranger were held prisoner by a tyrant, and I were told that if and only if I don't stick my finger in the light socket, my fellow prisoner will be forced to do so, then it would be a heroic thing to do it. [\[96\]](#)

Hence, if Lewis is right, then the act is heroic and praiseworthy. But if ordinary morals are right, the act is neither heroic nor praiseworthy, but insane. Therefore, Lewis's view in this situation contradicts ordinary moral notions. The fact that we are working in a world whose history allows for only one decision fork is irrelevant—our moral intuitions do not care about how many choices are available to how many people.

One might try the following reply on Lewis's behalf. Whatever I do, the numerically same people are shocked and unshocked. After all, the total space of possible worlds is fixed. Let "Smith" be a rigid designator of my counterpart (recall, that I define myself as my own counterpart) in w_P and let "Jones" be a rigid designator of my counterpart in w_R . Then, Smith gets shocked and Jones doesn't, whatever I do. So I shouldn't bother to shock myself, given that nothing changes by it. However, if this reply is made, then by the very same standard, in the case considered in the previous section, there is no point in my going to the trouble of e-mailing the dictator, given that nothing changes by it. Likewise, if it doesn't matter *to me* who gets shocked, neither should it matter in a case where I choose whether to murder someone or not whether it is *I* who do the murder or a counterpart—and this would undercut Lewis's reply to Adams' argument. Hence, the reply considered here is one that Lewis cannot make if he is to maintain ordinary moral views in other situations.

8.3.4 A general reply

One might also reply to the present case and to the case in the previous section that our moral beliefs implicitly depend on something that at first sight seems irrelevant to them, the fact that our world-history forks in more than one place (not just at the choice between doing *A* and refraining from *A*), so that there are many comparison worlds for counterfactuals whose antecedents are our actions. There is nothing surprising about the fact that there is an implicit dependence of our moral beliefs on a merely contingent fact about our world. However, apart from EMR, the multiple forkedness of our world-history seems to have no moral significance for the question whether I should stick my finger in a light socket or send an e-mail that would free an innocent prisoner. The fact that this contingent fact about our world has the moral significance that EMR forces on it itself shows that EMR is revisionary of our moral beliefs.

Of course, multiple forkedness of history is important for at least some of our moral beliefs (e.g., the fact that we will have future choices is relevant to the moral imperative for growth in virtue), so perhaps the dependence of our judgments in *these* cases on it is not so absurd. But now observe that in fact to escape from the paradoxes of Sections 8.3.2 and 8.3.3 one needs to assume a world-history that not only is multiply forked, but is *infinitely* forked.

For suppose that we are working within a world that is finitely many times forked. This means that the world-history is deterministic, except possibly for finitely many points in time when it forks indeterministically, and let us suppose that each fork is between finitely many alternatives. Suppose further that any possible free choice must (according to the way the instancial conditional and laws of nature are set up) happen at one these forking points and the choice is between those alternatives that the fork goes into. I shall only discuss the self-torture paradox here, but what I say goes over *mutatis mutandis* to the other case.

Suppose, e.g., that there are 460 worlds allowed by the laws of nature. One of these worlds is mine. I am now choosing whether to fry my finger in the light socket or not. Some of the 460 worlds are ones that are no longer available to my choice, having been closed by previous choices that I and others (and perhaps nondeterministic physical processes) have made. But there are, say, 312 worlds left open. Of these, my counterpart fries his finger in, say, 112, and avoids the pain in 200. By a "matching world" I shall now mean one of the worlds that are nomically open given all the choices and forkings in the past of the present world. There are thus 312 matching worlds. If I choose to stick my finger in the light socket, then it will be true that 111 people other than me in the

matching worlds will suffer this terrible electric shock pain—for, there are 112 matching worlds where my counterpart feels the pain, but on this choice, I am one of them. But if I don't stick my finger in the light socket, then 112 people other than me suffer this pain. By sticking my finger in the light socket, I bring it about that one fewer counterpart other than me in a matching world suffers the pain. This is a heroic deed.

But of course it is paradoxical for it to be heroic, given that it is crazy. This shows that if EMR is true, our moral notions of heroism (and not only—for a similar argument goes over *mutatis mutandis* for the case of saving the prisoner's life) depend on the fact that our world is *infinitely* forked. (The above argument needs there to be only finitely many matching worlds where my counterpart sticks his finger in the light socket.) But this is just too surprising. We can reasonably easily imagine living in a finitely forked world, and can see no reason why our moral notions should change there. Thus, EMR makes a completely morally insignificant difference, that between finitely and infinitely forked worlds, morally significant. And hence it is revisionary.

8.3.5 The need for consequentialistic considerations in ethics

As a general matter of fact, ethics does require consequentialistic considerations. Suppose you meet a stranger on the street who happens to tell you that he is thinking of giving \$1000 to the Helpers of the Hungry (HH). The stranger is not good at calculation and hates thinking about numbers. But you know that HH would save five lives with the \$1000. And, moreover, you know of a charity, the Efficient Helpers of the Hungry (EHH), that would indeed save ten lives with a \$1000 donation. The two charities serve the hungry in different distant countries, and neither you nor the stranger have any ties to either country or either country's residents. You also notice that the stranger is someone who is easily influenced. If you tell him: "Why don't you give the money to EHH instead?" he will do just that, without asking you any questions. He just seems to be the sort of person who gives to the last charity he has heard of. But if you try to explain anything about numbers of lives saved, he will not listen to you—he just wants to help, but does not care how many he helps.

Now, although one does not have a *duty* to do so, it is plausible that it would be *better* to make the suggestion that he donate to EHH. The natural justification of this judgment is that there is nothing deontologically wrong with speaking out whereas the overall consequences of speaking out are very good—five more lives are saved. Even if this be not admitted, at least the following will surely be admitted: In the "reversed" case, where the stranger originally planned to give money to EHH, it would be wrong to suggest that he give it to HH instead. [\[97\]](#) Again, the grounds for this judgments are consequentialistic.

Lewis cannot make these consequentialistic judgments, since whatever one does, there are infinitely many worlds where money is given to HH and infinitely many where it is given to EHH, and one's own actions make no difference to the overall result. Instead, he is committed to analyzing the matter in terms of causality. At first blush, one might try to say that what is good about telling the stranger to give to EHH in the original case is that one then causes ten lives to be saved. But this account fails, for although the saving of ten lives is a good thing, first of all one is at most responsible for five overall lives saved, but more crucially if one admits that in the original case the action is rightly morally described as "saving ten lives", which is good, then one will have to describe the action of suggesting that the stranger give money to HH in the reversed case as "saving five lives", which is *also* good, whereas it is evident that in the reversed case it is wrong to make the suggestion to the stranger. So the description "cause ten lives to be saved" is not the morally relevant one for the action of telling the stranger in the original case to give to EHH.

One might describe the action of suggesting the donation to be given to EHH as "causing five lives to be saved." However, that is inaccurate, because it is only five lives *overall* that are saved, while as a result of one's action five lives are lost that would otherwise have been saved. If one says "five lives were saved", without adding the word "overall", then one is begging for the unanswerable question: *Which* five lives were saved? For after all, there were *ten* saved, though only five overall. If the five that HH would save were a part of the ten that EHH would save, then there would be no problem with specifying which five were saved. But the supposition was that

HH and EHH serve different countries.

Now Lewis cannot say that by directing the donation to EHH one has “caused five lives *overall* to be saved.” For the “overall” brings in a consequentialistic consideration. The value of the action described in this way is that in the grand scheme of things more lives are saved. But Lewis cannot make this judgment, unless he makes it in a parochial way by restricting it to our world, which fails to do justice to our moral intuition that the lives of all *existent* strangers in need who are in objectively similar circumstances are of equal significance.

Lewis could say that by directing the donation to EHH one has “caused it to be the case that ten lives are saved and five are lost”. That is literally true on his counterfactual account of causation: were one not to have told the stranger about EHH, it would not have been the case that ten lives were saved and five were lost. However, this causal formulation again fails to do justice to the moral issues here. After all, a deontologist will say that there are actions that fall under the description “cause it to be the case that ten lives are saved and five are lost” that are in fact immoral: for instance, killing five innocent people to use their organs to save ten lives.

So for the description to carry the moral weight it does, it has to be modified as follows: one has “in a permissible manner caused it to be the case that ten lives are saved and five are lost.” Arguably, any action described in this way is a good action, and the description is one that Lewis can give. The problem, however, is in the question of *why* any action described in this way is a good action. In the case of an action where one simply causes it to be the case that x lives are saved one can say on Lewisian ethical grounds that what one has done is a virtuous thing, since what one has done is saved a life, thereby making oneself into the sort of person who saves lives, and this is good. However, in the case at hand such simple reasoning does not work, since the action likewise involves causing it to be the case that some lives are lost that otherwise would be saved, or at least it involves preventing the saving of lives that could have been saved, and this part of the action description is not one that in any obvious way contributes to having a virtuous character. Rather the only reason to think the action described as above is a good one is that the description entails that the action was one in which one acts intentionally and permissibly in a manner which results in the overall saving of five lives. But this consideration is once again a consequentialistic one.

One might try for a different kind of account. By suggesting that the stranger donate to the EHH, “one has caused it to be the case that the stranger performs a better action than the one he had intended to perform”, and that is a good thing. However, even if that is right, it certainly need not be the case that this description was the one under which one was performing the action. Very likely, instead of being motivated by the moral welfare of the stranger, one was motivated by the welfare of the existent starving people that are helped by various charities. But something stronger can be said. Because the stranger does not care about numbers of people saved, and one has not even told him that EHH saves more lives (because you knew there would be no point to this as the man does not reason numerically), he derives no moral benefit from the fact that his action was the saving of more lives. Here I will take the hard Kantian (or Thomistic) line that actions only have moral significance insofar as they are the actions they are intended to be, and do not derive moral significance from consequences that the agent does not know about. The stranger did not know that giving to EHH would be a saving of ten lives as opposed to the action he intended being a saving of five lives, and so this fact is morally insignificant if we evaluate the stranger’s action. Therefore, in suggesting that he change his action for one that unbeknownst to the agent involves more life-saving, you have not benefited the agent. Hence, again, we see that the action has no moral value except as far as the consequences are concerned.

It thus does not thus appear that Lewis can give a satisfactory account of why directing the donation to the EHH is good, because any satisfactory account will have to involve consequentialistic considerations.

8.4 The logical independence of free choices

Moreover, the following argument can also be made against Lewis. It is a conceptual fact that that one individual freely chooses A cannot entail that another individual freely chooses B , under appropriate

[98] descriptions. That one person freely chooses to stick his finger in the light socket cannot entail that another person freely refrains from sticking his finger in the light socket. But it is plainly logically possible that I exist in a world that has only one forking which corresponds to my free choice whether to stick my finger in a light socket or not. That I choose to freely stick my finger in the light socket entails that my counterpart in the other matching world freely refrains from this. But this contradicts the assumption that I and my counterpart are distinct free individuals, and hence our choices are logically independent. Indeed, if we are free, it should be logically possible for us *both* to stick our fingers in the light socket, a possibility that does not exist given EMR.

A Lewisian answer to this paradox will put weight on the fact that free choices are independent only under appropriate descriptions. “My counterpart in the other matching world freely refrains from sticking his finger in the light socket” fails to be an appropriate description, because it refers to the counterpart through the definite description “The counterpart of mine in the other matching world” rather than through a rigid designator. However, I argue, the definite description is quite appropriate for the action description because it is perfectly parallel to “The counterpart of mine in the actual world”, i.e., “I”. Admittedly, the individual picked out by “my counterpart in the other matching world” differs depending on whether I choose to stick my finger in the light socket or not, but he differs in the same way that the individual picked out by “I” does. If I stick my finger in the light socket, “I” will pick out my counterpart in w_P and otherwise my counterpart in w_R .

8.5 Fatalism

Andrew Beedle (1996) has argued that Lewis’s reply to Adams’ ethical objection implies fatalism. According to Beedle, Lewis is committed to the view that we do not choose which world we inhabit but merely discover it, and since the character of the totality of worlds is fixed, fatalism necessarily ensues. Note that this argument is closely similar to the reply to my self-torture example considered at the end of Section 8.3.3.

My arguments above assumed that we *do* choose which world we inhabit. I could, of course, pose them as a dilemma for Lewis: Either (i) we choose which world we inhabit, or (ii) we do not. In case (i), my argument works, and Lewis’s view runs counter to ordinary deeply-engrained notions of morality. In case (ii), my argument fails, but then Lewis’s view runs counter to ordinary deeply-engrained notions of morality even more strongly, since fatalism is contrary to these notions.

However, I do not need to present it as such a dilemma, because I do not think Beedle has proved his case sufficiently. The option is quite open for Lewis to say that we *do* choose which world we inhabit, and do not merely *discover* it. True, Lewis (1986a) does say that we “find out by observation ... what possibilities we are: what worlds may be ours, which of their inhabitants may be ourselves” (p. 492). However, it does not follow from this that *in addition* to discovering it, we cannot have a real role to play in bringing it about that we live in one world and not another. Indeed, on Lewis’s view of causation, it is simply true that our choices *cause* it to be the case that *this* world is actual. To see this, let w be a *de re* (i.e., rigid) designator of this world. Then by choosing to write this paragraph, I have caused it to be the case that w is actual. For, on Lewis’s view, to be able to say this, it is enough that w be actual, and that the counterfactual “Were I not to have chosen to write this paragraph, w would not be actual” be true. But w is actual (*ex hypothesi*), and the counterfactual is true (interpreted in a Lewisian sense) because in the closest world in which my counterpart does not choose to write this paragraph, w cannot be actual—since w *de re* includes the writing of this paragraph!

For Beedle to establish his case, he would have to say that the truth of the proposition expressed by the sentence “ w is actual” is prior in the order of explanation to all our actions; were that so, then indeed fatalism would result. But there is no reason why Lewis should need to be committed to such a claim. True,

it is evident that we do not have a choice about *some* of the matters in this world, so we do not have a completely arbitrary choice about which world to inhabit. But why can't our actions play *some* role in choosing which of the worlds we as a matter of fact inhabit?

One might object that this would involve some kind of "backwards causation" in that my future actions would determine which world I *now* inhabit. But no paradox results from *this* kind of backwards causation. "That I now inhabit *w*" is a temporally impure proposition, i.e., one that entails information about what happens at times other than now, and there is no absurdity in backwardly causing it to be the case that a temporally impure proposition was true. In 1992, the U.S. voters got to make it the case that it *was* true that in 1975 Hillary Rodham married the 42nd president of the United States. Any kind of an argument against this kind of backwards causation would be an argument for fatalism *simpliciter*, and those rightly unconvinced by such arguments will be unconvinced by Beedle.

If fatalism is to be avoided, an EMR-theorist has to say that when I choose between alternative actions, I am involved in the selection of which world I inhabit. There is, however, an apparent difficulty here. One is tempted to say that surely if our choices matter, the sum total of existence must be different. If it matters which world I choose to inhabit, then the alternatives must correspond to different totalities of all existing things. If I choose to shock myself, then the totality of existence is one where I inhabit a world where my counterpart shocks himself, and if I choose not to, then the totality of existence is one where I inhabit a world where my counterpart does not shock himself. Surely, the totalities in questions should be *different*, given that they correspond to different truth values of the proposition that I exist in a world where my counterpart shocks himself. But on Lewis's view, the totality of existence is fixed, and there cannot be different totalities of existence.

However, a good Lewisian answer to this difficulty is that the proposition that I exist in a world where my counterpart shocks himself is an indexical proposition, and as such its truthmaker is an ordered pair consisting of some aspect of the totality of existence and the individual who has expressed this proposition. [\[99\]](#) So the Lewisian can tolerate the idea that we choose which world we live in without denying the fixity of all existence. Fatalism is not the problem. The problem is in the subtler incompatibilities between EMR and our moral ideas, as argued above.

Section 9 Induction and actuality

9.1 Introduction

Forrest (1982) has argued against Lewis that we have no reason to think that our world is a nice one in its unobservable respects if Lewis is right. We can say that howsoever much regularity we might have observed in the world, if Lewis is right, we have no reason to think that the regularity will continue. For corresponding to any one world that continues to be regular (i.e., governed by the inductively observed natural laws) after this moment, there are infinitely many logically possible worlds that *were* regular but will no longer be so, and thus we have no reason to think that we're in a world that will continue to be regular. To the one world where the apple dropped tomorrow will fall straight down, there correspond infinitely many different ones with the same past but in which tomorrow the dropped apple will take some tortuous path in an unexpected direction. Call a world that shares our world's past [\[100\]](#) a "continuant world". Lewis then is committed to the claim that there are highly irregular continuants, because he is committed to a principle of recombination according to which anything could co-exist with anything, spatio-temporal layout permitting. In particular, any possible past could co-exist with any possible future.

Lewis's first reply to Forrest was that the set of continuant worlds that will behave regularly in the future is infinite and has cardinality equal to that of the set of irregular continuant worlds, and so that in fact one can pair up worlds in such a way that to each irregular continuant there correspond infinitely many different regular continuants

(Lewis, 1986a, pp. 119–120). One might reply that from the fact that the cardinalities of the sets of regular and irregular continuant worlds are infinite it does not follow that the cardinalities are *equal*. I shall argue in Section 9.4 that with a particularly strong view of what is involved in regularity, there *are* more irregular continuants than regular ones. Lewis's second reply is that Forrest-type problems persist on non-Lewisian views of possible worlds, too, and hence the problem should not count against Lewis (Lewis, 1986a, pp. 117–118).

Here, I shall allow Lewis his dubious assumption of the equality of the cardinalities between regular and irregular continuant worlds, but argue that even so, a Lewisian still cannot *know*, or even have reasons that make it more probable than not, that the world will continue even approximately regularly—the Lewisian in fact does not even have reason to think that gravity will function tomorrow. Assuming that, as seems obvious, there are at least as many irregular possible continuant worlds as regular ones, I shall argue that the problem is with Lewis's indexical notion of actuality. Moreover, I shall argue that the question of the theory of actuality is essential to discussion of the problem of induction. That induction holds places restrictions on what theories of actuality are tenable.

The structure of my argument will be as follows. First, I shall lay ground that will be neutral between various realist theories of possibility as well as between the Humean sceptic about induction and the defender of induction. Next, I shall show how a Humean might attempt to mount a sceptical argument from this ground. The sceptical argument will be seen to fail in general, but unfortunately for Lewis works if actuality is merely indexical. Therefore, since indeed we can inductively know that tomorrow things dropped will fall, *pace* the Humean, we can conclude that Lewis's indexical account of actuality fails. This attack on the indexical account of actuality can be considered as complementary to criticisms based on the analysis of language, as those of Gale (1996, chapter 5). However, the discussion will also show that more generally looking at what accounts of actuality give rise to scepticism about induction is a valuable way of looking at the problem of induction.

I am not claiming here that I will give an account of *how* inductive generalizations are to be justified. But I take it that we all think, and Lewis does as well, that they are. If not, we would have no reason to live as we do.

9.2 Neutral ground

For the moment abstract from the question of what theory of possibility and actuality is right, and assume only that we are going to be realists about possibility so that we agree that there exist such things as logically possible worlds. If it helps, we can even imagine possible worlds as physical books (of infinite size) in some heavenly library, each giving a maximal consistent description.

The notion of “a complete description of a possible world w up to time t ” will be needed. This is a maximal set D of propositions true at w such that: (a) the propositions in D do not jointly imply any proposition reporting what happens after time t at w , (b) every proposition true at w which reports a physical state at a time prior to or equal to t is in D , and (c) every proposition true at w which reports the occurrence of a phenomenal state in a finite sentient being at a time prior to or equal to t is in D .

Suppose now that we are given a complete description D of a possible world w_0 up to time t_0 . What w_0 is, e.g., a book in the heavenly library, or a concretely existing world, or some abstract entity, will depend on which theory of possible worlds is right. Moreover, we suppose that the propositions in D entail that induction has always held up to time t_0 .

I shall say that we have “reason to think that p ” if we are epistemically justified in taking p to be more likely true than not. Let S_D be the collection (set or proper class, depending on whether the collection of all possible worlds is a set or a proper class) of all possible world w of which D is a complete description up to t_0 . Now, because of the principle of recombination, Lewis will have to grant that there are infinitely many possible worlds in S_D , and that the number of these at which induction fails in the near future of t_0 is at least as great as the number at which it does continue to hold.

Remember now that all we know about w_0 is that it is a logically possible world at which all propositions in D hold, with D being a complete description of w_0 up to t_0 , and that the propositions in D entail that induction has

held up to t_0 . Does this by itself justify us in inferring, or even give us reason to think, that induction will hold at w_0 in the near future after t_0 ? The answer surely has to be negative. All we know about w_0 is that it is a member of S_D and we do not even know that w_0 is actual. The knowledge that w_0 is in S_D is surely insufficient to give us any reason to think that induction will hold at w_0 in the near future of t_0 .

The intuition here is particularly clear in the “heavenly library” view of possible worlds. We’ve picked out a book from the shelf, and read pages one through 1000, which pages give a complete description up to t_0 . Given our knowledge that there are at least as many books whose first 1000 pages are the same, and yet in which the pages after page 1000 describe induction failing in the near future of t_0 , as there are books with the same first 1000 pages and which then tell us of induction continuing to hold, we certainly are not justified in inferring that our book is one of those which will make it out that induction continues to hold. The exact same intuition holds on Lewis’s EMR. We are then talking of a concretely existing world w_0 about which all we know is that it is a member of S_D . Since we know that in S_D there are at least as many worlds at which induction fails in the near future of t_0 as ones at which it holds, we know that we have no reason to think w_0 is one of the “nice” worlds.

The above should not be controversial, being merely an observation about the structure of the collection of *possible* worlds, whatever these possible worlds’ ontological status might be. This examination of *logical* space is neutral territory on which both the contemporary defender of induction and the Humean sceptic must meet if they are to engage at all.

An over-eager defender of induction who wishes to dispute this might respond that the argument so far relies on the fallacy of thinking that in general if all we know about an entity x is that it is a member of a set C that can be partitioned into two disjoint subsets, A and B , where A has at least as many members as B , then we have no reason to think that x is in B . The fallacy lies in the fact that the sets A and B might have inner structure that would militate against this conclusion. (Suppose, for instance, that A is the set of all real numbers y satisfying $0 \leq y < 1$ and B is the set of all numbers y satisfying $1 \leq y \leq 1,000,000$. Then, A has just as many members as B , but if all we know about x is that it is a real number between 0 and 1,000,000, then it is *ceteris paribus* more likely that x is in B and so we overall have a reason in favor of the latter hypothesis.) However, if someone claims that it is more likely that x is in B , despite A having at least as many members as B , the onus of argument is on that person. Failing such an argument, we have no reason to think that x is more likely in B than in A . Even if the cardinalities of A and B are equal, we still cannot say that x is *equally* likely to lie in A and B but must suspend judgment about the probabilities involved, unless we have further relevant knowledge of the structure of A and B . But such suspension of judgment is all I need to establish that in the situation described we have no reason to think that x is in B as opposed to in A .

Thus, until a demonstration can be given of it being more likely that a possible world in S_D is going to have induction hold at it after t_0 , one has no reason to think that this is so. And no such demonstration has been offered. But one can do better than just to shift the onus of argument onto the opponent. For instance, one of the ingredients in law-like regularity confirmed by induction in worlds similar to ours—and it is only such worlds that we need consider—is that if a given material object x exists at both times t_0 and $t_0 + d$, where $d > 0$ is small, then its spatial positions at t_0 and $t_0 + d$ are close (in our world, the spatial positions cannot be apart by more than d/c , where c is the speed of light). However, there are many more—not in terms of cardinality but in terms of natural spatial measures—positions that are far from a current position than that are near to it, at least in large universes. Thus, although D might tell us that some particle x was always behaving regularly and not wildly jumping around from one end of the universe in w_0 to the other prior to t_0 , we still have more reason to think that x will after t_0 be very far away from where it is at t_0 than to think that it will be as close as regularity requires. After all, we are talking simply about logical possibilities here— w_0 is just known to be a logically possible universe in which induction has held up to time t_0 .

9.3 The Humean argument and a reply Lewis cannot give

Thus, we do not have reason to think that the possible world w_0 of which D is a complete description up to

t_0 is going to be approximately regular in the near future after t_0 . Someone might think that in fact this justifies scepticism about induction. For suppose we learn two more pieces of information: w_0 is actual and t_0 is now. Then, the Humean says, if *before* learning this we did not have reason to think that induction holds at w_0 after t_0 , surely we still have no such reason. For, after all, when we learned that w_0 is actual and t_0 is now, we did not learn anything relevant about what is true at w_0 . If p is a proposition true at w_0 when w_0 is actual, then p is also true at w_0 when w_0 is non-actual, and conversely. We do not have any new data to bring to bear on the question of whether induction will continue to hold in w_0 after learning w_0 is actual, since any data we may have in virtue of our own observations of the actual world is data that was already a part of D , since D contained descriptions of all our phenomenal states up to t_0 and also of all physical states up to t_0 given that w_0 is the actual world. Therefore, the objector continues, if prior to knowing that w_0 is actual and t_0 is now, we did not have reason to think that induction will hold at w_0 in the near future of t_0 , neither do we have such reason after learning this. But, the Humean concludes, our actual epistemic state is if anything *poorer* than that which is envisioned in this thought experiment, since in this thought experiment we would have a complete description of the past, while we only have a partial account, so that if in the thought experiment we had no reason to think that induction will continue to hold, in real life we *a fortiori* have no such reason either.

I take it that everything the Humean says is correct here, except for the claim that “when we learned that w_0 is actual and t_0 is now, we did not learn anything relevant about w_0 ”. For once this claim is granted, the sceptical conclusion that we have no reason to think induction will hold in the near future follows. We can imagine being given a complete description of our world up to now, and the above argument shows that were we given that description, we would not have reason to think that induction will continue to hold even in the near future. And surely it is right to say that if we would have no reason to think this given the complete description, likewise we have no reason to think it now.

The right response to the Humean, then, is to assert that we *have* learned something relevant when we learned that w_0 is actual and t_0 is now. When we make inductive inferences about the future states of the actual world, our sample space consists of the known states of the *actual* world. States of non-actual possible worlds, such as of the other members of S_D , are irrelevant. When we have learned that w_0 is actual, the problem of what will be true at w_0 after t_0 was transferred from questions about logical space, to questions about *actual* events. We *know* that induction holds, and will hold at least in the near future, in the actual world, and so by learning that w_0 is actual and t_0 is now, we have learned that induction will continue to hold in w_0 .

However, as we shall shortly see, Lewis’s indexical theory of actuality precludes him from being able to make this natural response to the Humean. Since the Humean argument succeeds as soon as it is granted that by learning that w_0 is actual and t_0 is now one has not learned anything relevant to the problem of induction, it follows that if Lewis is forced to grant this, Lewis’s view does entail scepticism about induction. But since we *know*, or at the very least have reason to think, that induction will hold in the near future [\[101\]](#), we therefore know that Lewis’s theory of actuality is false. Whether this implies that EMR as a whole is false depends on whether one is able to make EMR work with a different theory of actuality, which is a dubious prospect.

To see that Lewis cannot make the natural response to the Humean, observe that on Lewis’s indexical theory of actuality, in learning that w_0 is actual and t_0 is now, one has only learned an indexical fact, one akin to learning that something is *here*. However, learning a merely indexical fact cannot by itself give one any additional information about non-indexical claims such as “induction will continue to hold at w_0 in the near future of t_0 ”. The reason for this is the Principle of Impartiality of Reason (PIR) that says that whether a belief is epistemically justified does not depend on whether *we* are persons concerned in the matter and whether a time involved in the content of the belief is *now*. More precisely:

(PIR) Let K be a body of knowledge that is known to entail the existence of a finite person x existing at a time t_0 (where x and t_0 are definite descriptions or proper names). Let p be the proposition that I am x and that

the present time is t_0 . Let q be a proposition that does not involve indexicals. Then, K gives one reason to think that q if and only if $K \& p$ gives one reason to think that q .

Now, some may dispute PIR in the case of normative propositions, arguing that the fact that I am x can by itself (and not just by telling me other non-indexical facts about x) give me reason to make different normative claims about what x should do. With Kant, I take this to be wrong-headed, but for the purposes of this paper, I am willing to restrict PIR to non-normative claims.

PIR has been formulated to avoid the following kind of counterexample. Let K be the proposition that a man x alive at t_0 is going to win the lottery. Then, when I additionally learn that I am x , I learn the non-indexical fact that a philosopher is going to win the lottery, a fact I could not know on the basis of K itself. However, this is not a counterexample to PIR, because it is not just the proposition that ($K \& (I \text{ am } x \text{ and } t_0 \text{ is now})$) which gives reason to think a philosopher is going to win the latter, but rather this proposition conjoined with the additional proposition that I am a philosopher.

The application of PIR to the case at hand is follows. We let K consist of D and the claim that D is a complete description of a world up to t_0 . [102] We let q be the proposition that induction will hold at w_0 for a while after t_0 . Then, by the indexical theory, my learning that w_0 is actual is merely my learning that I am one of those persons who are asserted to exist by a proposition true at w_0 . Moreover, since I know some uniquely identifying properties of myself, I know which person x described in D is the one I am identical with. Since it has been shown to be true, in a way neutral between the views of the sceptic and the defender of induction, that K did not give me reason to think that q , by PIR neither does $K \& p$ give me such reason. Therefore, if learning that w_0 is actual is nothing but learning that I am one of the persons said to exist by a proposition true at w_0 , i.e., if Lewis's indexical theory of actual is correct, it follows that indeed learning that w_0 is actual and t_0 is now does not give us any relevant information with regard to the question of whether w_0 will continue to have induction holding in the near future of t_0 . Hence Lewis cannot make the only possible response to the Humean argument.

Note that the response to the Humean argument that I am envisioning is not an argument showing how induction is to be justified. Rather the response shows which premiss in the Humean argument can be rejected by the defender of induction.

Note that Lewis (forthcoming) allows some indexical information to affect one's credences. However, the case he is working with has the indexical information being added to a body of knowledge that is not purely non-indexical [103], and hence Lewis is not denying the PIR. And even if Lewis were to deny the PIR, then although the argument against him could no longer be an *ad hominem* one, it would still be a good argument in light of the plausibility of the PIR.

9.4 Conclusions

We see that indeed a reasonable principle of recombination together with Lewis's theory of actuality entails inductive scepticism. Recall now that Lewis has attempted to counter Forrest's argument that Lewis's theory overthrows induction by saying that exactly the same problem recurs on other theories of possible worlds. But as we have seen, the problem is specifically locatable in Lewis's theory of actuality, and so formally speaking it need not be present once that theory is dropped.

Our argument does, however, show something more general about the problem of induction. In order for induction to be justified as more likely than not, it must be the case that learning that a possible world w_0 is actual and that t_0 is now provides information relevant to thinking that it is true at w_0 that induction will continue to hold after t_0 . Since the temporal impartiality of reason shows that the determination of whether t_0 is *now* or not should not by itself have a significant role in determining whether one is justified in thinking that induction will hold after t_0 , it follows that specifically learning that w_0 is actual must be the crucial additional piece of information which gives

us reason to think that induction will continue to hold once we know that it held in the past of t_0 . This places a restriction on what theories of actuality are tenable. Theories of actuality on which the assertion that w_0 is actual cannot provide any such information are untenable given that we do know induction holds.

Lewis's *tu quoque* reply to Forrest will thus have value if and only if there are no theories of actuality on which learning that w_0 is actual gives us information relevant to the problem of induction. In Part VI such a theory will in fact be given. For the moment, however, consider the following theory which shows that indeed the problem is with Lewis's account of actuality. Strawson (1959, pp. 126–129) in his discussion of Leibniz considers, though ultimately rejects, an ontology on which all things that exist are maximally specific concepts of individuals, so specific that any given concept could only be instantiated in one possible world. Then, such a concept C is said to be actual if and only if C is a concept that is instantiated in the best of all possible worlds. In other words, Strawson is entertaining a theory of actuality on which to be actual is to be in the optimal world. On *this* theory of actuality, learning that w_0 is actual certainly gives one reason to think that w_0 will continue to be regular after t_0 . But, however good a solution this might be formally, it is useless in practice, because, as Strawson observes, this definition of actuality opens to doubt the question of whether *we* are actual, since we do not know whether we are members of the optimal world, and so the worry about whether induction will continue to hold in *our* world will continue to plague us.

The problem of induction, on this view, boils down to the problem of justifying a theory of actuality on which (1) learning that such-and-such a world is actual gives information supportive of induction in that world, while (2) we can *know* that we are actual.

Observe, finally, that the argument against Lewis can be formalized into a *reductio*:

- (33) Let D be a complete non-indexical description of the actual world up to the present (t_0) in temporally pure terms. (Definition.)
- (34) D contains the claim that gravity has always held prior to t_0 . (Premise.)
- (35) Conclusions about the actual world reached by reasoning in accordance with the canons of inductive reasoning are justified, and in particular knowing that gravity has always actually held prior to t_0 justifies one in believing it will continue to hold after t_0 . (Premise.)
- (36) There are at least as many worlds satisfying D in which the law of gravitation fails after t_0 as there are worlds in which it continues to hold. (Premise.)
- (37) * Therefore, knowing that an entity w is a world satisfying D does not by itself epistemically justify inferring that w is a world at which gravity holds after t_0 . (Premise, justified intuitively by appeal to (39).)
- (38) * Theoretical reason is impartial with respect to merely indexical facts: If knowing that x is F (where F is purely non-indexical and x is a definite description or proper name) does not epistemically justify inferring that x is G (where G is purely non-indexical), then neither does knowing x is F and that x is I (now, here, etc.: any pure indexical will do) justify inferring that x is G . (Premise.)
- (39) * Actuality is indexical. (Premise.)
- (40) Therefore, knowing that an entity w is a world satisfying D and w is actual does not epistemically justify inferring that w is a world at which gravity holds after t_0 . (By (40)-(42).)
- (41) * But knowing that the actual world satisfies D and w is actual epistemically justifies inferring that gravity holds in w after t_0 . (By (37) and (38).)
- (42) Therefore, knowing that the actual world satisfies D and w is actual both does and does not epistemically justify inferring that gravity holds in w after t_0 , which is absurd. (By (43) and (44).)

Here, the premisses marked with an asterisk form an inconsistent quadruple, all members of which are highly plausible except (42). Therefore, (42) is false, *pace* Lewis.

9.5 Cardinalities of regular and irregular continuants

I have already mentioned Peter Forrest's (1982) argument that if Lewis is right, then to any regular continuant of this world there correspond infinitely many irregular continuants, and hence EMR tells us that we ought to believe that the world will not be regular. Lewis (1986a) replied that to each irregular continuant there also correspond infinitely many regular ones, and hence the argument cannot be probative. I shall argue that Lewis's reply is unsound if the kind of regularity we have in this world is of a kind that justifies one in inferring that there are laws of nature roughly of the sort that contemporary physicists consider. My argument shall be formulated assuming that qualitatively identical worlds are numerically identical. The general case follows by using the argument below on equivalence classes, *if* each class of qualitatively identical worlds has the same cardinality. [104]

Note that physicalism is not required for the argument. If there are non-physical entities (as indeed I believe there are since I shall argue for the existence of a God in Part VI), there is still a "subworld" consisting of those aspects of the world that are physical. The argument below can then be applied to the physical subworlds of the actual world, or, more precisely, to equivalence classes of worlds under the equivalence relation \sim where $w_1 \sim w_2$ if and only if w_1 and w_2 are physically indiscernible. But for simplicity the formulation will be the physicalistic one.

9.5.1 The deterministic case

First, note that Lewis's reply completely fails if the *kind* of regularity that is present in the world is deterministic and if one thinks one could be justified in inductively inferring one specific set of laws of nature. For determinism ensures there is only *one* possible regular continuant (and not infinitely many), whereas certainly there are infinitely many irregular continuants. Hence Lewis's EMR leads us to say, rather implausibly, that even if (*per impossibile*) we knew with certainty that the world had always in every way behaved deterministically, nonetheless we would have good reason to expect that the determinism will no longer continue in the future, since there is only one world where it does and infinitely many which had been deterministic until now, but will no longer be so.

9.5.2 The non-deterministic case

We have thus seen that Lewis's reply only has any plausibility if the kind of regularity that the world is to exhibit is non-deterministic to such a degree that the actual world as it was up to now has infinitely many regular continuants. However, even so Lewis is wrong, and the reason is that as Cantor has shown, not all infinities are the same. Let t be the current time, and let R_t be the set of all regular continuants of this world, i.e., worlds which continue to behave in accord with laws the way (a way described by some rules or laws) that this world has behaved. Let us grant that R_t could be infinite. But it is nonetheless a *set* of some fixed cardinality. This point is not as trivial as it may have sounded to a reader before reading Section 7, since I have argued there that the collection of all possible worlds is not set. However, R_t *is* a set, because the rules or laws in accordance with which regularity is defined specify the *set* of all possible continuants for any given world at any given time. For, these rules specify what kinds of possible entities can come to be present in the world and how many of them can come in and when. The possibility of such specification is, I take it, analytically contained in the notion of a *regular world*.

To make this point clearer, suppose that the kind of regularity that the world has is that which some kind of quantum field theory posits. Then, the field theory posits that the basic entities in a regular continuant will be several kinds of fields governed by certain equations. Now, a field can be represented as a function from some fixed set A to some fixed set B (e.g., a scalar field on a four-dimensional Euclidean space-time can be represented as a function from \mathbf{R}^4 to \mathbf{R} , where \mathbf{R} is the set of all real numbers). The collection of *all* functions from A to B is a set denoted by B^A whose cardinality can be computed if the cardinalities of A and B are known (in the case of scalar fields on a four-dimensional Euclidean space-time, the cardinality of the set of functions from \mathbf{R}^4 to \mathbf{R} is a cardinality known as f which is bigger than that of the continuum). Or suppose that the kind of regularity that the

world enjoys implies, among other things, that it is composed of up to n (a finite or infinite number) particles, which come from k (a finite or infinite number) different kinds, each of which kinds of particles has at most p (a finite or infinite number) different properties that can each be described by real numbers (the properties might be things like charge, coordinate components of the position, etc). Then, the set of regular worlds, and *a fortiori* the set of regular continuants which is a subset of it, will have cardinality at most $c^n k^p$, where c is the cardinality of the continuum.

Once again, assume the Axiom of Choice, with the same justification as in Section 7.2, above. Let a denote the cardinality of R_t . Let n be any infinite cardinality greater than a , e.g., 2^a (the cardinality of the set of all subsets of R_t , which Cantor's diagonal argument has shown to be greater than the cardinality of R_t). We will let U_t be a set of continuant worlds in which tomorrow the earth and sun disappear counter-nomically for a day and are replaced by a big cloud of photons all dancing the polka [\[105\]](#), with the number of these photons having some value between \aleph_0 (the cardinality of the set of integers), inclusive, and the cardinal number \aleph_n (defined, with the synonym n^* , in Section 7.2 above), exclusive, which photons then disappear, the earth and sun return, and everything comes back to normal nomic order the day after. The number of different cardinalities m satisfying $\aleph_0 \leq m < \aleph_n$ is equal to n , and so we can choose the worlds in U_t so that U_t will have cardinality at least n and hence greater than a . Moreover, since n is an *infinite* cardinality, it must be that n is *infinitely many times greater* than a .

Hence, since all the continuants in U_t are irregular, there are infinitely many times more irregular continuants than regular continuants, so that to every regular continuant there correspond infinitely many different irregular ones. And there is no way of reversing this relation since there is no way of arranging the regular and irregular continuants in such a way that to every irregular continuant there should correspond a different regular continuant (if there were, then it would no longer be the case that $n > a$). And because we have chosen n to be *any* infinite cardinality greater than a , this argument in fact shows that one can point out as many times more (in terms of cardinalities) irregular worlds than regular ones as one might wish to. [\[106\]](#)

Not all infinities are the same. The infinite number of irregular continuants is greater (infinitely so) than the number of regular continuants, even if this latter number is infinite. Hence, if Lewis's EMR is correct, we have *a priori* rational reason to reject the possibility that the world, or even its physical aspect, is governed by the sorts of laws of nature that current science talks about. But it is irrational to reject this possibility *a priori*—at best, one can reject it *a posteriori*. Hence, Lewis's EMR entails that something irrational is rational, and hence EMR is refuted by *modus tollens*.

Section 10 Possible worlds and probability

If EMR is right, the probability that a proposition p will happen is the same as the probability that the actual world is a member of the set $S(p)$ of worlds at which p holds (which set is in fact identical with p , if we accept Lewis's account of propositions).

Suppose now that we are in a world which is deterministic except for one indeterministic throw of a loaded coin. According to the laws of nature in that world, the probability of the coin coming up heads is $2/3$ and the probability of tails is $1/3$. This is presumably conceivable. Suppose further we know the laws of nature and all of the state of the world up to the throw of the die. Let H be the proposition that the die will come up heads. Then, since we know the laws of nature, we know that $P(H)=2/3$. However, one might make the calculation differently. $P(H)$ is the probability that the actual world is in $S(p)$. But we know that in fact there are only two worlds, call them w_H and w_T , respectively, that the actual world could be—namely, that world compatible with the past we know in which the coin lands heads and that world compatible with the past we know in which the coin lands tails. It seems natural given that we all we know about that the actual world is that it is one of w_H and w_T to suppose that in fact the probability that it is w_H is $1/2$. But the epistemic probability that the actual world is w_H is $P(H)=2/3$. Hence, $1/2=2/3$, which is absurd.

One might think that this argument commits the simple fallacy of inferring from the fact that there are two possible situations that the probability that each one of these situations is actualized is $1/2$. If it did this, the argument would have the same weight against EMR as it does against other theories of possible worlds—i.e., none. But in fact the argument specifically depends on EMR. The point is that, as we shall see below, from an appropriate point of view there is a symmetry as to which ends up actual.

To see this, formulate the paradox for EMR as follows. Let ARP_H and ARP_T be my counterparts in w_H and w_T , respectively. I know that I am one of these two people. But which one? Both are concretely existing persons. All the information I have is logically compatible with me being either one of these persons. Moreover, both persons have the same background knowledge, being indiscernible up to the flip of the coin. Recall the notion of “the other matching world” from Section 8.3.1, i.e., the other one of the worlds with the same laws of nature and initial conditions as mine. Let “ I^* ” be my counterpart in the other matching world. Now, I know (since I know all the past information, it is supposed, and the laws of nature) that I^* is in the same epistemic position as I am. There is nothing that I know about myself that I^* doesn’t know about himself. The Impartiality of Reason then requires that I view the epistemic possibilities for I^* in the same way as I judge the epistemic possibilities for myself. It would be irrational for me to make some epistemic judgment about myself when I would not make about I^* , since the only difference between what I attribute to I^* and what I attribute to myself is purely indexical: I am myself and am not I^* . Note that this argument presupposes that Elga’s (2000) Sleeping Beauty counterexample to Van

Fraassen’s reflection principle fails. [\[107\]](#)

Consequently, the Impartiality of Reason requires that I say that

$$(43) \quad P(\text{I am } ARP_H) = P(\text{I}^* \text{ is } ARP_H).$$

But, necessarily, I am ARP_H if and only if I^* is *not* ARP_H . Hence,

$$(44) \quad P(\text{I am } ARP_H) = 1 - P(\text{I}^* \text{ is } ARP_H).$$

From (43) and (44) it follows that $P(\text{I am } ARP_H) = 1/2$. But I also know that $P(\text{I am } ARP_H) = 2/3$. Hence, $1/2 = 2/3$ if EMR is true, and thus EMR is false.

Other than challenging the Impartiality of Reason, one objection Lewis can make is to the intelligibility of the assumption that although there are two worlds that are epistemic positions for us, and although we know the laws of nature, nonetheless the probabilities of us ending up in one of these worlds are $2/3$ and $1/3$ (instead of $1/2$ and $1/2$). One way to put this objection is to worry that we are talking of the probability of a one-time event, and one-time events, the objection continues, do not have probabilities. But in fact the same argument goes through in a case where in fact the laws of nature have one to flip the loaded coin some large but finite a number of times, say 1,000,000,000. Then, we can construct an argument closely analogous to the above one to argue that the probability that we are in any of the $2^{1,000,000,000}$ possible worlds open to us through all the flips is equal to $1/2^{1,000,000,000}$ and thus the probability that the next flip will come out heads is $1/2$, since half of those worlds still open to one realize that probability (one can argue with multiple counterparts in much the way I have argued above with me and

I^*), whereas the laws of nature give the probability as $2/3$. [\[108\]](#)

We can put the paradox differently as follows. Assume EMR. It is a basic principle of rationality that if two persons are rationally required to assign some specific probability to something, and if the two persons both know what the other believes and what the other knows, then the probability they are required to assign is the same. But now my counterpart I^* *ex hypothesi* knows everything about me that I know, and *vice versa*. Thus, he will be rationally required to assign the same probability as I to the claim that I am the one in whose world the coin comes up heads. Hence, he will be rationally required to assign probability $2/3$ to this, as *my* knowledge of the laws of nature rationally requires this assignment of me. However, the laws of nature holding in *his* world rationally require that he believe there is a $2/3$ probability that the coin will come up heads in his world.

But as he knows that the coin comes up heads in his world if and only if it does not do so in my world, it follows that he is rationally required to believe that there is a $1 - 2/3 = 1/3$ probability that the coin will come up

heads in *my* world. However, as we saw above, he is rationally required to assign probability $2/3$ to this. But a person none of the beliefs on which a judgment is based are false (and in the case in question, it was assumed that the judgment was based on true beliefs about the past and the laws of nature), if rationally required to make some judgment, is not rationally required to make any incompatible judgment. Hence, since the judgments that the probability is $1/3$ and that it is $2/3$ are both required, they are both compatible. Thus, $2/3=1/3$, which is absurd, and so the initial assumption that EMR is true fails.

There are no such paradoxes if we do not accept EMR. For, the derivation of (46) fails then, since I^* does not exist at all, and Principles of Impartiality of Reason apply only to existing persons, at the pain of generating sceptical conclusions of the sort seen to result for Lewis's theory in Section 9. And if we reject EMR, the worlds inhabited by I and I^* are not ontologically on par. The inconsistency between my rational judgment that the probability of the coin landing heads being $2/3$ and the rational judgment I^* makes in his world of the probability of the coin thus landing in *his* world is resolved by the fact that the judgment I^* makes at his world fails to take into account the fact that his world is *in fact* non-actual. Of course, on Lewis's theory, it is still true *at my world* that I^* 's world is non-actual. But this is not good enough to resolve the paradox because on EMR it is true at my world that I^* knows this, and so this knowledge cannot explain the conflict between his judgments. But without EMR, there *is* something that the beliefs that I^* has at his world are absolutely wrong about: the actuality of his world.

As in the case of induction, it is not difficult to see that the basic problem with an EMR-based treatment of probabilities is that Lewis conflates logical order with the order of concrete existence.

Section 11 The epistemological objection

But even if all the above problems were solved by Lewis, some may think (e.g., Richards, 1975) we still have a pressing problem: How on earth are we supposed to know what goes on in the other worlds? The other worlds are causally isolated from us. To know modal propositions is to know what happens in other worlds. But how can we have knowledge of these entities that are causally isolated from us? How can we know, e.g., that it is possible that there exist unicorns and that it is impossible that there exist square circles, without employing some physically impossible telescope for gazing at worlds other than ours and finding that some of them contain unicorns but none have square circles in them?

Lewis thinks that this objection stems from thinking that all knowledge (with the exception of intentional knowledge?) needs to be causal—the object of knowledge must give rise to the knowledge. But it is impossible that something in another world should bring about any effect in this world. Hence, it is impossible for us to know modal claims if EMR is true. But Lewis goes on to observe that the same problem arises for mathematics (Lewis, 1986a, Section 2.4). The equality of the angles in a triangle cannot cause anything, and certainly cannot cause me to believe it. Hence, the causal account of knowledge is wrong, as it cannot account for mathematical knowledge.

But the case of EMR seems disanalogous to the mathematical case, precisely because mathematical knowledge is knowledge of abstracta whereas modal knowledge, according to EMR, is knowledge about things that have the same empirical status as the things we are familiar with. We know mathematical truths through *proof* but we do not have a proof that some other world contains a unicorn. However, this is incorrect. We *sometimes* know mathematical truths through proof, but sometimes by other means. For instance, I have tested the truth of the Andreev-Matheson Conjecture on tens of millions of randomly generated special cases. Since the conjecture makes an assertion about infinitely many cases, this is definitely not proof. But it might well count as providing *knowledge*, knowledge of mathematical objects, not through proof, and not through any causal contact with them.

But, one may ask, if we can only know whether there are unicorns in this world by empirical observation, why should we be able to confidently affirm that in *some* world there are unicorns, without having made any observation? However, there is a natural Lewisian answer to this question. To know whether there are unicorns in this world is to know whether unicorns are spatio-temporally related to us or not. If without any observation we had believed that in fact there are no unicorns, this would not count as knowledge because this belief would have been

wrong had we been living in another possible world where there were unicorns and our observations had not ruled out that possibility. But there is no world at which it is false that unicorns are possible. Thus it is impossible for someone to be wrong in thinking that unicorns are possible (cf. Lewis, 1986a, p. 113).

But the impossibility of being wrong about something does not suffice for knowledge. Otherwise, the mathematical cranks who believed that they had proofs of Fermat's Last Theorem would necessarily count as *knowing* that Fermat's Last Theorem is true, which is absurd. However, Lewis can always bring in some kind of reliabilism according to which we count as having modal knowledge providing we have some truth-directed faculty by which we generate modal beliefs, which faculty is in fact right most of the time. Thus the epistemological objection fails. However, it does show that Lewis's ontology of possibility is epistemologically irrelevant. This is not a fatal objection, but it will be a point in favor of the alternative account I shall eventually sketch that the ontology of possibility on that account will be such that it *could* be epistemologically relevant.

Richards (1975) goes even further than just saying that the Lewisian account gives no epistemology of possible worlds. If we could work out on the basis of things in this world what things in other worlds are like, the Lewisian theory would have "no explanatory value as a theory of meaning" (p. 110); indeed, Richards urges that since we would then in the next step want to move from what happens in other worlds to modal claims about this world, our account would be viciously circular. However, this objection confuses truth conditions with assertibility conditions. The move from this world to other worlds is a move justified by *assertibility* conditions: this-worldly situations *epistemologically* ground categorical facts about other worlds. But the move from other worlds to this world is a move justified by *truth* conditions: other-worldly situations, on Lewis's view, *ontologically* ground this-worldly modal facts.

Section 12 A final assessment of extreme modal realism

Lewis notes that the most common "objection" to his view is the "incredulous stare". How could such a crazily rich ontology be the case? I submit that one reason for an incredulous stare may be an intuitive recognition that such an enlargement of what we think reality contains is going to be revisionary to various ordinary ideas. We have seen that there *is* in fact such a revision entailed by EMR, and the revision is too expensive.

It would be nice to have a realistic discourse about possibilia. But if that is going to deny us standard moral notions, remove the possibility of having inductive knowledge, undermine objective probabilities, and force a revision of the notion that concrete substances can always be arranged into sets, then the cost is too large. For instance, without inductive knowledge of things in *this* world, we would have little use for possibilia at all, and so that alone is reason not to accept the theory. The cost of EMR is unacceptable, but perhaps there are other realistic theories of modality where the cost is lesser.

Part IV. Ersatz ontologies

Section 1 The general strategy

Lewis's theory has a number of useful consequences as, indeed, many theories of possible worlds do. If only the price, in terms of paradoxical consequences and ontological extravagance, were lower, one might adopt it. Many people have tried to find a cheaper replacement for EMR. All of them have in common the fact that they do not take possible worlds to be concretely existing worlds ontologically on par with ours. Rather, in Lewis's terminology, they prefer "ersatz worlds". These may be mathematical constructions, sets of sentences, collections of propositions and the like. All the ersatzists whose views I shall consider will take their ersatz worlds to be abstract. Their ideal would be to find a place for possible worlds within an ontology we already accept. Then, the hope is, one could talk about possible worlds without any ontological commitment.

There are two levels of ambition that an ersatzist may have. On the one hand, the Ambitious Ersatzist may claim that his worlds are things that we are really talking about in making modal claims. When we make a true modal claim, the truthmaker of the claim really involves one or more of the ersatz worlds, or a portion or portions thereof, or something standing in some objective non-projected relation or relations thereto. The Ambitious Ersatzist sets herself the task of answering the Parmenidean problem of what modal language is about. This is, after all, what Lewis was doing. The Ambitious Ersatzist cannot hope to avoid ontological commitment, because she is committed to the ontological claim that the truthmakers of true modal propositions are objectively related to her ersatz worlds. Even if the ersatz worlds are things we would believe in anyway, the claim that *these* things are the truthmakers of true modal propositions will be an ontological commitment we do not explicitly make.

The Unambitious Ersatzist, on the other hand, merely introduces surrogates for Lewisian worlds which are helpful logical constructions, theoretical tools, or useful maybe-fictions. While the Unambitious Ersatzist's ersatz worlds had better exist in some way, else in talking of them she is cannot be speaking truly, nonetheless they need not be at all related objectively to the truthmakers of modal propositions. The Unambitious Ersatzist makes no ontological claim beyond the fact that her ersatz worlds exist. She need not even deny that EMR is true. Unambitious Ersatzism thus strives to provide an ontological neutral way of talking of possible "worlds", but one that makes it clear that such talk can coherently be made sense of as talk *of* something and hence is not completely meaningless.

I shall not raise many criticisms against Unambitious Ersatzism. After all, every basic modal realist, of whatever stripe, will take it as plausible that the Unambitious Ersatzist's project can succeed, where "basic modal realism" is the view that there *are*, not necessarily Lewisian, possible worlds. *Whatever* we think possible worlds *really* are, whether concretely existent worlds, or books in heavenly library, it is likely that somewhere in Cantor's paradise ersatz versions of them can be found. For instance, suppose we accept EMR. Then, it is plausible that one can find some abstract proper class that stands in one-to-one correspondence to these worlds, and then we can find abstract relations T and I such that if p is a proposition, x a possible individual and w an ersatz world, then pTw holds if and only if p is true in the Lewisian world that w corresponds to and xIw holds if and only if x exists in the Lewisian world that w corresponds to. Then, T and I can be used to define the *true-in* and *exists-in* relations between propositions and individuals, respectively, and ersatz worlds. And then we have an ersatz theory.

The Unambitious Ersatzist's task is less trivial if basic modal realism cannot be taken for granted. I will argue that this task succeeds when worlds formed as classes of propositions are brought in. However, I shall argue that contemporary ersatzist approaches fail to conform to the more stringent criteria that Ambitious Ersatzism is to be judged by.

Note that Leibniz is an ersatzist, in that he believes he can make do with worlds that do not exist ontological on par with ours, but unlike the ersatzists' I shall consider here, Leibniz's worlds are not Platonic self-subsistent abstracta, but concrete thoughts in the mind of God. Therefore, discussion of Leibniz's ersatzist

approach will be postponed until Section 1 of Part VI.

Section 2 Linguistic approaches

2.1 The basic idea

On the first ersatzist approach, worlds are maximal consistent collections of sentences in something we can call a “worldmaking language” (cf. Roper, 1982 and Jeffrey, 1983). As discussed in the introduction (Section 4.3 of Part I), we cannot take these to be actual utterances of sentences. Nor, can we take them to be possible utterances of sentences, since if we are entitled to help ourselves to the possibilities that possible utterances are, we might as well help ourselves to the possibilities that possible universes are. Rather, we must take “sentence” to indicate type, not token. Presumably, the “type” can be some sort of set-theoretic constructions, e.g., a finite sequence of symbols arranged according to some grammar.

A proposition is then true in a world if and only if it is jointly implied by the abstract propositions expressed by the language. Some propositions not only jointly implied in this way, but are actually expressed by some sentence that is a member of the world. These fortunate propositions are said to be *explicitly represented* by the world; all the rest are implicitly represented. One might think that the fact that the worlds are maximal collections of sentences ensures that all propositions true in a world are explicitly represented by it, since otherwise we could just add to the collection sentences representing the implicitly represented propositions. But that is not the case, because the language in question may lack the resources for explicitly expressing all propositions. Thus, unless, as is highly implausible, the strong reductionistic thesis that each sentence of our language can be translated into a sentence of physics is true, the language of physics used as a worldmaking language will have to include implicit representation. But it is conceivable that propositions expressed by many sentences for which the strong reductionistic thesis fails would be *implicitly* represented by a world made with the language of physics and corresponding to our universe. Any proposition whose truth value supervenes on what can be described by physics would be thus implicitly represented.

A language is *propositionally complete* if and only if for every world made out of maximal consistent collections of sentences of that language and every proposition p , either p is true at that world or $\text{not-}p$ is true at it. For instance, if it is possible for there to be properties which do not supervene on physical properties, then the language of physics is propositionally incomplete. Unless we are willing to make do with partial worlds, bivalence ensures that the worldmaking language has to be propositionally complete. Note that while I shall continually talk of “propositions” in connection with linguistically-based approaches to modality, this usage does not presuppose any Platonic claims about there *being* any abstract propositions out there. Perhaps propositions are just a useful fiction, a useful way of expressing certain facts about language—e.g., facts about whether the language is complete or not. This issue is open. One of the advantages of the linguistic approaches is that they cannot remove this possibility.

Some of the criticisms of propositional approaches will apply to the linguistic approaches, and so now I shall only make the criticisms that do not apply to the propositional approaches as well.

2.2 Objection from alien properties

However, none of the languages available to us is propositionally complete. For, as Lewis (1986a, p. 159) notes, surely it is possible for there to be properties that cannot be reduced to the properties actually instantiated in our world. Just imagine the finite denizens of a world with elementary particles none of which exhibit electric charge. Electric charge being a *basic* property, none of their languages would be propositionally complete since no proposition reporting something's having that basic property would be implied by propositions expressed by sentences of their languages. But surely it is plausible that we are in the same position as the finite denizens of that impoverished world, just vis-à-vis some property other than charge that is instantiated in some world other than

ours. At the very least, clearly we cannot possibly know that we are *not* in such a position, and consequently we cannot know that the linguistic approach as based on one of our languages can succeed.

But perhaps the denizens of that world *can* specify electric charge. We bestow the name “electric charge” on that property which causes certain kinds of overt behavior of objects, paradigmatically the attraction of amber (Greek: *ēlektron*) that has been rubbed to pieces of fluff. Now there are two views of how this naming works. First, we could be picking (whether by ostension or definite description—at this point, it does not matter) out a particular complex of patterns of behavior of objects, and defining “electric charge” by the definite description “that property which causes those patterns.” This we shall call the “Russell-Frege view” of electric charge. Or, we could be picking out a particular complex of patterns (again, it does not matter how), and then defining “electric charge” as a rigid designator or name for that property which in our world causes those patterns. This will be the “Kripke view” of electric charge.

On the Russell-Frege view, necessarily, in any world in which those patterns of behavior that we have used to identify electric charge occur, that property in that world which causes these patterns of behavior will be “electric charge”. While on the Kripke view, it is at conceivable that there be a possible world with the same patterns of behavior exhibited but with a property different from electric charge responsible for them, just as it is possible for there to be a world with a liquid that behaves just like water but is not water because it is not H₂O.

If we adopt the Russell-Frege view of electric charge, then the example given fails. For the denizens of the impoverished world can describe the patterns of behavior that electric charge causes, and since, necessarily, wherever these patterns are exhibited, there electric charge is to be found, worlds described with their language have the resources for implicit representation of predications of electric charge.

If we adopt the Kripke view, then indeed the language in the non-electric world is propositionally incomplete. For it is then conceivable that there be two worlds exhibiting the same patterns of behavior, but in one world this behavior be caused by electric charge and in the other world by some other property. But is this in fact conceivable? Note that the language in the non-electric world is capable of specifying dispositional properties; nothing limits it to describing occurrent behavior only. Thus, we are asked to imagine two worlds, one of which has electric charge and the other some other property in its place, but the movements of things in those worlds and the phenomenological lives of their denizens are qualitatively the same *and* have the same counterfactual properties. If we think this is an impossible story, then we may be inclined towards the Russell-Frege view of electric charge.

But even if we adopt the Russell-Frege view of electric charge, can we use a similar approach for *all* properties? This seems implausible. Imagine a world impoverished by a total lack of electromagnetic radiation. This is a world without light. Is it plausible to suppose that finite denizens of that world could *both* express propositions that entail the existence of light *and* propositions that entail the existence of visual experiences? It is highly implausible that the Russell-Frege view, even if it works for the special case of electric charge, should work for both physical and phenomenal properties. But if it ever fails, then it becomes highly plausible that there are certain properties which no worlds made from our languages can represent.

Or consider finite beings in a world where there is no space. Could worlds made of their language represent spatial properties? Probably not. But we cannot know that there is no property to which we are related as those beings are to spatiality. Indeed, it is plausible to suppose there are such properties. Note that this objection applies even

But even if none of *our* languages are up to the task, perhaps there are languages that are. ^[109] Lewis discusses “Lagadonian” languages in which some words represent themselves. Thus, a stone might represent that very stone it is. A real number might represent itself. ^[110] If there are universals and basic properties are universals, then the world-making language might use these universals for representing themselves. It could even be an extension of our language. Suppose we represent types of sentences as set-theoretic sequences of characters (with each character being represented by a number, if we like, but I omit this for simplicity), so that “There is

snow” is set-theoretically represented as the set $\{ (1, \text{”T”}), (2, \text{”h”}), (3, \text{”e”}), (4, \text{”r”}), (5, \text{”e”}), (6, \text{” ”}), (7, \text{”i”}), (8, \text{”s”}), (9, \text{” ”}), (10, \text{”s”}), (11, \text{”n”}), (12, \text{”o”}), (13, \text{”w”}) \}$. Then a sentence saying that something has an alien property can be represented as the set $\{ (1, \text{”S”}), (2, \text{”o”}), (3, \text{”m”}), (4, \text{”e”}), (5, \text{”t”}), (6, \text{”h”}), (7, \text{”i”}), (8, \text{”n”}), (9, \text{”g”}), (10, \text{” ”}), (11, \text{”h”}), (12, \text{”a”}), (13, \text{”s”}), (14, \text{” ”}), (15, F) \}$, where F is the alien property. This presupposes a set theory with ur-elements, with universals being allowed to be such ur-elements.

This approach *does* answer the alien properties difficulty, but only at the cost of an ontological commitment to universals. And if one is willing to countenance *some* abstracta like universals, maybe it is better to go all the way for propositions. Of course, if one’s collection of universals is rich enough, there is little difference between an approach via universals and one via propositions. For, to any proposition p there corresponds the universal $U(p)$ which belongs to a cosmos if and only if p would hold in that cosmos. But then our language can express any proposition just by saying that the actual cosmos has $U(p)$.

2.3 Primitive modality

As Lewis (1986a, p. 151) notes, the linguistic approach does nothing to clarify what modality is about or to solve the Parmenidean problem, since it presupposes modality in two places: (1) in the requirement that the worlds be *consistent* collections of sentences, and (2) in the notion of implicit representation. For to say that a collection of sentences is consistent is nothing else than to say that these sentences are *compossible*. And to say that some collection of sentences implicitly represents a proposition is to say that the propositions expressed by these sentences jointly entail the latter proposition. But of course entailment involves modality since modal operators can be defined in terms of entailment: δp is true if and only if $(0 = 0) \text{ P } p$, while $\hat{a}p$ is true if and only if $p \text{ (} 0 \text{ } 1 \text{ } 0)$. This is only an objection against Ambitious Linguistic Ersatzism: the Unambitious Ersatzist will not mind it.

One might think one could get rid of implicit representation by positing a language sufficiently rich to express every proposition. If one believes that abstract propositions do exist, then of course it is trivially true that there is a language which can do this, namely the language whose basic symbols are the abstract propositions themselves, whose grammar specifies that each sentence consists of exactly one basic symbol, and whose semantics specifies that each symbol signifies that very proposition which it is. But if one does not believe in abstract propositions it is much difficult to see whether there is indeed a propositionally complete language.

One might also think that compossibility is not a big deal because it can be expressed syntactically. A collection of sentences is compossible if and only if there is no proof proceeding from these propositions according to certain predetermined inference rules and concluding with the negation of one of these sentences. If *both* this syntactic reduction could be done *and* the implicit representation issues could be handled at the same time, the linguistic approach would solve the Parmenidean difficulty. For then asserting that something is possible would be making an assertion about actual reality rather than speaking about nothing: it would be asserting that actual reality does not include a proof from that proposition to its negation. And, proofs are the sorts of things that exist in actual reality, being nothing but certain sequences of sentences, and sentences being nothing but certain mathematical objects (sequences of symbols, say) that, if we are realists about mathematics, had better exist.

Lewis (1986a, Section 3.2) argues, however, that one cannot both do the syntactic reduction and solve the implicit representation issues simultaneously. It is plausible that there are certain arrangements of elementary particles such that when they are present, necessarily a donkey is present. But it is implausible to think that the word “donkey” is to be understood as an abbreviation for a description in terms of arrangements of elementary particles. There is then no syntactic way of inferring from a sentence of the form “The joint wave-function of a bunch of particles here is Y ” that “There is a donkey here.” If the worldmaking language is rich enough to include both elementary particles and donkeys, this shows that syntactic definitions of consistency will not do, as there is no syntactic way of gauging whether “The joint wave-function of a bunch of particles here is Y ” and “There are no donkeys here” are consistent or not. If, on the other hand, the worldmaking language is not rich enough to include both, then there will have to be implicit representation which, for the same reasons, cannot be expressed

syntactically.

There are two possible responses. The first, which Lewis considers, is that perhaps there are a bunch of additional axioms such as “Necessarily, when the joint wave-function of a bunch of particles is Y , there is a donkey there” whose use is permitted in the syntactically based proofs. But the axioms are not mere definitions because the strong reductionistic thesis that one can *define* macroscopic objects in terms of microscopic ones is highly implausible, and so they are substantial. As such, these axioms participate in an unanalyzed primitive modality. And, contrary to Parmenides’ strictures, they are making assertions whose truthmakers involve non-existent things, namely non-actual wave-functions that are non-actual donkeys.

The second approach is to complain that these Lewisian examples rest on a weak reductionistic thesis that the presence of macroscopic objects like donkeys supervenes on the presence of elementary particles. One might take an Aristotelian view according to which the presence of a substantial form is a necessary condition for the existence of a donkey. It is not enough that there be elementary particles arranged in a certain way: it is necessary that there be a certain objective *telos* to their arrangement. It is, perhaps, an essential property of a donkey that it be the product of nature rather than an artifact, because being a product of nature endows it with a certain *telos* which is different from the *telos* that an artifact has, and for an Aristotelian the innate *telos* of something is an essential property of it. It is necessary to x ’s being a donkey that certain normative claims about x be true, such as that x *should* be four-legged, because satisfying these normative claims is part of the *telos* of a donkey, which *telos* is specified by the substantial form. Now, clearly, it is possible for there to be an entity whose arrangement of particles is just like a donkey’s, even a four-legged donkey’s, but where it is not true to say about this object *should* be four-legged. For instance, it is possible that this entity was a normal member of an alien species which otherwise is somewhat donkey-like and for which species it is normal to be six-legged, but the entity underwent some unpleasant operation which modified it, and its DNA, so it would resemble, particle-by-particle, a four-legged donkey. If, nonetheless, we can say that the resulting entity is still the same individual as before the operation, then it will arguably be abnormal for this entity to be four-legged. But since, necessarily, it is normal for a donkey to be four-legged, this entity will not be a donkey. Or, one might bypass all mention of normativity and simply rely on an intuition that being a donkey is an essential property while the individual after the operation is the same as the individual before, so that if the individual before wasn’t a donkey, neither is the one after.

However, even this very controversial objection will not save the Ambitious Linguistic Ersatzist. It only relocates the problem. Call an aggregate of particles a “*donkey” if the arrangement of particles in this aggregate is such that a normal donkey could be composed of these particles. Admittedly, not every *donkey is a donkey. In fact, no *donkey is a donkey because the criteria of identity for aggregates of particles are different from those for donkeys. However, Lewis’s argument can now be rephrased in terms of *donkeys. For it is still true that there being a certain arrangement of particles there entails that there is a *donkey there. And the same problems as before ensue.

The second objection against the Ambitious Linguistic Ersatzist to take a syntactic approach to compossibility is similar to an objection levied against EMR in Section 2.2.2 of Part III. One might well wonder why the truthmaker of the proposition that asserts Smith can perform A should have to involve syntactic claims about sentences. It seems natural to say that this proposition is just about Smith. Why should its truthmaker involve *sentences* at all? This leads to the next objection.

2.4 The arbitrariness objection

The Ambitious Linguistic Ersatzist claims that facts about sentences in a language are involved in the truthmakers of modal propositions. But facts about sentences in *which* language? Any particular human language we choose seems implausibly arbitrary. (This objection is much like the one made in Section 4.2 of Part I. Section 5 against Lewis’s quasi-linguistic structured propositions.) It is clearly absurd to think that facts about English sentences are involved in the truthmaker of a Frenchman’s, or a space alien’s, modal claim. Nor will it do to speak

of an idealized language, for there are many ways of idealizing a language, and any particular way of doing this will be just as arbitrary as any other.

One might attempt to consider some super-language which involves all actual languages. I have argued elsewhere [\[111\]](#) that in contexts in which the interlocutors know more than one language, there is in effect a super-language whose grammatical rules govern contextually how each utterance is to be understood, e.g., in the case of speakers who know German and English, whether “gift” is to be understood as signifying a present (as in English) or poison (as in German). Extending this, we can consider the totality of all communicative utterances as a single super-language. True, nobody knows all the words and grammar of that super-language, but then none of us know all the words and few know all the grammar of our native language. It is less implausible to suppose that such a language should be *the* world-making language. However, this, too, is parochial and arbitrary. For surely, in the spirit of S5, what the possible worlds are should not itself vary across possible worlds, even though what languages are actual does.

Thus, we should extend to considering a totality of all actual and possible languages. But, of course, there is a circularity problem here, since we were supposed to be elucidating possibility, and yet one is appealing to a certain class of possibilia, namely possible languages. The same objection can be made against a variant view according to which the world-making language has as its sentences equivalence classes of synonymous sentences in all possible languages. [\[112\]](#)

2.5 Mathematical constructions

One might also try to model possible worlds mathematically (see, e.g., Heller, 1998). For a Democritean model, suppose S is a set of basic physical objects, microscopic or macroscopic, and suppose, contrary to fact [\[113\]](#), that it is impossible for two objects to occupy the same point in space. One might then model a world as a function f defined on the product space $S \times R$, where R is the real line, that assigns to each member of this space an ordered pair consisting of a position in Euclidean three-dimensional space and an orientation.

A simple model of that sort will not do, of course. For one, different possible worlds will have different space-time structures, and using a Euclidean space will be inadequate—even for our world. Perhaps, though, part of our models can pick out the dimensionality of the space and the metric. But even so, the more serious objection is that not all properties are positional. First of all, it is surely at least logically possible that there exist immaterial entities, and we have no idea if these *can* be modeled mathematically. And even if such entities were impossible, nonetheless there are non-positional properties, such as charge, mass, and the like. Considerations similar to those applied in Section 2.2 make it implausible to think that we can find a model in which all possible properties find their place.

2.6 Combinatorialism and Armstrong

A somewhat similar approach is to construct possible worlds as recombinations of actual-world entities (Quine, 1968; Cresswell, 1972 and 1973; Armstrong, 1989; see the discussion in Lycan, 1994, Section 3.2). E.g., Armstrong constructs possible worlds as rearrangements of a subset of the objects of the actual world and of arbitrarily many duplicates of the objects of the actual world. Consequently, it is impossible for a property to be instantiated that cannot be reduced to properties found in the actual world. One way to object to Armstrong is to say that therefore S5 is violated, whereas S5 has intuitive plausibility (see Section 3 of Part I). For it is logically possible, on Armstrong’s account, that there exist a world w_1 in which there are no electrically charged entities. Assuming electric charge is a property that cannot be reduced to other properties found in w_1 , were w_1 actual, no rearrangement of the entities in w_1 and/or of their duplicates would yield a world with electric charge. Thus, were

w_1 actual, the existence of a charged entity would be impossible. But it is possible for w_1 to be actual, and hence it is possible for it to be impossible that there is a charged entity. But there is in fact a charged entity, and so the Brouwer axiom, and *a fortiori* S5, is violated.

However this does not worry Armstrong. After all, it *is* Armstrong's basic intuition that possible worlds are to be built out of the ingredients of the actual world: "The Combinatorialist ... is an actual-world chauvinist" (Armstrong, 1989, p. 56), and on this intuition, as Gregory Fitch as noted in conversation, it becomes plausible to think that what is possible depends on what is *in fact* actual, and hence that S5 is false.

But there is a much stronger argument against Armstrong, in that if his account is correct, then it is possible that there is a proposition p which is simultaneously physically possible and logically impossible, which is absurd. To see this, consider a world w_1 with laws of nature just like those of our world, but in which, although there are photons, there are never any charged entities like electrons or positrons. There are surely plenty of such worlds. Now, were w_1 actual, then by the argument above, on Armstrong's view, it would be logically impossible that there ever be a charged entity. However, the laws of nature holding in the actual world were assumed also to hold in w_1 , and according to those laws it is physically possible for a pair of photons to turn into an electron-positron pair. Hence, in w_1 it is physically possible that a pair of photons turn into an electron-positron pair, which would entail that a charged entity exists in w_1 at some time. Hence, in w_1 it is physically possible that there be a charged entity and it is logically impossible that there be a charged entity. But this is absurd, and hence Armstrong's account of possibility is wrong.

The one premiss here that Armstrong can dispute is the claim that there can be a world w_1 in which there are photons but no charged entities and in which the same laws of nature as hold in the actual world hold. On Armstrong's view, laws of nature are relations between universals. Thus, a law according to which a transition from a pair of photons to an electron-positron pair is possible is presumably some kind of a relation between universals including the universals *photonicity*, *electronicity*, and *positronicity*. However, in any world in which there are no electrons, the latter two universals do not exist, and hence neither does the law. Therefore, Armstrong's account of laws of nature implies that were there no charged entities, there would be no law of nature in virtue of which it is possible for a pair of photons to make a transition to an electron-positron pair. Indeed, as has been shown above, Armstrong *must* claim this if he is to avoid the absurd conclusion that something can be both physically possible and logically impossible.

However, this claim is itself absurd for two reasons, and hence cannot be maintained. First of all, for on Armstrong's rearrangement view, and likewise by our own intuitions about possibilities, it is logically possible that there be a world containing photons but no charged particles. However, plausibly, the property of *being physically able to produce an electron-positron pair upon collision with another photon* is entailed by being a photon. Armstrong's rearrangement theory of possibility entails that there is a world containing photons but no charged particles. But as seen above, Armstrong must also claim that in such a world it would not be *physically* possible for an electron-positron pair to be produced, which is then a contradiction. One way Armstrong could get out of the difficulty would be to claim that although being a photon may entail being physically able to produce an electron-positron pair upon collision with another photon, nonetheless being a photon is not an essential property. Thus, in the world we are imagining, there are entities that are identical with the photons of this world, though they are not themselves photons. This is of course rather implausible.

The second objection to the claim that in a world where there are no charged particles, there is no law of nature making it possible for a photon pair to make a transition to an electron-positron pair assumes the following plausible principle together with indeterminism:

- (45) For any possible world w and any collection of non-actual events which physically could have occurred in the history of w it is true at w that were this collection of events to have occurred, the actual laws of physics would remain unchanged.

This principle formalizes our intuition that the laws of physics govern what happens in the universe, and no event

which they make possible can affect their content. Of course, a Humean will likely deny this principle ^[114], but Armstrong is not a Humean given his acceptance of universals in which the laws are grounded. Assume (48). Imagine then a physically possible world w where there *eventually* come to exist charged particles but where initially there are only photons and in which world the same laws of physics as govern our world hold. It would be difficult for Armstrong to deny this possibility. ^[115] Then, if this world were actual, there would be a collection of non-actual physically possible events—indeterminism in the laws of nature might be necessary for this hypothesis—that are physically incompatible with the processes by which charged particles in fact come to be in w . By (48), were these processes to take place, the same laws of nature would still hold as in w . Hence there is a world which contains only photons but in which the same laws hold as hold in our world, and this is what our argument needs.

Section 3 Platonic approaches

3.1 Platonic approaches

On a number of Ersatzist accounts, possible worlds are Platonic entities, such as collections of propositions (Adams, 1974), maximal states of affairs (Plantinga, 1974), constructions from properties (Castañeda, 1974 and 1989; Parsons, 1974 and 1980) or *sui generis* Platonic entities (Stalnaker, 1976). I will specifically focus on the Adams propositional account, though the criticisms made will apply to all theories that explain possible worlds in terms of Platonic abstracta.

3.2 What are propositions

One can introduce, though not define, the theoretical entities called “propositions” and the relation of “expressing” between assertings and thinkings, on the one hand, and propositions, on the other hand, by (a) the claim that our assertings and thinkings express propositions and precisely one proposition is expressed by any one asserting or thinking, (b) the claim that synonymous assertions are precisely those that express the same proposition, (c) the claim that one says what one thinks precisely when one’s asserting expresses the same proposition as one’s thinking, and (d) by using the idea that the intuitive notion of the “content” of a given asserting or a thinking is the same as the notion of the proposition being expressed by the asserting or thinking. Moreover, to further expand the theoretical account which introduces “propositions”, one can tell a story about a relation between propositional attitudes, such as that between the attitude of *desiring that p* and the proposition *that p*, analogous to the relation of “expressing”. To give an account of truthmakers, one can say that when an asserting is correct, it is correct in virtue of there being something in the world (e.g., the assertion “Socrates died” is correct in virtue of our world being one in which Socrates’ dying existed), and in that case the proposition expressed by the asserting is said to be *true* and the item in the world because of which the asserting is correct is said to *make* the proposition true, be a truthmaker for the proposition.

But not all propositions are actually expressed by an assertion we make or an act of thinking we engage in. However, given the intuition that there are things that we *could* mean or express but in fact which we do not mean or express, and given the Platonic intuition that these things are not different in kind from those things we in fact mean or express, we conclude that the class of propositions is wider than the class of things that our assertions in fact express. Moreover, it is plausible to suppose that there are some things that none of us is capable of expressing: after all, it would be surprising to think that our expressive capacities should be capable of covering the collection of all possible things that can be expressed. These considerations lead us to realize that there are propositions that we cannot even express.

It is plausible to suppose, then, that propositions are *necessary* entities, since they do not appear to depend

on our existence or on the existence of other contingent expressive beings, and it is not at all clear what else they could depend on. We can do better than this plausibility argument in the case of some propositions, namely those expressing necessary truths. It is plausible to suppose that what is expressed by “ $2+2=4$ ” would be true no matter what. Let p be the “what is expressed by ‘ $2+2=4$ ’”. Then, p is true no matter what. But then, in particular, p exists no matter what, since Parmenides tells us we cannot talk of what does not exist, and in particular cannot say that what does not exist is true. Hence, p is a necessary being. But in fact a similar argument can be made for all propositions. For, if p is a proposition, then p is true or p is false, no matter what. But what does not exist cannot have properties, by Parmenides’ stricture, and hence cannot be either true or false. Hence, no matter what, p exists.

The collection of propositions then far outstrips us. We *introduced* them as theoretical entities which are the things we express by our claimings and the things that synonymous claimings have in common. But this introduction does not exhaust the collection, any more than the fact that we may have introduced the collection of electrons for explaining a bunch of phenomena in this galaxy means that we have no right to suppose electrons in other galaxies. Note also that the collection of propositions is not a set (see Grim, 1986; Chihara, 1998, p. 125ff).

It is reasonable to suppose that all propositions stand in logical relations to one another, since only then do they belong to a kind of things which can be meant or expressed. The ersatzist can now say: “Let us consider maximal compossible collections of propositions and call them ‘worlds’.” But wait: It must first be demonstrated that there are such collections.

Well, first of all, we know there is at least one maximal compossible collection of propositions, namely the collection of all propositions that are actually true. This collection is maximal, because if p is a proposition, then either p is actually true or actually false. If p is actually true, then p will be a member of the collection. If p is actually false, then $\text{not-}p$ will be a member of the collection, and hence p could not be added to the collection without introducing inconsistency.

But perhaps there is only one such collection? This will not make the ersatzist happy. In order for modality to be definable in terms of quantification over possible worlds, the ersatzist needs to claim that for any possible proposition p there is a world containing p . This is a nontrivial claim. One way to try to prove this claim would be to take a union of increasingly larger consistent sets; unfortunately, such approaches fail since the union of consistent sets can be inconsistent, e.g., consider the consistent sets $\{x \text{ is finite, } x > 0\}$, $\{x \text{ is finite, } x > 0, x > 1\}$, $\{x \text{ is finite, } x > 0, x > 1, x > 2\}$, ..., whose union is evidently inconsistent.

Another approach would be to argue for the existence of a maximal collection using “subjunctive conditional bivalence” which says that for every possible proposition p and every proposition q either:

(46) Were p true, q would be true

or:

(47) Were p true, $\text{not-}q$ would be true.

It is not difficult to see that given conditional bivalence one would have the desired conclusion that every possible proposition is contained in a maximal collection of compossible propositions, viz. those propositions which would be true were it true. Unfortunately, subjunctive conditional bivalence is false. It is logically possible that there now be a unicorn in this room. But there is no fact of the matter as to how many hairs such a unicorn *would* have. Thus no counterfactual of the form

(48) Were there a unicorn in this room, it would have n hairs

is true. For a *reductio*, suppose subjunctive conditional bivalence is true. Then it follows that every counterfactual of the form

(49) Were there a unicorn in this room, it would not have n hairs

is true. But the conjunction of subjunctive conditionals with the same antecedent entails a subjunctive conditional with the same antecedent and having as its consequent the conjunction of all the consequents. Therefore, since all propositions of the form (52) are true, it must be true that:

(50) Were there a unicorn in this room, then it would be true that " n (the unicorn does not have n hairs).

But the consequent of (53) is impossible, and no subjunctive conditional can be true if its antecedent is possible and consequent impossible. Hence we have a *reductio* of subjunctive conditional bivalence.

There is, fortunately, another construction that does work. Let P be the collection of all propositions, let F be the collection of all functions from P to $\{true, false\}$, and define v as a short form for the definite description [116] “the unique function such that for every proposition p , the function’s value at p is *true* if p is true and *false* if p is false”. Note that “ v ” is a non-rigid designator of a function in F . Thus:

$$(51) \quad \delta \ \$f (f \text{ is in } F \text{ and } v = f).$$

I.e., necessarily, there is a truth-value assignment in F which matches the truth-value assignment that reality provides. To avoid unnecessary confusions about merely apparent contingent identity claims between v and f , define a predicate V which holds of a member f of F if and only if it is the case for all p that $f(p) = true$ if and only if p is true. Then, (54) becomes:

$$(52) \quad \delta \ \$f (f \text{ is in } F \text{ and } Vf).$$

But now suppose p is possible. Then, since the conjunction of a possible truth with a necessary truth is also possible, it follows from (55) that:

$$(53) \quad \grave{\text{a}}(p \text{ and } \$f (f \text{ is in } F \text{ and } Vf)).$$

Hence:

$$(54) \quad \grave{\text{a}}(\$f (p \text{ and } f \text{ is in } F \text{ and } Vf)).$$

Thus, since p and Vf entails that $f(p) = true$ by definition of the predicate V :

$$(55) \quad \grave{\text{a}}(\$f (f(p) = true \text{ and } f \text{ is in } F \text{ and } Vf)).$$

But now functions are necessary entities and one can interchange the order of quantifiers over necessary entities and modal operators so that:

$$(56) \quad \$f (\grave{\text{a}}(f \text{ is in } F \text{ and } f(p) = true \text{ and } Vf)).$$

Now let f_0 be a rigid designator of a function such that it is possible that f_0 is in F and $f_0(p) = true$ and Vf_0 holds. Then, in fact $f_0(p)$ is equal to *true* since rigidly designated functions have their values necessarily, and also in fact f_0 is in F since collection membership if possible is necessary for rigidly designated collections. Thus, $f_0(p)$ is *true* while Vf_0 is possible. Now, let C be the collection of all propositions q such that $f_0(q)$ is *true*. Then, because Vf_0 is possible, C must be compossible, since, necessarily, Vf_0 holds if and only if every member of C is true. And, because $f_0(p)$ is *true*, p is a member of C . Moreover, C is maximal, because for every proposition q if q is not in C then $f_0(q)$ is *false* and hence $f_0(\text{not-}q)$ is *true* and hence $\text{not-}q$ is a member of C .

Thus, the propositional ersatzist seems to be entitled to her maximal compossible collections of propositions. Since, moreover, we have shown that any possible proposition is contained in a maximal compossible collection of propositions and evidently no impossible proposition is, it is true that a proposition p is possible if and only if it is a member of some world. But of course this cannot be intended as an analysis of possibility, at the pain of vicious circularity as the concept of possibility is presupposed by the concept of compossibility.

One can also vary the above construction slightly and define worlds as conjunctions of the propositions in a maximal compossible collection of propositions. [117]

While neither construction yields an analysis of possibility, the ersatz propositional worlds are very useful for conceptual purposes. At the very least, they are useful for making complex modal arguments clearer. [118] But, furthermore, they do provide a conceptual tool that mere modal operators do not, at least not in a straightforward way: these ersatz worlds can be used for analysis of supervenience, world-similarity and counterfactuals just as Lewis’s worlds can.

3.3 The set-theoretic objection

One might worry that the collection of all propositions is an impossible collection. After all, it seems it is

not even a proper class. For, there is no proper class of all proper classes, while there is a different proposition corresponding to every proper class, viz. the proposition that asserts that the class is self-identical (cf. Grim, 1986). However, this objection is not fatal. One can distinguish theoretical levels, with the class of all “propositions” being considered on a metatheoretic level, without these “propositions” being allowed to concern classes at the metatheoretic level. Nothing relevant is lost in restricting propositions to concern things at a lower level, because contingency is rooted at a lower level (e.g., the level of first order propositions about the world). Alternately, one can try to find a different set-theoretic axiomatization for which things would work out.

3.4 Kripkean objections

One objection to this account is Kripkean. Recall Kripke’s argument to the effect that one cannot say that unicorns are possible. For, unicorns would be a specific natural kind, and one cannot possibly specify what natural kind unicorns are, since such kinds are to be specified by ostension. Taking this argument a little further, one might want to claim that some propositions are not necessary entities. The proposition that something is a member of some natural kind on this view does not exist outside of a world where this natural kind is instantiated. But if possible worlds were defined in terms of those propositions that exist in *our* world, there being none other we can use without falling afoul of the Parmenidean objection, then there is something descriptive lacking to these possible worlds—namely, that which would be expressed by the propositions that do not exist in our world.

This objection has deep Aristotelian roots. Recall that according to Aristotle, to claim that something is a member of a natural kind is to predicate its essence of it. But unlike Platonic Forms, Aristotelian essences are bound up with the individuals that they are essences of, and the only way they exist apart from these individuals is in the minds of the persons who have abstracted these essences from the individuals they are in. But if this is so, then there will be no essences of non-exemplified species, and therefore no propositions reporting membership in these species. Therefore, there are propositions in other worlds that do not exist in ours if there are possible species that are not exemplified in our world. Aristotle himself did think that it was necessary that precisely those species be exemplified that are exemplified, but this is implausible given the contingencies involved in evolution.

One reply to Kripke could be in terms of genetic reductionism. Let H be the set of all possible sequences of DNA that a horse can have. Then, arguably, necessarily anything that has a DNA sequence in H is a horse and nothing else is. But if this is so, then it is plausible to say that the claim that x is a horse expresses the same proposition as the claim x has a DNA sequence in H , just as the claim that x is water expresses the same proposition as the claim that x is H_2O . Now, for any set S of DNA sequences and any designator x of an individual there is a proposition reporting that x ’s DNA sequence is in S . But if saying that x is a horse is nothing but saying that x ’s DNA sequence is in H , then analogously any alien DNA-based species will have its species membership expressible by a claim of the form “ x ’s DNA sequence is in S ” where S is some set of DNA sequences. But those *sets* of DNA sequences that exist in an alien world will also exist in our world, though they may be unexemplified, and hence so will propositions reporting membership in this species.

There are two difficulties with this approach. The first is that it fails for a more radical Aristotelian who will deny that any kind of reductive account can be given of species membership. For instance, species membership is in part a normative notion, and one cannot get to normativity from a mere description of atomic positions.

But unless one makes Aristotelian forms into completely mysterious things, there is good hope that the normativity in them can be specified in terms of necessary propositions: “A horse is a being that *should* seek to reproduce its kind, that *should* have DNA structure in H , that *should* have four legs,” Normativity itself does not make for an impossibility of propositional expression. In fact, quite the contrary—often it is easier to describe exactly what something is supposed to be than exactly what it is. If I intend to draw a circle, then it is easy to describe exactly what I intended to draw: the aggregate of possible pen marks in the plane equidistant from some fixed center. But to describe the exact irregularities of the shape drawn is an impossible task for us finite

[119]

humans.

The more serious difficulty is that the account will not work for completely alien properties. We can talk of alien DNA-based natural kinds, but what about alien natural kinds based not on DNA and ultimately carbon, but on some completely alien particles equipped with completely alien properties? Here the propositional ersatzist needs to stand her ground. The above account has made plausible that there is no in principle bar against there being propositions in our world reporting membership in some alien world's natural kind. The present objection neglects the fact that propositions are not the things that we can express, but are things of the same *kind* as the things we can express. There are doubtless many kinds of things that we cannot express. But unless one has some general principled view that membership in alien natural kinds cannot be implicitly represented by propositions in our worlds, which general view is refuted by the genetic argument, there is no reason to think that the class of propositions and the class of things we can express is the same. Against this is the intuition that whatever had turned out, every possible proposition would still have been possible (this is just S5), and hence would still have been a proposition.

3.5 How do propositions represent?

David Lewis discusses propositional ersatzism very cleverly (1986, Section 3.4) First he describes a variant of propositional ersatzism without ever using the word "proposition". There just is some collection of entities with certain formal properties and relations, namely those that propositions have but without their usual names like "reporting", "being true", etc. He then wonders why this collection should have anything to do with possibility, and asks whether calling the members "propositions" and bestowing the usual names on their properties makes it any less mysterious what these entities are and why they have anything to do with modality. A special case of this puzzlement is the wonder of how it is that the propositions represent. What is it that makes some entity p represent there being unicorns as opposed to representing there being horses?

However, as I suggested, propositions should be thought of as theoretical entities analogous to those brought in by science. Forget what physics you know for the moment, while remembering mathematics. Imagine that I gave an account of the following sort. There is a set S of entities each member e of which has four quantifiable properties, which we shall call $u_1(e)$, $u_2(e)$, $u_3(e)$ and $u_4(e)$, and binary relations between them which I shall neutrally call "binrels" that ensure that the mathematical relations

$$\frac{\partial^2 u_i(e)}{\partial t^2} = \sum_{e' \in S - \{e\}} \frac{u_4(e')(u_i(e) - u_i(e'))}{((u_1(e) - u_1(e'))^2 + (u_2(e) - u_2(e'))^2 + (u_3(e) - u_3(e'))^2)^{3/2}}$$

are satisfied for $i=1,2,3$.

Suppose I then said that this explains various astronomical phenomena. Obviously, as far as the account went, my claim would be crazy (if you do not think it is crazy, this is only because you recognize what interpretation the quantities should be given). The account as given explains nothing because it is not sufficiently interpreted: there are many collections of entities that one could imagine singling out that satisfy this relation, and few of them explain astronomical phenomena. It is only when I further specify that $u_1(e)$, $u_2(e)$ and $u_3(e)$ are position coordinates and $u_4(e)$ a mass, all expressed in appropriate units, that this account becomes at all useful. (I can then go on and say that the "binrels" are gravitational forces, though this will not be helpful to my interlocutor if my interlocutor has never before met the concept of a "force".)

Suppose I now make a claim like that which Lewis makes about propositional ersatzism in this case: to say that $u_1(e)$, $u_2(e)$ and $u_3(e)$ are "position coordinates" and $u_4(e)$ a "mass" does not make the account in any way more explanatory or clear. Clearly, my claim will be out of line, because prior to my giving the account we had a certain grasp of position coordinates and of mass. (If we did not, then of course the criticism would have been perfectly just.) Likewise, claiming that it is something magical about some property that makes it a "mass" is out of line.

But much the same defense can be made of propositional ersatzism as of Newtonian gravity. Prior to

introducing worlds built up out of propositions we had a concept of a “proposition”, as a generalization from the concept of that entity which is what our language express. Just like “mass” and “position” had a prior role in our language game, and were introduced for good explanatory reasons, so too “proposition” had a prior role in our language game and was introduced for a good explanatory reason. Thinking that propositions represent “magically” is just like claiming that gravity is magic (not that that hasn’t been claimed, e.g., by Leibniz).

This does not mean that there is no real question asking what it is about propositions that makes them represent certain states of affairs. There *is* a substantial question there, just as there is a substantial question as to what it is that makes “mass” and “position” interact in such ways as satisfy the laws of gravitation. However, given how propositions were introduced, there is no problem with thinking that they *do* represent. This is something that the defender of propositions is entitled to, and there is no requirement that she give an account of how propositions represent prior to her being entitled to propositions.

This reply to Lewis is why it was essential that I first introduce propositions as fulfilling a different theoretical role from the one that the propositional ersatzist needs them for, just as it was essential that notions of “mass” and especially “position” have had a role to play prior to the advent of the Newtonian theory of gravitation.

Furthermore, there is an *ad hominem* answer to Lewis’s concern (van Inwagen, 1986; Jubien, 1991). His own theory admits abstract entities such as sets. But the relation between a set of concreta and the concreta it is a set of is no less mysterious than the relation which would hold between a proposition and its truthmaker, were the proposition true. To the question of why such and such an abstract entity represents there being horses rather than there being dogs one may reply with the question of why such and such an abstract entity is related in the set-to-members way to cows rather than to chairs.

There is, however, a certain point to Lewis’s objection. It is true that the ersatzist is entitled to assuming propositions can represent. But the claim that propositions can represent may be incompatible with further claims that the ersatzist may wish to make about propositions, such as that they are self-standing Platonic entities, just as the claim that material things have mass may be incompatible with a geometricist’s claim that material things are nothing but subsets of a Euclidean space-time. This particular objection will be examined more carefully in Section 1.1 of Part VI.

3.6 ***What do modal propositions represent? The unmet Parmenidean challenge and the need for an analysis of possibility***

However, the Ambitious Propositional Ersatzist ought not rest satisfied with the above answers to Lewis. For there is still Parmenides’ challenge to be faced. What real thing are we talking about when we make modal affirmations? The answer our ambitious ersatzist offers is: propositions. But what is it about these propositions to make them suitable for being the truthmakers of modal claims?

This is only a question for the Ambitious Ersatzist. The Unambitious Ersatzist can simply reply that the ersatz worlds constructed out of propositions *model* the truthmakers of modal propositions but need be involved in these truthmakers. But the Ambitious Ersatzist insists that these ersatz worlds, or at least their propositional ingredients, are involved in the truthmakers of modal propositions, and hence provide an answer to Parmenides.

Lewis raises this problem under the head of “primitive modality.” Our ersatzist has presupposed modality in the very definition of her worlds, when talking of compossible collections of propositions. It is Lewis’s contention that the EMR-theorist can do without *any* primitive modality, and if this is right, then a view that presupposes primitive modality is less theoretically desirable. However, I have argued in Section 4.2.1.b of Part III that even Lewis needs to presuppose some primitive modality, and my preferred account of possible worlds will also involve irreducible primitive modality. Thus, in order to make a “primitive modality” objection against the ersatzist without opening oneself for a *tu quoque* reply, one needs to explain what is particularly objectionable about the ersatzist’s primitive modality.

The answer to this is that the ersatzist fails to give any illuminating answer to Parmenides. If Parmenides’

argument against modality is to be escaped, something existent must be pointed out which can plausibly be said to be the truthmaker of modal claims. The Ambitious Ersatzist has pointed out the necessarily existent entities known as “propositions”. When we say it is possible that there are unicorns, we are predicating the property of being possible of the proposition that there are unicorns. Perhaps this is all that can be said, but if so, this state of affairs is disappointing. An account that could say more would be preferable theoretically.

But what kind of a property is this property? What is it about the proposition that makes it true to say that it has it? Syntactical characterizations of possibility being ruled out by arguments like those in Section 2.3, the only option for the ersatzist is to take some or all of these modal properties to be unanalyzable basic properties of propositions. But it does not provide an illuminating answer to Parmenides to say that the proposition that unicorns are possible is made true by the property of possibility which belongs to the proposition that there are unicorns. For it is at this point that the defense of the ersatzist made in Section 3.5 fails. Just as the Newtonian account does not answer the philosophical question of what the gravitational relation ontologically consists in, so neither does the ersatzist’s story answer the question of what the property of possibility ontologically consists in.

The problem is that as soon as we say that the truthmaker of the proposition that it is possible that unicorns exist consists in the proposition that unicorns exist having the property of possibility, we realize that we have not provided any illumination. To grasp what it means for the proposition that unicorns exist to have the property of possibility simply *is* to grasp what it means for it to be possible that there are unicorns.

A related worry is as follows. We can deduce that it is possible that there are horses from the fact that there are horses. Let p be the proposition that there are horses. Then, the claim is that possibility is a basic property that p has (it might be that one takes necessity to be a basic property, with possibility defined in terms of it; but *mutatis mutandis* the same objections can be made, so this is an option I shall not explore further). But how is p ’s having this property connected with p ’s being true, i.e., with p ’s having a truthmaker? In general, necessarily, whenever a proposition p has a truthmaker, then p has the property of possibility. One might think therefore that there should be some connection between the truthmaker and the property of possibility. But it is inexplicable on the “basic property” view of possibility what this connection is. The truthmaker of the proposition is, often, a concrete contingent entity while the proposition itself is an abstract Platonic one that cannot causally interact with the truthmaker.

That fact that propositions cannot be actually true without being possible but can be possible without being actually true suggests that having the property of actual truth is ontologically dependent on something related to what makes it have the property of possibility. But what makes it the case that proposition is actually true is nothing else than its truthmaker, something about the actual world: it is only on the truthmaker of p that p ’s property of actual truth depends. Now, the truthmaker of a proposition is not dependent on any abstract entities, since it is a concrete object in the world, being dependent at most on its causes—and abstracta are not causes. Now, if the property of possibility is a *basic* property of propositions, and hence if it is an abstract fact about something abstract that a proposition is possible, surely then no concrete worldly entity such as a horse can *depend* on some proposition’s having this property. But now we have a problem: actual truth here ontologically depends on possibility, an abstract property of an abstract thing, but actual truth depends only on the concrete truthmaker and on its concrete causes, if any. The truth of the proposition that a horse exists depends only on the horses of this world and on their causes. Therefore, it cannot be that the property of possibility consists in an abstract fact about abstracta.

One place to attack in this argument is the claim that a proposition’s being true should ontologically depend on its being possible. This dependence is brought in as an explanation of the fact that no true proposition can fail to be actual. Perhaps an alternate explanation could be given. But it is not at all clear what other explanation could be given within the framework of the assumption that possibility is a basic property of propositions. If one thinks that possibility is a property that can be analyzed in terms of concrete things about this world, as I shall eventually argue it can, then there is less of a problem, because there need then be no categorical

difficulty in seeing how the existence of a truthmaker could depend on that worldly property that possibility is.

All of the difficulties I am now raising illustrate that the “basic property of propositions” account of possibility fails to be an illuminating answer to Parmenides’ worry. A yet different way of giving voice to this concern is as follows. Many ordinary statements of possibility, if they are statements about anything, seem to be statements about *concrete* aspects of this world. When I say Hitler might never have been born, I am making a claim about the contingencies in the reproductive process, the potentialities for the miscarriage of children, and the like. When I say Smith could have done something, I am speaking about Smith’s own abilities. To claim that a proper analysis of these assertions will make them be ultimately statements about abstracta is to raise the worry: So why do we *care* about these statements? Why is it relevant to figuring out whether Smith is actually guilty or innocent to figure out whether some abstract proposition has the basic property of possibility? (Cf. Section 2.2.2 of Part III.) There is a force, then, that pulls us in the direction of thinking that assertions of possibility are assertions about concrete things in this world. But at the same time, that line of thought seems to run up against Parmenides’ objection leaving us in apory.

Roy (1993) gives an account that attempts to explain what a Platonist can ground modality in. Properties have inner structure so that “for example, we might suppose BEING COLORED is BEING RED OR BEING BLUE OR ... and/or that BEING WATER has among its constituents BEING HYDROGEN and BEING OXYGEN” (p. 342). Thus, necessarily, if BEING WATER is instantiated, likewise BEING HYDROGEN and BEING OXYGEN are instantiated, and necessarily anything that is red is colored. One difficulty with this view is with *de re* modality. Suppose that x has BEING WATER *essentially*. We want to say that x could not fail to contain hydrogen. On Roy’s account this is to be analyzed by pulling out another Platonic entity, *the essence of x* , which x essentially has nothing else has at all, and saying this entity is structurally related in an appropriate way to BEING WATER, which in turn is related in another appropriate way to BEING HYDROGEN.

Now, *the essence of x* is some property F . But why, one can ask, is it the case that x has F *essentially*, as opposed to having it merely non-essentially? One is tempted to say that the fact that x has F essentially is evident if we spell out what F is: We could just say that F is *the property of being x* —indeed that is just a paraphrase of “the essence of x ”. But this property appears to have no inner structure, and the structural account above seems to fail then.

In fact, we have a dilemma. Either F has as one of its ingredients BEING WATER or it does not. If it does not, then perhaps F is something structureless like *the property of being x* and it is unmysterious why x has F essentially. But then Roy’s account fails for the *de re* modal claim that x could not fail to be water. But if F does have BEING WATER as an ingredient, then presumably, on the model of Roy’s account of BEING COLORED, F is a property of the form: BEING WATER AND G , where G contains all the other ingredients of F . But now it has once again become quite mysterious why x has F essentially. Roy wants to explain all modalities in terms of structural facts about the Platonic realm. It is quite unclear how it is that having BEING WATER as an ingredient could explain F ’s being had by x essentially. Nor does it seem promising to say that G is some property of a sort that makes it clear why x has F essentially; for while it might be that having G would explain why x has G essentially (e.g., if G were *the property of being x*), it is quite mysterious how it is that having G would explain why x has both G and BEING WATER essentially. And it is mysterious how putting both BEING WATER and G together and saying x has both would explain why x has these two properties essentially.

In fact, things can get even more mysterious than that. There are two ways of a property being essential. A property can be non-essentially essential, meaning that it is possible that something has it essentially and possible that something has it non-essentially (e.g., a lump of plutonium has *containing plutonium* essentially but a human may have it non-essentially), or it can be essentially essential, meaning that necessarily anything that has it has it essentially (e.g., *being a lump of plutonium* might be an essentially essential property—perhaps, necessarily, anything that is a lump of plutonium is essentially a lump of plutonium). The property F is an essentially essential property, since, necessarily, anything that has it is x and x has it essentially. We can then ask: What is it about F

that renders it an *essentially essential* property? It is an essential property of *F* that nothing has it non-essentially. But we will fall prey to a vicious regress if we account for this modal claim using Roy's method, by positing *the essence of F*, and so on. The regress will be vicious because that for which we were attempting to account will reappear at each stage of the regress. (Compare this to the argument in Adams, 1981.)

One might of course say that a given thing has two ways of instantiating a property. It can instantiate it *essentially* and it can instantiate it *non-essentially*, and these two ways of instantiating we can take to be a primitive distinction. But once we say this we are no longer engaging in the Platonic project of explaining modality in terms of the structure of the Platonic realm—we are producing an account much closer to the Aristotelian one I will advocate in Section 2 of Part VI.

Alternately, a supporter of Roy's Platonic account might wish to get rid of *de re* modality by denying the existence of essential properties. Because of the arguments in Section 4.2.1.b of Part III, however, this does not seem promising.

3.7 Affirmations of necessary existence

If there are propositions, then there are entities which are necessarily existent. One might object to this on the Humean grounds that anything that can be imagined as existent can be imagined as non-existent. However, there are straightforward counterexamples to Hume's claim. E.g., one of my undergraduate students once observed that the totality of all existent things is a counterexample. There not being anything is not a possible state of affairs. For suppose that it were a possible state of affairs. Then, were it the case that there wasn't anything, this possible state of affairs would *be* the actual state of affairs. But then there would *be* something, viz. a state of affairs. Another way of putting this is: There is no possible world at which there is no world.

3.8 What are propositions?

One might of course object to the very idea of propositions on the grounds of their ontological obscurity. This objection would cut just as much against Unambitious as Ambitious Propositional Ersatzism. Propositions are entities, some of which—namely, the true ones—are connected with extant states of affairs, which are worldly entities that exist when and only when the proposition with which they are connected is true, with the connection being a made-true-by relation. They are immaterial, because they are necessary entities and it is one of our basic modal intuitions that nothing material is necessarily existent. They are unchanging. Moreover, if we think that temporality itself is a contingent feature of the world, then we will have to say that propositions are atemporal.

These Platonic features may seem ontologically extravagant in a way that EMR was not, because although EMR had a plenitude of concrete things, the propositional ersatzist has new items that are different in kind from the concrete things. However, this is not a fair objection. First of all, our ersatzist has not brought in the propositions just for purposes of making modal claims true. Rather, she initially posited them for a different explanatory role. Secondly, Lewis's own theory also involves Platonic abstract entities, such as sets, collections, abstract relations like the counterpart relation, and the like.

One might, however, have an objection against the way that propositions are brought in. Recall that Aristotle's main objection against Plato's forms was that they did not *do* anything: they lacked causal efficacy. One might argue that entities can only be brought in for explanatory purposes when they are *causally* relevant to the things of this world. Otherwise, one could imagine these entities disappearing and the things of this world continuing to exist with nothing being different. But propositions are not causally relevant to anything—that is, in fact, one of the main characteristics that defines them as *abstract*. So away with them! Observe that brave indeed must be the philosopher who levies this objection, because it does away with mathematical objects, too.

This objection can be put in two ways. The first way is as an Ockham's razor type of objection. Propositions are not causally relevant to anything, so there is no reason to believe in them. A standard reply to this

objection is that causal explanation is only one of many kinds of explanation. E.g., there is such a thing as mathematical explanation, and it is not causal. There is also such a thing as philosophical explanation, which may or may not be causal, and propositions are brought in through such an explanation.

The second way of putting the objection from the causal superfluity of propositions is to argue that we can only meaningfully refer to entities that have some causal influence on things we observe. We can meaningfully refer to electrons because we can pick them out as “those entities that cause such-and-such patterns of behavior.” But propositions do not cause anything and hence we cannot meaningfully refer to them. We cannot, for instance, individuate them from other abstract entities that might exist. However, it can be argued that the principle that we can only meaningfully refer to entities that have some causal influence on things we observe should in fact be replaced by the weaker principle that we can only meaningfully refer to entities that have some explanatory role in the things we observe, with “observe” being understood in a very wide sense. Propositions do play an explanatory role related to the meaningfulness of sentences, and the meaning of sentences is something that, at least according to McDowell’s plausible account, we perceive. The restriction to causal explanation is not justified here.

Moreover, is it really a category mistake to think of propositions as having a causal role? Unless our minds are purely physical, is it absurd to think that propositions might influence it? Perhaps propositions are also the sorts of things that can be found in our minds (in more than one mind at a time, if necessary), playing a causal role in our mental life. Despite Aristotle’s claims, it *is* plausible to suppose that the sun-analogy in the *Republic* gives the Form of the Good a causal role to play in our mental life.

It does not appear, then, that one can argue against the notion of a proposition just on the grounds of “abstractness”.

Section 4 Conclusions

Linguistic ersatzism fails, whether in an ambitious or an unambitious form, unless the language is something like the “language” whose terms are propositions or universals or the like. The main objection against linguistic ersatzism is that it cannot take account of alien properties.

The objections against Unambitious Propositional Ersatzism fail. That propositions are abstract is no reason to disbelieve in them. That it is not clear *how* they represent the states of affairs they represent is also no objection, just as it is no objection to various scientific views that they fail to explain how it is that certain objects have properties like mass.

However, Ambitious Propositional Ersatzism does not succeed completely, because it fails to give a satisfactory account of what we are talking about when we are making modal claims. It either is completely unilluminating in connection with Parmenides’ problem, simply moving the bump in the carpet to the problem of what it is about a proposition that makes it have the property of “possibility” or “necessity”, with this being exactly the original problem in different terminology. Granted there is an abstract proposition that there are unicorns, and another abstract proposition that it is possible that there are unicorns. But what makes the second proposition true? Well, what makes it true is that the first proposition is possible. And in virtue of what do we say the first proposition is possible? In virtue of it being possible that there are unicorns. Or, if our ersatzist more ambitiously insists that “possibility” is a basic unanalyzable property of abstracta, the account fails to explain how *this* particular basic property has anything to do with our ordinary claims about possibility, what its connection to actuality is, and how it is relevant to concrete facts about the world.

Part V. Rescher's critique of the very notion of possible worlds

Section 1 The argument

Rescher (1999) has recently criticized the very notion of a complete possible world, holding out the possibility of partial “schematic” worlds. The argument begins with the observation that “real particulars”, with “real” understood as opposed to “merely possible”, are descriptively inexhaustible. For one, we can observe the contours of a physical object need a description that specifies infinitely many coordinate positions, and we finite humans could not possibly produce such a description. But the problem is deeper than that: it isn't just a question of the messiness of physical objects. For, physical objects have dispositional properties. And not all dispositions are going to be realized, even over an infinite amount of time. Specifying all the dispositional properties is a clearly impossible task.

Now consider a possible individual (or a possible world, for that matter—a world being after all, from a certain point of view a certain kind of individual).

To validate the idea that it is a genuine individual that is at issue in a discussion of ours we must be in a position to indicate to our interlocutors just which individual is in view: we must specify the particular

individual that is the object of our discussion. ... *Particularity demands identification* [\[120\]](#)

But no individual that is of a kind that could be real can be specified completely, whereas surely a complete definite specification would be necessary for individuating possible individuals. For it is not enough to say we are talking of a man just like me but with green hair: the exact shade needs to be specified, as do various dispositional properties of the dye, as well as the exact causal process by which the man came to have green hair. For, if there are possible individuals, different individuals correspond to different values of these parameters. So we cannot individuate possible individuals. However, without individuation, there is no identity, and without identity, no entity. Hence, there are no possible individuals that we can meaningfully talk about.

Section 2 The individuation principle in the argument

However, in fact, many of us *do* believe in entities where individuating criteria cannot possibly be given by us. For instance, many physicists believe in a continuum consisting of uncountably many points. It is impossible *for us* to specify each one of these points in a way that sets it apart from all others. There being countably many definite descriptions in our language, we can only specify countably many points. Most of the points of a continuum can only be talked about set-wise (e.g., “the set of all real numbers strictly within distance 0.01 of 12.2”) or aggregatively (e.g., “the mereological sum of all real numbers greater than 3”) and not individually.

But this does not of itself prevent us from making intelligible reference to the individuals in the continuum. Often when we talk of these individuals, the referring expressions will be bound variables governed by some quantifier. Thus, we can say things like: “Let x be a real number that cannot be uniquely specified in some language E which is a formalization of English. Then, x is an irrational number. Moreover, x^2 is also a number that cannot be uniquely specified in E .” To say this *we* do not need to be able to pick out and individuate x , because the reference to x is within an implicit universal quantifier: “For all x , if x is a real number that cannot be uniquely specified in E , x is an irrational”

However, one might say that the observation that we cannot individuate all numbers shows that they are *not* real, and then go on to give a non-realist rendering of mathematics. Or, temptingly, one might adopt Aristotle's approach to the continuum and say that only finitely many (or perhaps, only countably many) points in a given continuum are *real*. The others “exist” only potentially. They could be made to exist should there, e.g., be some object at them that would point them out. This approach does have apparent difficulties. After all, the claim that the continuum is potentially infinite seems to imply that there are infinitely many “potential points”, and again the

unspecifiability problem comes back for these potentialities. One might insist that in the case of merely “potential points” the criterion that we must be able to individuate entities is inapplicable. But this Rescher cannot afford to do—for by the same token in the case of merely “potential individuals” the criterion will be inapplicable. So Rescher’s strong individuation objection to possible worlds does not appear to be compatible with realist mathematics, even of an Aristotelian variety.

Rescher’s own reply in conversation to this criticism is that in the case of the continuum we can at least specify *some* points, whereas in the case of merely possible worlds, Rescher claims we cannot specify any. In the next section, however, we will see that it might be quite possible to completely specify some—though probably not that many—worlds distinct from the actual one.

Moreover, it surely is *possible* that there is an omniscient being capable of individuating all possible individuals. The existence of such a being would entail that there are possible individuals. Thus, since it is possible that there is such a being, it is possible that there are possible individuals. But it would be strange indeed to allow there to be possible possible individuals but not possible individuals.

Section 3 Counterexamples to Rescher’s argument against possible worlds

But even given Rescher’s strictures, there are possible worlds that can be specified. For even if the individuals of this worlds are descriptively inexhaustible, that need not be true of the individuals in other worlds. Consider for instance, a physical universe that exists from time 0 sec. to time 100 sec. (both inclusive) and consists of two perfectly round balls, exactly three centimeters in diameter, whose only properties are (1) perfect hardness and elasticity, (2) a mass of 10 kg each, (3) a relative velocity of 10 m/s at time 0 directly towards each other, (4) the satisfaction of Newton’s laws of motion including the law of universal gravitation with G exactly equal to $6 \cdot 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$, and (5) any properties that are the logical consequences of the preceding. I have given a complete specification of a physical universe that seems possible. [\[121\]](#) One might, of course, deny that it is possible to have a material that has *only* the properties listed. But I can do a similar construction with some elementary particles which may be thought to have no properties beyond those that elementary particle physics enumerates.

Or consider a physical universe which consists of two causally isolated copies of the actual universe, except that one of these copies has matter where our universe has antimatter and antimatter where ours has matter (this condition is to ensure that there is no violation of the principle of identity of indiscernibles, should that principle be in fact true).

The universes in the above examples are not very interesting, perhaps. They won’t help us with counterfactual questions about our world, for instance. But they do suffice to show that even with Rescher’s unnecessarily restrictive individuation condition, there are possible worlds. And the strangeness of a view according to which the only possible worlds are only the actual world *and* worlds corresponding to the curious universes of the above paragraphs makes Rescher’s condition implausible.

Part VI. Aristotelian-Leibnizian theistic ontology

Section 1 Leibniz's approach

1.1 Leibniz's argument for the existence of God and the explanation of the nature of abstracta

Finally, having discussed in a negative way a number of accounts of possibility and possible worlds, it is time to sketch a positive approach. This approach will combine ideas from Leibniz and Aristotle. However, it will be a mere sketch, leaving many details to be worked out in the future. How plausible this account will be will depend in large part on how well these details can be worked out.

First we need to discuss the Leibnizian half of the account. After having set out his account of how the *possibilia* vie with one another for existence, as it were trying to persuade God to create them, Leibniz writes:

But, you say, this comparison between a certain determining metaphysical mechanism and the physical mechanism of heavy bodies, though it seems elegant, is faulty insofar as the heavy bodies striving really exist, while possibilities or essences before, or rather outside of existence, are imaginary or fictional, and therefore, one cannot seek a reason for existence in them. I respond that neither those essences nor the so-called eternal truths pertaining to them are fictitious. Rather, they exist in a certain realm of ideas, so to speak, namely, in God himself, the source of every essence and of the existence of the rest. The very existence of the actual series of things shows that we seem not to have spoken without grounds. For the reason for things must be sought in metaphysical necessities or in eternal truths, since (as I showed above) it cannot be found in the actual series of things. But existing things cannot derive from anything but existing things, as I already noted above. So it is necessary that eternal truths have their existence in a certain absolute or metaphysically necessary subject, that is, in God, through whom those things which would otherwise be imaginary are realized, to use a barbaric but graphic expression. [\[122\]](#)

This provides Leibniz with yet another argument for the existence of God, over and beyond his ontological argument and the cosmological arguments based on the Principle of Sufficient Reason. Necessary truths must have some kind of a reality, and since, Leibniz thinks, it is inconceivable that ideas should have a self-standing existence outside of a mind, it follows that there must be a mind that contains the necessary truth. Since ideas of the form “A world containing seventeen donkeys, the first of them lame in the right foot, 6.54 feet tall, ... is possible” are also necessary truths, ideas of *possibilia* are included in this argument. Observe that it is essential to this view that the “ideas” be taken to be divine *thinkings*, rather than merely *thinkables* which are the *objects* of divine thought [\[123\]](#), since arguably once one countenances thinkables as such, one is already a Platonist and Leibniz's desire for grounding these thinkables in a mind will no longer impress one. [\[124\]](#)

The crucial assumption in this argument is that there cannot be self-standing necessarily true propositions. One way to attempt to argue for this would be along the lines of Lewis's doubts about how propositions could represent. One might hold the very controversial view that representation is something that is necessarily dependent on there being minds and that only minds that have original intentionality. If X is a non-mental item, one can argue, using Putnam-like shuffling arguments (see Putnam, 1978 and 1983) that it is completely up to the minds that exist to decide what they want X to represent if X is to represent anything at all—even if X is some sort of a Platonic entity.

Even if there is some relation, like Platonic “participation”, between X 's and the states of affairs that make X true, something like the noetic rays that Putnam (1978) disparaged that are supposed to emanate from a sentence to that which makes the sentence true, it would still be possible to take X to represent something completely different, simply by choosing not to see *this* relation as “the relation of representation” but choosing to see another relation as being representation. We can reinterpret the noetic rays themselves: perhaps each noetic ray really points to the complement of that to which we first took it to point. Arguably, the only thing we cannot reinterpret in

this way is something that has intentionality by virtue of being in someone's mind. For plausibly everything else *can* be reinterpreted, and realism requires that there be something that *cannot* be reinterpreted. If this obviously controversial theory is right, then at least in the case of clear thought, we must then assume that *our* thoughts are intrinsically tied to meanings through some faculty of intentionality that we have (and if the controversial theory is wrong, then Leibniz's argument fails). But at the same time we need to extend the same courtesy to thoughts in the minds of other persons: thinkings in other people's minds have an intrinsic intentionality that we can understand and cannot reinterpret because these thinkings stand in an objectively analogous relation to their minds and the world to the one that our thinkings stand to our minds and the world.

It will be objected that of contemporary semanticists only Searle believes that minds alone can have original intentionality, and one can give accounts of intentionality, especially functionalist ones in terms of causal interconnections within us (either understood individually or collectively) as well as between us and the rest of the universe, that do not require any such original and intrinsic intentionality in the mind. However, two replies can be given to this. The first is that in the Leibnizian context we are not talking of intentionality in general, but specifically the intentionality of necessary propositions, understood realistically. Unless one is to subscribe to a questionable social theory of mathematics, it is unclear how causal interactions between a person and the world give intentionality to mathematical propositions or, for that matter, how such an account of intentionality can distinguish between thoughts of different logically equivalent propositions. After all, it is plausible that one can construct a Quine-type indeterminacy of translation argument that would show that we can systematically retranslate the meanings of the thoughts attributed to the mental components of the functional we-and-environment system in such a way that each thought gets translated into a logically equivalent thought (e.g., as in Quine's original approach, the thought "There is a rabbit there" could be translated into the logically equivalent thought "There is an attached rabbit part there"), while keeping functional (i.e., causal) interrelations fixed. But as has been seen in Section 4.1 of Part II, we should distinguish logically equivalent propositions.

Secondly, most sufficiently complex physical deterministic systems can be reinterpreted preserving causal relations (at least understood as subjunctive conditional relations) in such a way as to be seen to be isomorphic to any given functional system, as I shall argue in the Appendix. Thus, if we just have a photon shot out from a fixed photon gun, and if the functionalist is right, then the photon-gun system will have in it a mind having intentionality as complex as ours. But this is absurd, and hence functionalism is to be rejected.

Observe, however, that although I *have* argued against functionalism in the above, this is not crucial to the defense of the above argument for why specifically positing Platonic propositions is not enough. For even if there can be original intentionality obtained through a means other than that of an intrinsically intentional faculty of a mind, nonetheless re-shuffling argument work vis-à-vis the relations of *propositions* and the world, and that is all that a Leibnizian would need in an argument against the Platonic theory of propositions as existent abstracta. For even if our intentionality is derived from a functional system that we either are or are a part of, nonetheless the intentionality of propositions cannot be derived in this way, since abstracta such as propositions do not stand in any

causal relations. ^[125] However, an attack on functionalist systems *is* needed here for other dialectical purposes. For the divine mind that Leibniz posits is one that lacks functionalistic complexity and does not enter into back-and-forth causal interactions, and hence if it is to have intentionality at all, it must have it in some other way than functional systems are said to have it.

I am not attempting to argue here that *no* possible functionalist system can work. The shuffling and reinterpretation arguments would not work if in addition to merely preserving causal relations one were to require other kinds of preservation. For instance, a system that involves Aristotelian teleology or, equivalently, Aristotelian normalcy considerations, with its parts having innate purposes not reducible to categorial facts about them, is not subject to the same kind of objection. For instance, if some neuron not only is often caused to fire after a red patch is presented to a person, but in fact intrinsically has the property that it *ought to* fire after a red patch was visually

presented to the person, then one might want to say that the neuron has intentionality. Nor could one in any obvious way flexibly reinterpret a system involving these kinds of teleological causal relations.

Note, however, that such a system will not appeal to the scientifically minded philosophers who like functionalism, unless the teleology is understood in an evolutionary way. But an evolutionary account of intentionality is implausible on two counts. The first is that evolutionary stories can be told, arguably with no loss of informative value (though this is a controversial point), without any teleological language, simply in terms of non-teleological causal relations (those genotypes that lead to phenotypes that aid transmission of genotype survive, whereas such-and-such mutation processes produced such-and-such genotypes in such-and-such environments, which led to the survival of these but not other organisms and hence transmission of their genotypes, etc.) But if this can be done, then the *whole* causal system involving evolutionary history can still be reinterpreted in various ways, just as an isolated system can. The second implausibility to this form of the objection is that would be physically possible via random quantum phenomena for human DNA to have arisen by a *single* very large scale mutation in some primitive organism. It would be highly implausible to claim that were this to have turned out to be the true history of human beings, it would also have turned out that human beings would not in fact have thoughts. [\[126\]](#)

In any case, since the only reason to consider functionalistic accounts of intentionality is to consider the anti-Leibnizian argument that a divine mind *could not* have intentionality, and since the teleological functionalism has not been worked out in detail and even if it were, one would not have much reason to think that every kind of original intentionality *must* come from a complex functional system of this sort, I will not consider such a theory further.

Brian Leftow (1989) has noted that even if we accept that necessary truths must be found in a mind, it does not follow that the mind must be a necessary being. Perhaps there is a different omniscient mind in every world, each such mind being contingent but with it being logically necessary that some such mind exist. One way to argue against this possibility would be to claim that surely it is logically possible that no contingent person exists, and hence if Leftow's suggestion were true, then there would be worlds without necessary truths, namely those worlds that do not contain omniscient minds, which is absurd. But even so, one can claim that different necessarily existent minds are omniscient in different worlds, none being such that it *necessarily* contemplates all necessary truths.

Against this one can bring the argument that surely what ontologically grounds necessary truths had better be the same in every world. The ontological grounds for it being true that $2+2=4$ will, surely, be the same in every world. If necessary truths are grounded in divine ideas, then the same mind must contemplate them in every world. But perhaps, as, at least according to Adams (1994, Chapter II), Leibniz did, the Leibnizian will not say that the *truth* of necessary truths is grounded in God's having them as ideas.

One might also argue that there could be multiple divine minds, each of which has the ideas of some, but not all, necessary truths, and together they have the ideas of them all. A less than completely satisfactory solution to this objection would be just to define as "God" the aggregate of all these minds, leaving it as a topic for further investigation how united God's substance is. A more satisfactory solution is to follow Adams' (1994, Chapter II.7) exposition of Leibniz, according to which the unity of the divine mind follows from the fact that the propositions are supposed to be interrelated, and it is mysterious for Leibniz how they could be related if they are in different minds. Indeed, after all, the idea of a possible world will contain the ideas of all other possible worlds since at that world it will be true that they are possible. Likewise, one might also argue that if p and q are necessary truths, so is p -and- q , and hence p -and- q will have to exist in some divine mind. On the assumption that one cannot have the idea that p -and- q without having the ideas that p and that q , it easily follows that if there are finitely many divine minds, then one of them contains all propositions, and we can reasonably call that one *God*. If there are infinitely many divine minds, and *if* there is a conjunction of all necessary truths (a rather complicated beast, since it will be a conjunct of itself), then we can use the same argument as above.

1.2 A solution to the problem of how the possible worlds represent

If the notion of divine ideas or thinkings can be made to work coherently, Leibniz can define a possible world as a coherent maximal divine idea. Recall now Lewis's objection to the problem of how it is that the possible worlds built up out of propositions represent states of affairs. This problem is now partially solved: they represent in a way analogous to the way *our* ideas represent states of affairs. Of course how *that* happens is itself a difficult question, but one cannot coherently doubt that it *does* happen.

1.3 The Parmenidean challenge revisited and an assessment of the Leibnizian approach

However, Leibniz's account does not answer the Parmenidean challenge any more than the propositional account did. It is a necessary truth that it is possible for there to be unicorns. There is an idea in the mind of God which expresses its being possible for there to be unicorns, since there are ideas in the mind of God that express the necessary truths. This idea is true or correct. But what makes it correct? Ideas are correct if they accurately reflect reality. Saying that a necessary truth is an idea in the mind of God does not let us off the hook of explaining what the truthmakers of these ideas are. If we say that the truthmaker of the idea consists in the idea that there are unicorns having the property of possibility, we are seemingly no further ahead that the Ambitious Propositional Ersatzist was.

Though, actually, we do have a gain. We do not have to deal with abstracta. We do not have to worry about how we can meaningfully refer to entities that are of a kind that plays no causal role in anything. For, the divine ideas as aspects of the divine mind presumably can be thought of as playing a causal role in creation. Moreover, if we have a substance-based ontology according to which all things that exist are substances or things that ontologically depend on substances, then this approach is more tenable: the divine ideas ontologically depend, in some way, on the substance that God is.

Furthermore, Lewis's worry of how it is that there can be entities like propositions which somehow manage to represent states of affairs is alleviated. For, although it is a difficult problem to say *how* the thinkings of a person represent the world, it would be self-defeating to doubt that they *can* do it.

One objection that could be made, however, is that that in virtue of which the thinkings of a human person represent the world would not be available in the case of the thinkings of God. For instance, one might think that intentionality is produced by, or at least requires, two-way causal interaction with the world, [\[127\]](#) whereas God (at least as traditionally thought of) is not subject to mutual interaction. If this objection were correct, it would constitute an argument for the non-existence of God, of course.

In response, first of all one could try to argue that God is at least in one-way causal interaction with the contingent parts of the cosmos, by being their cause, and moreover has the potential for non-actual causal interactions, and the connection of the divine thinkings with this potential may be sufficient to give them intentionality. But this would be only a promissory note, with it being unclear whether it can be filled out.

Alternately, one can argue that there is no reason to believe that the correct account of intentionality will have to involve two-way causal interaction. For instance, the failure of the existing accounts that construct intentionality out of mere efficient causal interactions to handle problems of reinterpretation gives one little confidence for thinking that the correct account of intentionality will involve *solely* such interaction. And as to the intuition that causal interactions account are necessarily at least a *part* of any correct account of intentionality, one might counter it with the intuition that intentionality at least sometimes can be had outside of causal interaction, namely in mathematical thought. It is true that one might hold out some hope for an account of intentionality that uses primitive Aristotelian teleological causation which is not subject to reinterpretation, but once one admits irreducible *telê*, it is not a far stretch to admit an irreducible faculty of intentionality. After all, is there a significant

difference between saying that a thinking x is *of* a state of affairs s and saying that assertorically engaging in x is *normal* only when s obtains (I assume here that one can go back and forth between talk of teleology and talk of normalcy)?

Finally, one might try to counter the objection by using a panoply of the standard arguments for theism, since once it is established that there *is* a God, it follows that indeed it is possible to have intentionality without two-way causal interaction.

If the objection can be countered in one of the above ways, or in some other way, then the Leibnizian can also define “propositions” in general as divine ideas of states of affairs, and get the theoretical benefits of propositions at no greater ontological cost than the account of necessary truths. Since Lewis’s account of propositions was argued to be implausible (Section 4 of Part II), this means that the present account has an application which his does not.

Section 2 Aristotelian possibility and causality

2.1 Proximate matter, proximate cause and Parmenides

Now consider a very different theory by recalling Aristotle’s theory of change in the *Physics*. Change involves matter and two forms, a positive form and a privative form. The matter in question is directly capable of having both forms. This matter is termed in the *Metaphysics* “proximate matter” (see, e.g., L.3, 1070a19). It is proximate because with no further prior changes, it can take on either of the two forms. Thus the proximate matter for a brazen statue is not the elemental earth. Nor is it the copper and tin that the elemental earth is worked up into. Rather, the proximate matter is the bronze that is produced from the copper and the tin. The copper and tin are not the proximate matter of the statue because copper and tin require a further process in addition to the process of taking on the form of the statue in order to gain that form, namely the process of alloying.

Another way to put this is to say that Aristotle sees the possibility of there being a statue as *materially* grounded in the capabilities of the proximate matter, viz. the bronze. The proximate matter is the matter found at the last stage in a process that can branch into either of the two forms that the matter is capable of. But along with the proximate matter responsible for the material grounding, there is doubtless what one might call the “proximate cause”, which is the cause capable of operating on the proximate matter and of producing in it the forms that it is capable of. In the case of the statue, this proximate cause will be the sculptor’s decision whether or not to make a statue. Aristotle does not distinguish the proximate cause in the explicit way he does the proximate matter, but it is implicit in his account.

Armed with this terminology, let us imagine that we have before us a lump of bronze and a sculptor. It is possible that a statue will be there in the future, but it is also possible that a shapeless mass will be there in the future. Suppose that in fact no statue will exist here. Nonetheless, it is possible that a statue will exist here. And this is an assertion of possibility in connection with which we can give an answer to the Parmenidean problem. A Parmenidean will wonder what it is that we are talking about when we say a statue is possible, since the statue talked about does not exist, has not existed and will not exist. Aristotle can answer that the assertion that it is possible that a statue will exist has as its truthmaker the *actual* capabilities, namely second potentialities, of the sculptor and the bronze: the sculptor’s actual know-how, the actual strength and dexterity of her limbs, the malleability of the bronze, and the like. All these things are *actually* possessed by the sculptor and the bronze. We are only talking of the actual world here. And being *second* potentialities, these capabilities have in themselves all that is needed for producing the effect.

But is this not a circular answer? After all, are not the “actual” capabilities of the sculptor and of the bronze *dispositional* properties and as such ones that cannot be spoken about if one is restricted to speaking of what is merely actual? However, it is here that the importance of speaking of the *proximate* matter and cause shows up. The proximate matter is the matter which *in its own right* (*per se, kath’ hauto*) is capable of taking on the various

forms. It needs nothing else, except the proximate cause, for this. Thus the items which actually exist in this world, namely the proximate matter and the proximate cause are sufficient to guarantee the possibility of the statue's existence. They must be sufficient for producing the statue because they are sufficient for a complete explanation of the statue's existence. And because they are sufficient to guarantee this possibility, it is plausible to say that when we are talking of this possibility we are talking of them. Actuality is prior to possibility in the sense that assertions of possibility are in fact grounded in, have as their truthmaker actual states of affairs.

Of course this is not a completely reductive account of the possibility of a statue's existing. We are analyzing a possibility in terms of things that are themselves through-and-through modal: capabilities, causal powers, dispositions, etc. However, these things are actual and concrete in a paradigmatic way: they enter into causal explanations of actual phenomena.

2.2 The basic Aristotelian view

One way in which the present account is a mere sketch is that it has yet to be determined precisely which concrete modal things such as capabilities, causal powers, dispositions, etc. are needed for the account. This will indeed have to be investigated in the future.

However, as described above, this is not the only problem, even if the account might be helpful in the case of the statue. For it would still be true to say that it is possible that a statue will exist even if the proximate matter of the statue had not come to be, but, say, had remained as tin and copper, and even if there had never been a sculptor on earth. Moreover, as noted in the Introduction (Section 4.5 of Part I), the Aristotelian view runs into an apparently decisive objection in its inability to handle global possibilities.

First suppose that in fact in the course of human history, sculptors had never arisen, while tin and copper had never been alloyed into bronze. We can still give an Aristotelian account of why a statue could have existed. The possibility of the statue is grounded in the possibility of bronze and a sculptor which would be the proximate causes, while the possibility of bronze is grounded in the capabilities of tin and copper which could be the proximate material causes of bronze, and the possibility of a sculptor is grounded in more general capabilities of human beings which could be proximate causes of the existence of sculptors. The possibility of the statue is just seen to be ontologically grounded in a longer chain of possible causes, which chain nonetheless starts with some *actual* items in the world. Moreover, the items at each point in the chain were sufficient for producing the items in the next stage in the chain. General capabilities and the free will that human beings have are sufficient for making some human being into a sculptor.

What this account suggests is the generalized Aristotelian view that the claim that it is possible that *B* take place is made true by there having existed something, *A*, such that when it exists it has causal capabilities enabling it to be the first step in a chain of causes (perhaps the chain having indeterministic links and perhaps it being a null-length chain, i.e., one that consists just of the initial thing itself) terminating at *B* and such that each step in the chain would have sufficient capabilities to be a full explanation of the next step should the next one be taken. This truthmaker is something actual, not in the temporal sense of being *now*—for it might be in the past—but in the sense of being something in the actual world. Observe that trivially anything actual is possible since it leads to itself with a trivial null-length chain. Observe also that because each step in the chain is capable of being a full explanation of the next (albeit perhaps an indeterministic explanation), all the possibilities in the chain are inchoately present in the first actually existent member of such a chain.

This chain-based definition takes care of the difficulty with possibilities whose proximate causes are not in actual existence and gives an account roughly like that of branching modalists like Penelope Mackie (1998).

Note, too, how this approach entails S5. First, observe that S4 follows quickly. Suppose it is possible that it possible that *C* occurred, but not possible that *C* occurred. Then, there is a state of affairs or substance *A* which had the power of initiating a causal chain leading to it being possible that *C*. Were it to have initiated that causal chain, then by assumption *C* would not have occurred (since it is not possible that *C* occurred), but the chain

would have led to there being an item, B , capable of initiating a causal chain leading to C . But then surely A is capable of initiating the chain leading to B , and B is capable of initiating the chain to C , and putting these two things together, A is capable of initiating a chain leading to C (though A may not be capable of *ensuring* that B initiates a chain leading to C). Hence, if it is not possible that C occurred, it is possible that C occurred, and so it is possible that C occurred.

The Brouwer axiom also easily follows from the Aristotelian view, on the assumption that if A initiates one chain C_1 of events and is capable of initiating another chain, C_2 , then were it to have exercised this capability for initiating C_2 , it would still have been capable of initiating the chain C_1 . It is important that what the powers of an item (state of affairs or substance) at t are, the actualization of which powers grounds various possibilities, should not itself depend on which of these powers are actualized. For then the powers would not be prior to the actualization.

2.3 Global possibilities: a difficulty

However the difficulty about global possibilities remains on this view. Surely, the universe could have been radically different at every instant of time in the past. Perhaps our cosmos might not have been temporal at all. Certainly our universe might always have been twenty-two dimensional. Our universe started existing after a Big Bang. But, *pace* William Lane Craig (1994) ^[128], it might also have always existed. But these are possibilities which it is difficult to see how they are to be grounded in an actuality, since in the Aristotelian account the possible world and the actual world in question always have something in common, namely the causes in which the possibility is grounded, e.g., the sculptor capable of producing or not producing a statue.

It is here that combining the Aristotelian account with Leibniz's theistic story would solve the difficulties. On this combined account, there would be a God outside the universe, though not outside the cosmos (recall that in the present terminology "the cosmos" is the aggregate of all existing entities while "the universe" is the aggregate of all spatio-temporal ones), who has in his mind ideas of all kinds of possible worlds. We get to retain from Leibniz's account the intuition, itself an Aristotelian one, that propositional-type entities can only exist in minds. But what we now add is the fact that this God is not just all-knowing but also all-powerful. He is capable of putting his divine ideas into action. He can produce a twenty-two dimensional universe. If he himself is atemporal, then he is capable of creating beings that are atemporal. Whether he is temporal or not, he can make a universe that has always existed—by *creatio continua* in the former case and a timeless act of creating time and a universe that had always existed in the latter case.

As far as this argument goes, though, the theistic story is not the only one available to the Aristotelian. The Aristotelian might, with the historical Aristotle who thought that the existence of the actually instantiated natural kinds to be necessary, simply deny the possibility of the universe having been radically different. On this view, though for any time t it might have been that the universe was different at t from the way it actually was at t —for this is a possibility we can ground in the powers of things at a time t^* prior to t , nonetheless it is false that it might have been for all t different at t . I take this to be an implausible view, since it seems plausible that if the state of the universe on any time slice is contingent, it might surely have been that the state of the universe was different on *all* time slices. But it is indeed a view an Aristotelian might take. Alternately, the Aristotelian might accept something atemporal prior to the course of history which non-deterministically causes the start of our history, but deny that this atemporal being is a person—perhaps this being is just some event in a supercosmos, it might be said.

The Aristotelian thus has three options: (1) deny that the whole course of history was different, (2) accept that it could have been different due to the causal agency of an ahistorical non-person, or (3) accept that it could have been different due to the causal agency of an ahistorical person, by all account a God. We will see an argument similar to the one behind this trichotomy in more detail in Section 2.5.

2.4 Primitive causal modality and counterfactuals

One might argue that the Aristotelian view, whether in its theistic version or in a version that bites the bullet and says that the whole history of the universe could not have been different, has a serious disadvantage over a view like Lewis's, because it presupposes primitive causal powers, whereas Lewis manages to reduce both powers and causality to his possible worlds.

However, in fact, a Lewisian account must also presuppose causal relations, and indeed not only actual causality but potential causality. For, it is an integral part of an account of modality that it give an account of counterfactuals. We have seen, however, in Section 3 of Part II that Lewis's account of counterfactuals must make a distinction between past and future: agreement of a possible world with the actual world in the past counts for more than agreement in the future when evaluating closeness of worlds. This, however, presupposes causality, because I will argue that causality is the best way of distinguishing past and future. Briefly, event A will count as being in the past of event B if and only if something happening in the same place and time as A is at could have affected what happens at the same place and time as B is at.

The reason why this is the best account comes if we try to extend our natural notion of a past to a relativistic context. The relativistic analog of the past of an event is the backwards half light-cone with apex at the event.

Now imagine someone who disagreed and said: "No: the past of an event is not the backwards half light-cone centered on the event, but something else, such as the intersection of the backwards half light-cone with the world-line of the galaxy in which the event happens, if the event happens in a galaxy, and the backwards half light-cone otherwise." The best answer to this is that then an event in our galaxy would be affected by many spatio-temporal events that are not in the past of the event—namely, by the multitude of events in the past half light-cone of the events that do not happen in our galaxy—which is absurd. This answer shows, I take it, what it is that we think is significant about the backwards half light-cone of a space-time point x . This backwards half light-cone is simply the region of space-time events at which, according to current physics, can causally affect what happens at x . If we had different laws of physics, we might end up choosing a region of space-time different from a past half light-cone to be "the past of x ", but we would choose this region by the same definition: the set of all points in space-time at which an event can causally affect what happens at x . It is because backwards half light-cones satisfy this definition that they are significant and can be counted as the "the past" of the space-time point at which an event happens.

Thus, the relativistic definition of the past, even though it seems purely geometric, presupposes causal relations. The reason it presupposes it is that the purely geometric definition would be merely stipulative in an objectionable way unless the causal story is given. Observe also that the causal story involves not merely actual causal relations but *possible* ones: we are looking at points in space-time at which an event *can* affect the given event.

It is also worth noting that we can probably give an account of counterfactuals in terms of possible worlds without supposing a space-time framework, working in terms of the raw ingredients of the Aristotelian view, namely causal powers, providing we also have an account of possible worlds on the scene (e.g., the Leibnizian one). One way to do this is to work in terms of substance-slices, namely time-slices of substances, and think of states and causal powers as things had by substance-slices. If s is a substance-slice and w is a possible world, then define $\text{past}(s,w)$ as the set of all substance-slices in w which have the capability for initiating a causal chain affecting s . Then, we say that two worlds w_1 and w_2 *match in the past of s* if s exists in both worlds and every member of $\text{past}(s,w_1)$ is a member of $\text{past}(s,w_2)$ and has the same state, and vice-versa. Then, let $\text{match}(w_1,w_2)$ be the set of all substance-slices in the past of which w_1 and w_2 match. We can then use the old Lewisian criteria for similarity of worlds, but add a criterion that trumps them all: World w_2 is closer to w_1 than w_3 is providing $\text{match}(w_1,w_3)$ is a proper subset of $\text{match}(w_1,w_2)$. I leave it as an exercise for the reader to examine the consequences of this

[129]

approach, but conjecture that this approach will indeed give correct answers to counterfactual questions.

Unfortunately, this account may not work for counterfactuals with disjunctive antecedents or existentially quantified antecedents, as is shown by a counterexample to Lewis cited by Edgington (1995). Suppose E is an event at a time prior to the time of an event F , and that neither event is actual. Then, on this account it is automatically true that were E or F to have occurred, then F would have occurred. After all, a world where F occurs but where the past of F matches the actual world will be closer by the proposed criterion to our world than one which only matches exactly in the past of E . But of course there are cases where we do not want to say that were E or F to have occurred, then F would have occurred. To slightly sharpen the example Edgington cites, suppose two people enter a room, one before another, before noon. My coat was in the room all along, and at noon it is still there. Then it is automatically true that were one of these two people to have stolen the coat, it would have been the later one who did it. But if the later one is someone known for her moral rectitude while the former is a notorious coat thief, then this counterfactual would presumably be false. Thus, the proposed account has to be limited to counterfactuals with antecedents that do not involve disjunction or existential quantification. It is not clear what exactly is to be said about the disjunctive antecedent case. A case can be made that in fact neither of the counterfactuals (1) were one of the two people to have stolen the coat, it would have been the later one and (2) were one of the two people to have stolen the coat, it would have been the earlier one is true. Instead, what is true is a probabilistic counterfactual that were one of the two people to have stolen the coat, it would probably have been the notorious coat thief. However, probabilistic counterfactuals, though most interesting, are beyond the scope of this discussion. Note that these problems are problems for Lewis, too.

2.5 The necessary existence of a first cause: a proof

Observe now that the Aristotelian account makes God a necessary being, given that by definition God has no cause. For in fact the Aristotelian account implies that, necessarily, every contingent being has a cause. To see this, suppose for a *reductio* that it is possible that x is a contingent being that has no cause. Let p be the proposition that there is an uncaused contingent being. Then, p is possible. There is then a possible world (e.g., an unambitious and hence uncontroversial propositional ersatz possible world—I am just using possible worlds as an aid to expression) where p is false. But by S5 (which the Aristotelian account entails as has been seen), since what is possible is necessarily possible, p is still possible in that world. But by the Aristotelian account, this possibility would have to be grounded in a cause or causes capable of making p true. But it is logically impossible that the proposition that there is an *uncaused* contingent being should be made the case by a *cause*. Therefore, there cannot be a cause or causes capable of making p true, and hence p is impossible on the Aristotelian account, and yet possible by assumption, which is absurd. Consequently, there cannot be an uncaused contingent being.

Note that this argument for why an uncaused contingent being is impossible in fact presupposed nothing more than the basic Aristotelian analysis of possibility in terms of causes capable of producing the possible effects. This argument can in fact be made into an argument for the existence of a first cause along the lines of Thomas Aquinas' Third Way.

We need first two assumptions:

- (57) The Aristotelian account of possibility is true: Something non-actual is possible if there actually exists something that could have initiated a causal chain leading to it.
- (58) If S is a set of actually existent contingent beings, then it is a contingent fact that *any* of the beings in S exist.

Now, let S be the set of all actually existing contingent beings, and let w_0 be the actual world. Then, let w_1 be a possible world in which no member of S exists. It is then true at w_1 (by S5) that it might have been that w_0 was actual. Then, by the Aristotelian account, there is a being x in w_1 which could have initiated a causal chain that could lead to w_0 being actual. But x cannot start any chain of causes that can lead to actuality's not including x . Therefore, x must also exist in w_1 . Thus, x exists both in w_0 and w_1 . Since no member of S exists in w_1 while x exists in w_0 , it follows that it is true at w_0 that x is not a member of S and hence is a necessary being. Moreover, x

must be the first cause of all contingent beings in w_0 . For were x to initiate a chain of causes that leads to w_0 's actuality, it would necessarily be the first cause of all contingent beings in w_0 . Therefore, since x can initiate a chain of causes that leads to w_0 's actuality, it cannot be the case that w_0 is a world in which x is not the first cause of all contingent beings in w_0 : for x evidently could not lead to the actuality of such a world. Therefore, x is indeed the first cause of all contingent beings in w_0 .

And if we were Thomas Aquinas we would say that after we have argued for the necessary existence of x “et hoc dicimus Deum.” But we must be careful here. No argument has been given for x being a personal first cause. It has not even been argued that x is a singular entity rather than an aggregate of substances. [\[130\]](#) But what the argument does show is that the Aristotelian must either reject (61) and insist that though any one contingent being could have failed to exist nonetheless it is necessarily true that at least one of the contingent beings that exist exists, or she must accept there is a first cause. The first alternative is not an attractive one. Note that the claim the Aristotelian who rejects (61) makes is not just the claim that it might have been that no contingent beings existed—that claim indeed would be a controversial one [\[131\]](#)—but the less controversial claim it might have been that none of *those* contingent beings that exist existed.

The Aristotelian who accepts (61) is committed to a necessarily existing first cause. Does this Aristotelian need to identify this first cause with the God of theism, or can she allow it to be some event in a super-universe atemporally prior to the Big Bang or something like that? This will depend on other issues. For instance, Koons (1997) has argued that once one has an argument for the existence of a first cause, teleological arguments gain in plausibility, and so one might have reason to accept that the first cause is intelligent. The plausibilistic considerations in Gale and Pruss (1999) as well as the arguments of Gellman (2000) are also relevant here.

Alternately, and closer to the issues of this dissertation, if one wants the benefits of a theory of possible worlds and finds the Leibnizian approach plausible, then in order to have *both* possible worlds *and* a grounding for alethic modality on the scene one might point to the theoretic utility of a theory on which there is one God whose powers ground alethic modality and some of whose thinkings are the possible worlds that he can initiate causal chains leading to the actualization of. This would closely merge the Aristotelian and Leibnizian accounts. But of course it would require that the controversial parts of the Leibnizian account—the intentionality of the thinkings of a non-natural being—can be made sense of.

2.6 The capability of producing an effect makes it true to say that the effect is possible

We can now solve the difficulty which was the main objection to Leibniz's account of possibility. Granted, propositions expressing modal truths are ideas in the mind of God. But what, we asked, makes these propositions true? The answer is that they are made true by the capabilities of the God whose action-guiding ideas they are [\[132\]](#) and by those of beings that he might have created. What makes it true to say that it is possible for there to exist a world with unicorns is that the idea of such a world is an idea in the mind of an all-powerful God capable of acting on it and creating such a world.

This account also improves on the answer to Leftow's (1989) concern that perhaps different minds know the necessary truths in different possible worlds. The present account entails that the same God must be the first cause of all things in all possible worlds. For suppose that there was a world w where the God who is the first cause of all things in our world was not the first cause of all things, but instead Smith was, and Smith was not created by God. Then, it is true at our world that it is possible for Smith to exist with Smith not identical to God and with Smith being a first cause. Hence, God's creative action can initiate a branching process one of whose branches is going to lead to the existence of Smith, not identical with God, and with Smith being a first cause. But this is absurd, for in the case of such a process, Smith would no longer be a first cause. Since modal truths are to be analyzed in terms of ideas in the mind of a first cause, and since the first cause must be the same in every world,

it follows that at least the same mind contemplates the modal truths. (Cf. the arguments in Gellman, 2000.)

Moreover, the Aristotelian account, even in its non-theistic form, gives an argument against the Platonic view of modality. On the Platonic account, what makes it possible that I was a biologist is that the abstract proposition that I was a biologist has the abstract property of possibility. But *there are* concrete capacities and dispositions in the universe that are sufficient to make it possible that I was a biologist, because they are sufficient for causally making me a biologist—the Platonist should not deny this. We thus have two different ways of characterizing possibility: one is via concrete this-worldly Aristotelian properties of concreta which really do exist—the Platonist should not deny this—and the other is via some abstract Platonic primitive properties of abstracta. Moreover, anything that is possible on the Aristotelian grounds will be physically possible, and hence also logically possible, and hence possible on Platonist grounds (though perhaps not conversely). But why is this so? Why is it there this coincidence that anything made possible by this-worldly powers and capacities and dispositions happens to correspond to a proposition in the Platonic realm that has a certain abstract property? The Platonist is unable to explain this coincidence between powers in our universe and abstract facts about the Platonic realm.

2.7 The Humean objection from the compatibility of all things

If things are as described, then some worlds that are apparently possible are impossible. For if God is necessarily omnibenevolent, omniscient and omnipotent, then it is a necessary truth that any evil that exists is *justified*, i.e., there is a morally extenuating reason why an omnibenevolent, omniscient and omnipotent deity would allow this evil. Richard Gale (1996, p. 229) has in fact used this observation to argue against the possibility of a necessarily existent omnibenevolent, omniscient and omnipotent deity, since, he claims, it is plainly possible for there to be an unjustified evil. Moreover, Gale claims that theists by taking seriously worries about the problem of evil are admitting this logical possibility.

One will be particularly confident about Gale's argument if one accepts a Humean view that anything could co-exist with anything else. We could, thus, have a universe in one corner of which there is a perfectly innocent person who co-exists with a horrid pain in her mind for all eternity, and both of which co-exist with the rest of the universe being set up in such a way that no one can benefit from this pain. But such a pain, surely, would be an unjustified evil—yet, the Humean claims it is possible. Hence, there cannot be a necessarily existent omnibenevolent, omniscient and omnipotent deity.

One can, however, challenge the Humean intuitions by bringing in the global nature of the Aristotelian-Leibnizian account of modality. To make a claim of the possibility of some non-actual event is to make a claim about how something actual could have brought it about. To be sure that what Gale claims is possible is indeed possible one would have to give some sort of a causal account of how it could come about. But no such causal account can be given without begging the question against the existence of a necessarily existent, omnibenevolent, omniscient, omnipotent creator of all things. For if there is such a creator, then any full causal account will have to ultimately go back to him, perhaps in a non-deterministic way, and if the evil is unjustified, then no such story going back to him is available. The confidence with which one affirms such possibilities withers when one realizes that in asserting the possibility of a proposition one is asserting something's capability for ultimately bringing this possibility to actuality.

One might try to give an account that does not go all the way back to God, but goes back to some free creature that has the capacity for inflicting an unjustified evil (i.e., one that *God* would not be justified in permitting the creature to inflict) and ground the possibility of an unjustified evil in that. However, we do not know that there in fact was any such creature ever in existence (it will not do to claim such a creature is *possible* because then the considerations in the previous paragraph come into play). God might have had contingency plans to bring justifying goods out of all evils that he had given creatures the capability for and did not give more freedom than he was morally justified in permitting the exercise of. Of course this gets one back to the “standard” problem of evil,

[133]

which, alas, is beyond the scope of this dissertation.

Of course there is a simpler answer if one is not interested in defending all of traditional theism. One could just unambitiously say that the Aristotelian-Leibnizian account of possibility does not require the assumption that God is necessarily omnibenevolent, and hence Gale's argument is irrelevant as an objection to this account of possible worlds. But the more ambitious answer above is, of course, preferable.

2.8 The free will of creatures

Things would be particularly simple if it were necessarily the case that God's creative act determined the truth values of all contingent propositions. Then, for a world to be actual would be precisely for that world to be the one that God actualizes, to be possible is to be one that God can actualize, and that would be it for the analysis. However, although the analysis of actuality and possibility thus stated is correct, it must be carefully understood. After all, it is logically possible for there to be free contingent beings. Now, on apparently reasonable principles, it is impossible that God determines a free action of a contingent being, since an action which is completely determined by another person is not free.

To take into account creaturely freedom, we have to distinguish between "strong" and "weak" actualization of a propositions, and in the claims of the preceding paragraph it is weak actualization that should be understood. God *strongly* actualizes a proposition if God's creative act determines the truth-value of this proposition to be *true*. God *weakly* actualizes a proposition if God's act of creation is the first item in the causal series leading up to the

proposition's truthmaker. It is left open whether God knows what *would* be the truth-value of this proposition [134] at the logical point of his creative decision, but of course he knows what the possible results are since he has the divine ideas of them. Consider now a proposition of the form: "It is possible for there to be an intelligent dragon that freely chooses to scorch an oak tree." The truthmaker of this claim is God's power to *weakly* actualize the proposition that there is a dragon that freely scorches an oak tree. This proposition is of course one that it is logically impossible for God to *strongly* actualize, since were he to do so, the dragon's action would not be free. We can say that God is the first cause of the truthmaker of the proposition that a dragon freely scorches an oak tree, but not that he is a determining cause of it.

It seems that the ground of possibility of the possibility that a dragon freely scorches an oak tree should, then, be sought in the powers of dragons, and not in the power of God. That is true as far as it goes, except that there in fact are no dragons, and hence no draconic powers. The possibility of a dragon freely scorching an oak tree is *ultimately* grounded in the fact that God can create a dragon that has the power of freely scorching a tree. If there in fact did exist dragons, then we could also find a truthmaker for the possibility in the power of some actually existing dragon. This illustrates that in fact the same claim can have more than one truthmaker. That is not absurd: Every horse is a truthmaker for the proposition that there exists at least one horse.

However, in the case of possibility, if dragons existed, although we could take draconic powers as the truthmakers of claims about what it is possible that dragons do, this possibility is still ultimately ontologically grounded in God's power to have brought the dragons into existence.

What we have done is to see possibility in terms of a branching system, with God's creative choice being the first branching in the system. We can say that all possibility is ultimately grounded in this first branching, because the *power to branch* at subsequent stages itself comes from God's bestowal of powers on the entities whose activities determine these subsequent choices.

2.9 Branching worlds all the way down

Indeed, we can say that the present system allows one to have a branching theory of modality, rather like those that one might have in a tensed logic system, except that *this* theory allows for global possibilities—for the

whole temporal sequence to have been different—because it has a God who of logical necessity is the first cause and can create different cosmic systems.

Note that the branching system here is not necessarily temporal: God might not be in time. However, that should not be a problem, unless one thinks that causality is always temporal (in which case one will, *pace* my considered view, simply have God in time, too, maybe locating his creative act at time $-t$ or making it a continual act).

A major advantage of the branching view is that one can handle different modalities in a closely parallel way. Ordinary language does not concern itself with logical possibility as much as with physical possibility (cf. Place, 1997), and logical possibility is a generalization from physical possibility. If we see physical possibility and necessity as grounded in the causal powers of actual things, then it is difficult to see how logical possibility could be a generalization of this, unless one takes the Aristotelian-Leibnizian view. For on that view, there is no radical difference between logical and physical possibility. The only difference is that when we talk of the capabilities that ground physical possibility, we restrict our discussion to take into account only the capabilities of physical objects, whereas when we talk of logical, or (as I have argued in Section 2 of Part I, equivalently) metaphysical, possibility this restriction is relaxed and the capabilities of *all* entities are now permitted to be talked about. Similarly, a temporalized notion of possibility are arrived at when we restrict ourselves to talking of the *future* exercise of the capabilities of actual entities. The fact that the account makes available such a common thread in all non-epistemic notions of possibility is a major point in its favor.

There is no problem with grasping the nature of the modalities involved in this branching account, because this nature is nothing else than the nature of non-deterministic causation, a causation that we ourselves ordinarily take ourselves to exhibit, and which we can introspectively understand. [\[135\]](#) Note that for this grasp it perhaps need not even be required that we *exhibit* such causation, but only that we *experience*, perhaps non-*veridically*, the exercise of such causation. (For an account of causation that starts with agency, see Swinburne, 1997).

Section 3 The challenges to Lewis's ontology no longer a problem

Observe that most of the challenges to Lewis's ontology do not apply to this account. Consider them one by one.

3.1 Primitive modality

The present view certainly depends on a primitive Aristotelian modality: substances' capabilities for bringing about effects. However, this primitive modality is not objectionable in the way that the primitive modality involved with Lewisian counterpart theory was (see Section 4.2.1.b of Part III). For, it can be argued *pace* Hume that we have a direct grasp of this sort of modality. Leibniz has claimed that because we are substances, we have a grasp of what it is to be a substance. Similarly, because we are, or at least experience ourselves to be, active substances, we have a grasp of what it is to bring about effects. But even more than this, we have a primitive grasp of what it is to be capable of doing something, even if we do not do this. This grasp is, for instance, in play when we see ourselves as responsible for having neglected a duty. Our knowledge of ourselves as guilty involves a knowledge of ourselves as having neglected duty.

Recall that what made the primitive modality in Lewis's counterpart theory particularly objectionable is that one relation among many had to be picked out as *the* counterpart relation (or a proper subcollection of the collection of all relations had to be picked out as *the* subcollection of acceptable counterpart relations), and there was no way we had of specifying *which* relation (or collection of relations) this was. We can, however, arguably specify what kind of a thing a causal power is by ostension, if we can overcome the problems with ostensive definitions in general. [\[136\]](#) Note that if this approach were to work, then the "grasp" of causal powers would not be

objectionably magical. What is objectionably magical is the specialness of a property such that we cannot give a story about what distinguishes *this* property from all others: a property that is supposedly special but yet one which we cannot individuate from all other properties (cf. Lewis's [1986a, p. 178] criticisms of the Platonists which accuse them of this kind of magic).

It might be argued that, on the contrary, by taking the causal chain of modalities back to God, I have brought in the *mysterium tremendum et fascinans*, and there is nothing more mysterious than that! However, the mystery involved in God is of a different kind than that in Lewis's view. The objectionable "mystery" in Lewis's view consists in the fact that an account of modality needs transworld identity or at least an objective transworld counterpart relation. Lewis's view is unable to pick out such a relation—indeed, Lewis is unwilling even to try—and the most one can do is to say that such a relation is "primitive". But this statement does not let one non-circularly pick out *which* primitive relation is *the* counterpart relation. True, one could pick it out as "the one that matches the true judgments about counterfactuals, counterparts and transworld identities", but since this relation was part of the *analysis* of counterfactuals, counterparts and transworld identities, saying this will be viciously circular and will fail to make clear that there is any such relation that in itself has some characteristic that makes it fitting to be the counterpart relation, as opposed to simply happening to be the one that matches our independent judgments about what is a counterpart of what. In the case of God there is no parallel problem, because we can pick out God uniquely, e.g., as "The first cause of all contingent beings" or "A personal being that is necessary." And there is no problem with picking out the relation of "causation", because it is one that, *pace* Hume, we have a direct understanding of in virtue of our experience of our own causality. [\[137\]](#)

Further objections concerned with the mysteriousness of God will be considered in Section 6.

3.2 Counterpart vs. identity theory

3.2.1 Transworld identity

We no longer need to espouse the paradoxes of counterpart theory, because the only serious objections to identity theory in Section 2.1 of Part III were based on the Lewisian assumption that all possible worlds concretely exist, which is no longer the case.

3.2.2 Haecceities

How exactly one cashes out transworld identity will depend on whether one thinks there are haecceities or individual essences. If for every possible individual there is a haecceity, a property that exists in every possible world, such that the individual has that haecceity essentially and necessarily no one else does, then we can simply define an identity between the individuals in terms of identities between haecceities.

But of course the theory of haecceities is a dubious one. For instance, Adams (1981, p. 11) observes that there is a problem about the relation of me to my thisness, namely the property of being me. It is not just that I have my thisness essentially. Rather, I have my thisness essentially and my thisness essentially has the property of not being had by anyone other than me. Thus my thisness has a property that cannot be specified without specifying me *de re*. It will not do to specify me *de dicto* here, e.g., by saying that my thisness essentially has the property of being had only by beings that have the property of being me, since that would be just making the trivial claim that my thisness essentially has the property of being had only by beings that have my thisness. But that would not distinguish my thisness from *having green-dyed hair* since *having green-dyed hair* essentially has the property of being had only by beings that have green-dyed hair. But if my thisness has a property that cannot be specified without specifying me *de re*, then it cannot exist if I do not exist. Alternately, my thisness is essentially related to *me* and hence can only exist if I exist since nothing can be related to what is not (*ibid.*). Therefore, my thisness cannot be a haecceity, since a given haecceity exists in all possible worlds. But my haecceity, were it to exist,

would indeed be my thisness, and hence I have no haecceity. ^[138]

The haecceitist might try to bite the bullet and say that if H is my haecceity, then I simply am *that entity that has H*. And there is no need to further specify that H can only be had by me, because that would be tautologous. This would in fact allow a reduction of *de re* modality to *de dicto* modality: the claim that I could not be a horse could become the claim that, necessarily, anything that has H is not a horse.

3.2.3 The identity of indiscernibles and a causal-historical account of identity

On the other hand, if we reject haecceities we have a more major problem. Kripke suggests that one can simply *stipulate* which individual in one world is identical with which individual in another world. That will not do, however, because the possible worlds under consideration are supposed to be complete representations, whose content God can actualize. Take two divine ideas of worlds, say w_0 and w_1 , and suppose that they respectively represent the existence of the individuals x_0 and x_1 . One cannot merely stipulate whether x_0 and x_1 the same, for that would mean that there would something more to God's making a world actual than actualizing the divine idea: God would have to *specify* which individuals he creates for which spots. And in this stipulational setting this can only be done if there are haecceities, which at the moment I am assuming do not exist.

Without haecceities it is difficult to see how on the Aristotelian-Leibnizian view one can avoid a strong principle of the identity of indiscernibles (at least for non-world-bound individuals, where an individual is world bound if it exists only at one possible world—it is unclear whether there are any such individuals). Suppose some world w_0 contains two indiscernible individuals, say x_1 and x_2 . Then, unless there are haecceities, the two individuals are represented in the same way by the divine idea that w_0 is. Suppose w_1 is another world containing at least one of these two individuals (and such a world will exist if the individuals are not world bound), and let y be one of these two individuals in w_1 . Then there must be a fact of the matter as to whether y is identical with x_1 or with x_2 , a fact that had better be a part of the content of the worlds if they are representationally complete, and yet a fact that there cannot be because w_0 represents x_1 and x_2 in the same way and hence any reason for taking y to be identical with one is also a reason for taking it to be identical with the other.

The fact that the proposed approach is committed either to haecceities or to the identity of indiscernibles does in fact count to some degree against it. Haecceities are hairy and controversial, while it seems *prima facie* possible for there to be two distinct individuals that are indiscernible. In regard to the apparent possibility of indiscernibles, I will make a similar reply to that I made to the Hume-Gale compossibility objection. Once we see possibility as tied to causality, the pressure to think indiscernibles possible will no longer be there. On the causal account, given that all things ultimately come down to one cause, namely God, if there are to be indiscernibles, then somewhere along the line the same cause acting at the same time and in the same manner must produce two indiscernible effects. Apart perhaps from some sub-atomic phenomena which can be variously interpreted ^[139], we do not have cases of the same cause acting at the same time and in the same manner to produce two indiscernible effects. So the source of confidence that this is possible cannot be from empirical observation. Rather, it must be from some intuitive notion that such a thing would be possible for an omnipotent God.

However, it is not clear that such a thing would indeed be possible for God. For it is not clear whether it would make sense to say that God has instantiated some complete individual concept C (a complete description of a possible individual) more than once as opposed to just instantiating some other concept once. I have defended this kind of a view in a Leibnizian context against Strawson's objections, and it is worth giving this defense at length.

Thus suppose that seven entities e_1, e_2, \dots, e_7 satisfy C . This can only lead to a non-Leibnizian ontology, but the implications will be explored within a framework as close to Leibniz's as possible. In particular, as in Leibniz's system, the concept C is to be given in purely universal terms.

First one can ask: In what sense are there seven entities satisfying C ? Suppose that I am observing these entities. It seems I would not be able to determine that there are seven of them, because the process of counting requires that I be able to see differences (at least in temporal or spatial position) between the

entities I am counting. To count them, I would have to first point out one of the entities. This is already a problem, because by pointing to one indiscernible entity, I point to all, and by describing one, I describe all (this also holds for counterfactual pointings, since an argument based on the Principle of Sufficient Reason would show that in all possible worlds in which one of the indiscernible entities exists, so do the others and they all have exactly the same properties ^[140]). But suppose that somehow I have managed to point out, say,

^[141]
 e_3 . I then utter the noise “One instantiation of C ” (like someone starting to count sheep: “one sheep”, “two sheep”, etc.) Now, as a child learns at an early age, one of the difficulties in counting is that one must beware not to count the same entity twice. Thus, I must next in some way point out an entity other than e_3 and utter the noise “Two instantiations of C ”. But because the entities are indiscernible, it is impossible for me to tell whether the second entity I have chosen is the same as e_3 or not.

Thus it is logically impossible for me (and perhaps even for God) to c o u n t the entities in the usual way by successive (possibly counterfactual) pointing out of the entities. In what way could I then possibly determine

^[142]
 that there are seven of them? The principle of pre-established harmony does entail that I must be able to do so, at least in principle, since my complete individual concept, including as it does the whole state of the universe, must also include the sevenfoldness of the instantiations of C .

Consider a specific example. Suppose I know that the concept C discussed above is a concept describing a point-source of light with brightness index one. I look in the direction of the instantiation of C , and see that I am faced with a point-source of light with brightness index seven. I might thus conclude that I am looking at seven instantiations of the concept C describing a point-source of light with brightness index one, though I am not able to distinguish them. However, I could equally well describe myself as facing a single instantiation of a concept C_7 which describes a point-source of light with brightness index seven. In other words, from the point of view of any other entity in the universe other than God and perhaps e_1, e_2, \dots, e_7 , there is no way of even in principle empirically recognizing that there are seven instantiations of C (this case will be called (a)) rather than one instantiation of C_7 (this case will be called (b)). The principle of pre-established harmony tells us that this is a major problem.

But perhaps it is possible for e_1, e_2, \dots, e_7 t h e m s e l v e s to recognize (after all, if they are to be separate monads, they will have perceptions) whether (a) or (b) is the case? It is here that Strawson’s insistence that we should talk of the points of view on analogy with consciousnesses strengthens the claim that Leibniz’s ontology need not preclude the possibility of multiple monads (i.e., entities analogous to consciousnesses) with the same point of view. After all, intuition suggests that if I were e_1 , surely I could tell that e_2, e_3, \dots, e_7 were not the same as I by means of a self–other distinction. But could I really? Suppose that I went into a room and I had the sense impression of a person looking exactly like I do, and performing exactly the same acts at the same time as I do. Assuming a lack of hallucination, the most reasonable inference would n o t be that I have an indiscernible copy. Rather, the most reasonable inference would be that I am looking at myself in a mirror or through some video-imaging system.

One can argue that the case of the indiscernible e_1, e_2, \dots, e_7 would be analogous to the mirror/video-camera case. If I were e_1 , I could never tell while looking at e_2 that I am not just perceiving my own self, i.e., e_1 , in some reflected way, and so I could not conclude whether (a) or (b) was the case. This point will be particularly forceful if we suppose that there is no intrinsically private first-person authority: that my knowledge of my own self, i.e., my governing monad’s knowledge of itself, is simply more distinct than my knowledge of other selves in light of the fact that my governing monad forms a clearer impression of itself than it does of other monads, but this self-knowledge is not different in k i n d from the sort of knowledge that other persons can have of me (after all, if there is perfect pre-established harmony between my states and the rest of the universe, then others will in p r i n c i p l e have enough data to infer my inner states). But if there is no intrinsically private first-person authority and if we further suppose that e_1 perceives e_1 equally distinctly as it does e_2, e_3, \dots, e_7 , then there is no way for e_1 to distinguish itself, even from its own point of view, from the other instantiations of C . In the point-source of light example, then, e_1 could not tell whether it is a single instantiation of C sitting together with six other such instantiations, or whether it is a lone instantiation of C_7 , if e_1 perceives itself equally clearly to e_2, e_3, \dots, e_7 . But even if we do not assume a lack of essentially first person authority and do not grant the supposition that e_1 perceives e_1 equally distinctly with e_2, e_3, \dots, e_7 , still e_1 could conclude that its perceptions of e_2, e_3, \dots, e_7 are simply self-perceptions with a different degree of distinctness, and e_1 (were a light-source capable of thought!) could reasonably think that it is a lone instantiation of a concept $C_7 \neq$ like C_7 , but with some further internal complexity caused by its seeing of itself with a greater distinctness in one way (what one might call, “ e_1 seeing e_1 ”) and with a lesser distinctness in another way (what one might call, “ e_1 seeing e_2, e_3, \dots, e_7 ”).

I now wish to suggest that there is no real ontological difference within Leibniz's system between God saying (if this were logically possible), "Let there be seven instantiations of C " and God saying, "Let there be an instantiation of C_7 (or maybe $C_7\phi$)." On both creation possibilities, there is a single monad by the proposed definition of a monad as "that which perceives the universe from a given point of view." On the first possibility, this monad is comprised by the "seven instantiations" e_1, e_2, \dots, e_7 of C —it is the compound e_1, e_2, \dots, e_7 that perceives the universe from the given point of view. On the second possibility, this monad comprises one instantiation of C_7 (or maybe $C_7\phi$). The first possibility does not contradict the simplicity of the monad, because for Leibniz the simplicity appears to be primarily a lack of spatial parts and Leibniz is certainly willing to allow for internal complexity of state in the monad; moreover, perfect indiscernibility makes for perfect indivisibility, since divisibility presupposes a potential for difference which is not available

[143]

here. But if we consider the concept C_7 (the case of $C_7\phi$ is a little more complicated, but a similar argument can still work), then in the above example this just describes a source of light which emits seven units of brightness. This concept does logically involve the concept of emission of a single unit of brightness, and a seven-fold instantiation of a single unit of brightness just is a single instantiation of seven units of brightness; in either case, by the proposed definition of a monad, this will be happening within a single monad. Talk of creating seven instantiations of C on this view is then just imprecise vernacular short-hand for talk of creating a single instantiation of C_7 (or maybe $C_7\phi$). Since in Leibniz's system it is a monad which is the one and only possible kind of entity, if we take seven instantiations of C to be literally seven different entities, we arrive at a contradiction. Thus, e_1, e_2, \dots, e_7 cannot be different entities—their "existence" can at most indicate different inner aspects (i.e., qualities) of a single monad.

The distinction between (a) "seven instantiations" of C and (b) a single instantiation of C_7 (or $C_7\phi$) is one that can be argued to be a distinction simply about words. A similar kind of distinction is discussed by Leibniz in a letter to de Volder, where the question is to distinguish between the theory that substances persist over time and the theory that God continually produces new substances which replace old ones (i.e., the view that substances are not temporally extended, but on each time slice there are different substances). Leibniz comments that to disagree about this would be to quarrel about a mere name, since "there is no further

[144]

principle in things by which such a controversy can be resolved". It would appear that the distinction between (a) and (b) is precisely a distinction of this kind.

One could object that while what was said above is plausible for light sources given that brightness of light (at least on the old wave theory) is considered a continuously variable phenomenon, it is not so obvious for higher level monads like the governing monads of human beings. Leibniz's commitment to the principle of continuity may help to bridge the gap here by saying that our apperception is also something that varies continuously in degree. Also, once we get away from Strawson's own view of consciousness, perhaps it will not be so unreasonable to suppose that if one of us were "instantiated twice" (in the same spatio-temporal location, of course) then still there would not be two monads, but a single indivisible one, perhaps with additional internal complexity. If "two" consciousnesses have the same perceptions (and thus also appetitions, since appetitions are just changes in perceptions), then the entity defined as "that which perceives the universe from a given point of view" is a compound entity consisting of the "two." But if they cannot be split apart, these "two" will simply be a single simple monad. There is, moreover, no problem with unity of apperception, since the "two" have the same view. In other words, it is not completely unreasonable to accept that in the place of the above description of "two", we should instead talk of a single monad (perhaps with additional internal complexity caused by the fact that it may have two ways of perceiving itself, one direct and one indirect—but certainly normal everyday mirrors allow one to have such multiple ways of perceiving oneself without vitiating one's unity of person). There are "two" consciousnesses there only in the sense that a car moving at 60 km/hr is undergoing "two" movements, each at a speed of 30 km/hr. Just as in the case of the moving car (and similarly in the case of the light-source described above) one cannot split apart or distinguish the two movements and therefore one considers them a single whole, so too the presence of the "two" consciousnesses in the proposed system simply indicates a single consciousness which at most can be said to be "twice as intense" in some way; in the case of the counting of indiscernibles, it appears that instead of an extensive or enumerative magnitude, there is an intensive magnitude.

Leibniz, thus, could simply deny that it makes sense to speak of "two" consciousnesses if their internal states are the same. In this way, he would strongly depart from Strawson's view of consciousness. (Pruss, 1998, pp. 117–121.)

Admittedly, there is an intuitive plausibility to the possibility of two distinct but indiscernible consciousnesses that are non-embodied, or embodied in a single body, or embodied in two indiscernible bodies. But unless we can give an account of how they could be caused, and of necessity in an indiscernible way since otherwise one could distinguish them by their causes, this intuitive plausibility can be rejected. The intuition may

count somewhat against the Leibnizian-Aristotelian account, but it is by no means decisive. However, recall that the account can also escape the criticism by espousing haecceities.

But in fact, even without haecceities, we can also try turn this issue into an asset for the proposed theory by also giving an explicit account of transworld identity that naturally comes out of the spirit of the Leibnizian-Aristotelian account. The usefulness of having such an account will count in favor of the theory, outweighing the somewhat counterintuitive consequence above which needs to be tolerated if we are to have a theory of transworld identity. The proposal here is that identity of the causal chain (the causes and the manners in which they act) leading up to the coming-to-be of an individual is the essential individuating property of individuals at the moment of coming to be with ordinary criteria of persistence being applied afterwards. This is a strengthening of Kripke's weaker claim that the matter an individual was originally made out of is individuating for that individual. For a recent discussion of this, see Mackie (1998).

To decide whether an individual in one world and an individual in another world are identical, then, we need to decide whether the same chain of causes led up to them. This is an approach that can be taken only by someone who believes that all chains of causes necessarily go back to a first cause. For the criterion gives us no way of deciding whether a chain of causes that has no first cause is identical with such a chain of causes in another world. Likewise, it gives no account of the identities of first causes, and hence works best with an account on which there is necessarily only one such.

This account of identity *does* have counterintuitive consequences. For instance, one might think that the same mass-produced chair could have been made by a different workman had the one who made it called in sick that day, assuming the work is routine. This objection can be answered in a two-fold way. First of all, there is no fact of the matter about the identification of artifacts between or even within worlds. One of the things that makes something an artifact is that its essence depends on the purposes that its artisans and users have. If I should suddenly choose to treat the chair as a work of art, then the identity of the workman would be the identity of an artist, and hence an essential property. If I should choose to treat the chair as a mere lump of matter, then the identity of the workman would be irrelevant—only the identity of the matter would be relevant. Objective reality does not decide between these options—our attitudes may, but need not do so completely.

However, the identity of artifacts does yield useful intuitions which is why philosophers have always employed it. Now if in cases where there is an objective matter of fact as to the identity of an object its identity is determined by the identity of the chain of causes leading up to it, then it would be plausible to suppose that at least one should say that in cases where the identity depends on our attitudes, we should see those causes that we consider *contextually relevant* to be determinative of identity. So the example of the chair cannot be dismissed entirely. However, in fact, this consideration acts in favor of my proposed account. The reason we do not take the identity of the workman to be an essential property of the chair is that we do not consider the workman to be one of the contextually relevant causes of mass-produced items. Rather, in cases of mass production, we see the designer as acting through the workmen, with the workmen being almost interchangeable tools. For moral reasons, one wishes people did not see it this way, but in the case mass-production they do. The identity of the workmen is contextually irrelevant here as part of the chain of causes. But should the designer have been different, then because we see her as a *relevant* cause, we would say the chair is different. If, on the other hand, we took the workman to be more of an artist, someone whose causal role we would take to be more important, then she would become a more relevant cause and we might see the identity of the chair as depending on her identity. Hence, a causal account of identity may even apply in the case of the contextually-identified artifacts, providing we restrict ourselves to the contextually relevant causes.

On the other hand, the identity of *natural* entities, such as the matter making up the chair, or human beings or animals is an objective property. Certainly, this is so in the case of human beings if certain moral notions are objective and not dependent on our attitudes. Thus, if we think there is an objective matter of fact whether *x* is morally bound by some promise, we will likewise have to think that there is an objective matter of fact whether *x* is

identical with the individual who made the promise. Likewise, if there is an objective matter of fact whether one is responsible for some deed, then likewise there will have to be an objective matter of fact whether one is identical with the deed's doer.

One might of course argue that it *is* logically possible for the same human being to have been conceived by different parents, *pace* this view. I know of no better answer to this than Kripke's (cf. Kripke, 1980, p. 114n56). Suppose in the actual world w_0 I am conceived by TW (my father) and IZ (my mother) and a world w_1 where I am conceived by AB (a man) and CD (a woman), who are not identical with TW and IZ. Now, surely there is a possible world w_2 in which TW and IZ have a child at the same time as they do in the actual world and which child has the same genetic makeup as I do, while AB and CD *also* have a child at the same time as they do in w_1 . But then, surely, that child that AB and CD have in w_2 is the same child as they have in w_1 . After all, the actions of TW and IZ should not be relevant to the identity of the child of AB and CD (assuming enough causal isolation). And likewise, surely, that child that TW and IZ have in w_2 is the same child as they have in w_0 . Hence, in w_2 , both the child of AB-CD and the child of TW-IZ is I, since it is me that AB and CD have in w_1 and TW and IZ have in w_0 . Hence in w_2 I am conceived by two different sets of parents—and that surely is absurd.

The above argument only shows I could not have been conceived by a pair of parents having no one in common with my parents. But this seems to be enough to show that the exact identity of the pair of parents I am descended from is an essential property of me. For, for a *reductio* suppose that I could have been descended from AB and IZ. Let X be the descendant of AB and IZ in that world. Then, by the same token, I could have been descended from TW and CD, and let Y be the descendant of TW and CD in that world. Now, we can imagine a world where both AB and IZ have a child, and this child is identical with X, and where TW and CD have a child, and this child is identical with Y. But in that world X and Y are non-identical. But X and Y are both identical with I. Hence I am not self-identical, since X and Y are non-identical, which is absurd. Hence, I could not have been descended from AB and IZ.

On this approach, it does however appear to follow that there cannot be indiscernibles—or even indiscernible initial stages!—that are directly created by God, since there would be nothing whereby the causes, namely the direct acts of God involved, could be distinguished. If this is right, and if one does not believe in haecceities, then one will have to embrace the perhaps counterintuitive consequences of the Principle of Identity of Indiscernibles for entities that are the second step in the chain of causes that starts with God. This counterintuitive consequence counts against the theory, but because in ordinary thought we have no reason to think of universes which at the start have indiscernibles in them, the intuitions gone against here will not be ones whose abandonment would require revision of anything that is of importance to us.

3.2.4 The divine ideas as differing between worlds

Alternately, one might try to resolve the problem that indiscernibles pose for the proposed account of possibility by denying that the divine ideas of possible worlds encode all the information about possibilities. What the divine ideas fail to encode is information on which member of a set of indiscernible individuals in one world is identical with which individual, perhaps no longer indiscernible, in another world. To see how an account along these lines would work, consider an example. Jones and Smith are indiscernible individuals in the actual world—they always do the same thing, and so on, and they presumably exist on two mirror planets or the like. Suppose next Saturday both will mow their lawns. Nonetheless, it is possible

(59) that Jones will fail to mow his lawn while Smith still mows his.

And it is possible

(60) that Smith will fail to mow the lawn while Jones still mows his.

These are two distinct possibilities which correspond to a single general divine idea of a world, assuming there are no haecceities. However, if the world containing the indiscernible Jones and Smith is actual, then we have the resources to individuate the two possibilities in terms of that which grounds them. The first possibility is grounded

in Jones' power to refrain from mowing the lawn while the second is grounded in that same power of Smith's. Since these two powers belong to different individuals, howsoever individuated, they are distinct powers.

One way to handle this is through relativizing certain possibilities. Suppose that it is possible for there to be two indiscernibles, Smith and Jones, in some non-actual world w_0 , and neither Smith nor Jones exists in the actual world. Then unless there are haecceities, there is no necessarily existent proposition, i.e., divine idea, that handles the two asymmetrically—alas, it seems not all propositions are necessary beings on this view. For instance, the proposition that says that Smith mows the lawn cannot be a necessary being, though there is a proposition that says that at least one of Smith and Jones mows the lawn. To see this, suppose that the proposition p that asserted that Smith mows the lawn existed in all worlds. Then Smith would be the unique person who possesses the property of being identical with that person in w_0 whose action makes p true (or false, as the case might be). Moreover, this property would be one that would exist in all worlds if p exists in all worlds. Hence, this property would in effect be a haecceity, since it would distinguish Smith from all his indiscernible copies and would exist in worlds where Smith does not.

In the actual world, then, all existent modal propositions involving boxes and diamonds can be explicated in terms of quantification over possible worlds even if the possible worlds fail to distinguish between worlds that verify possibility (62) and those that verify possibility (63), because there are no propositions that distinguish between these two. There is one divine idea corresponding to both possibilities: the idea of a world where one of Smith and Jones mows the lawn and the other does not. On the other hand, in w_0 , there are two divine ideas corresponding to the possibilities, because there are now more propositions in existence. Since Jones and Smith exist, there are propositions about them, and God has individual ideas about them. So in w_0 there are two possibilities.

The only possible paradox here is that what ideas God has now differs from world to world. This does present a *prima facie* difficulty given doctrines, which I in fact accept, that say that there is no contingency in God himself, but there is no reason to think it presents any more difficulty than the fact that in some worlds God believes unicorns exist and in others God does not—something that can be reconciled with the doctrine of the lack of contingency in God if one holds that the same divine state can count as a belief in one proposition in one world and as a belief in another in another, or if one holds that the divine ideas are not in fact parts of God but some sort of dependent beings. Given that these theological issues are also present in views that do not involve possible worlds, I shall not discuss them further.

Observe, finally, that this view is contrary to S5. According to S5, if it is possible that p then it is necessary that it is possible that p . This now would have to be revised to claim that if it is possible that p , then it is possible that p at every world in which p exists. For propositions that do not make reference to indiscernibles, the latter is no restriction, however, and so S5 still holds for them. This might be enough to keep advocates of S5 happy. In my discussion in the rest of the thesis, I shall not specifically consider this view, so if it turns out to be the right one, slight modifications may have to be made to a few statements.

3.2.5 Conclusions about identity

So there are in fact three possible solutions to the problem of indiscernibles. (1) Allow haecceities. (2) Bite the bullet and deny the possibility of indiscernibles while presenting a simple historical-causal account of transworld identity. (3) Allow the collection of divine ideas to differ between worlds.

3.3 Attributions of ability

The objection to Lewis's view (Section 2.2.2 of Part III) as well as to the ersatzist approaches (Section 3.6 of Part IV) on the grounds that statements of the form "I am able to f " should only be about me and not about persons in other worlds or abstracta of course disappears immediately on the present approach, of course, since if I can be the cause of something, those capacities whose exercise made me the cause are a truthmaker for the

proposition that reports the possibility of that thing's existence. That is no surprise, of course, given that it is precisely to handle ability claims that the Aristotelian account was developed.

3.4 No set of all possible worlds

I have argued in Section 7 of Part III that on no reasonable account of possibility is there a set of all possible worlds and that this is a serious objection to Lewis's view, since on his view worlds are physical objects at ontological ground level so that there *should* be a set of them. After all, they are paradigmatically the sort of thing that makes up a set (assuming, of course, a set theory with ur-elements). One might think that the very same argument works against the theory I have proposed. After all, the worlds on this theory are in a sense concrete: they are ideas, not of the Platonic variety, but actually existent in a mind. Should there not be a set of them, then?

To this I answer that while physical-type entities *are* paradigmatic of the sorts of things from which we build sets, mental things are not. There is no such thing, for instance, as the set of all the pains I have. The reason for this is that pains are not something about whose individuation there is a fact of the matter. (This should not worry us. Pains are not substances, after all.) If my legs hurt, is the pain in my left leg something different from my pain in my right leg? If so, what about the pain in the lower three inches of my left leg and the pain in the next three inches? But there *is* always a fact of the matter as to how many members a set contains: having a cardinality is one of the basic properties of sets that can be proved from any reasonable axioms for sets. Neither is there such a thing as the set of all of my ideas. Is my idea of the bachelor different from my idea of the married man? When is an idea that is implicit in another idea to be counted as separate from it? There is little plausibility to thinking that such questions always have determinate answers.

Thus, ideas and the like are *not* paradigmatic ur-elements of sets in the way that physical objects, or more generally substances, are. Now, the above *specific* reason why my ideas and my pains do not form a set does not apply to the divine ideas of worlds. Those ideas can be individuated by the logical maximality and mutual incompatibility of their contents (unless there is a problem with individuating worlds like those in which (62) and (63) respectively hold—see Section 3.1, above). However, having seen that ideas in general are not paradigmatic ur-elements of sets since not all collections of ideas are sets and having seen that the paradigmatic ur-elements are substances, the burden would be on one claiming that some special ideas *do* form a set to argue for this claim. I am not aware of such an argument. Moreover, the argument would have to have premisses more plausible than the evidence for the theory under consideration.

3.5 The ethical and probability-theoretic objections

The ethical objections to Lewis's theory have no force here. It is clear that we have no reason to be concerned in the same way about things in worlds that God has not actualized as about things in the world that God has actualized, since the former things do not exist—only ideas of them do.

The probability theoretic objections do not apply either. The world in which some event that in fact occurs and that had a $2/3$ probability of occurring is not on par with the actual world, because it is objectively non-existent—it is an unactualized idea, one to which nothing physical corresponds.

3.6 The inductive objection

The inductive objection to Lewis's theory in fact turns into a positive asset for the present theory. For it is an asset of a theory of actuality that it should be clear why it is that, if the theory is true, one has more reason than not to believe in induction. God is a personal being, a rational being, and so there is reason to think that he would act as one would expect rational beings to do so. One would not expect rational beings to act haphazardly. One would expect a certain order to the states of affairs they bring about. The fact that we have found such a state in the past gives us reason to believe that God *does* value a certain amount of order in the universe, just as orderliness in

any rational being's past activity gives reason to believe that a certain amount of order is valued by the being.

One might try to make a Goodmanian objection. Consider two theories. On one, the universe is always orderly. On another, it is gruesome, i.e., it is orderly *except* that emeralds are grue as opposed to green: they are green before the year 2002 (we need to update the dates, given that the flow of time has disproved the original gruesome hypothesis!) and blue thereafter. Now all the evidence we have about God is just as compatible with God valuing ordinary order as with God valuing gruesome order. Hence we have no more reason to think emeralds are grue than that they are green.

However, this reply neglects the fact that part of what makes a being rational is that the being responds to objective values. Order in the universe *is* an objective value. Gruesomeness is not. It is not the sort of thing to which a rational being is sensitive. Thus we have *a priori* reason to think that God would be more responsive to order than gruesomeness as a value. This is not enough to give an *a priori* proof that God would create an orderly universe. After all, God can be responsive to a value but choose to act on an incompatible value, such as the value of an utter diversity of things and events in the universe. However, that *a priori* reason taken together with our past observations of the universe which showed that there were consistently obeyed laws there *does* give one reason to prefer the hypothesis of the universe being ordinarily orderly over the hypothesis of gruesome order.

Section 4 Rescher's challenges no longer a problem

Neither are Rescher's challenges to the very notion of possible worlds a problem now. True, we cannot specify possible worlds or individuate them in the strong sense of actually stating individuating properties of them. But there is no similar problem about *God* individuating these worlds.

Section 5 Is God omnipotent if logical possibility is defined in terms of his power?

One might worry that if logical possibility is defined in terms of God's power, then the claim that God is omnipotent, i.e., capable of being the first cause of any possible state of affairs, is tautologous, since to be a possible state of affairs and to be capable of having God as first cause are the same. ^[145] As far as that goes, this is true. However, this claim together with the Aristotelian-Leibnizian analysis of possibility does have content.

Traditional theism does not shy away from allowing important statements about God to be tautologies. Christianity, for instance, is willing to say that God is identical with perfect divine love, and does not consider this claim to be empty by virtue of the fact that, like all necessary identity claims, it is tautologous. The reason the latter claim is not empty is that we have some independent epistemic access to what love is. And, likewise, we have independent epistemic access to what is logically possible, so that saying that the logically possible is nothing but the possible for God *does* convey some information to us.

We have very good reason to believe, for instance, that it is possible for the universe not to have contained any living things, and so upon learning that logical possibility is nothing but divine ability to be the first cause of a given state of affairs we learn that very likely it is possible for God to have created a world without any living things. *How* we know that it is possible for the universe not to have contained any living things is, of course, a more difficult question. But it is difficult on *any* view of possibility, with the Aristotelian-Leibnizian view not being an exception. Though we can come up with an argument once we have the Aristotelian-Leibnizian view firmly in place. A universe without living things seems quite conceivable to us. What seems conceivable does not *by itself* give us solid reason to believe that the thing is possible. But if additionally it seems to us that we could tell an outline of complete coherent causal story of such a thing, then we have good reason to think such a thing is possible. And we *can* tell such a story. Perhaps God chose not to act on the value of there being contingent life-forms, but simply on the value of there being contingent beings. After all, God had no imperative to create a universe at all, and to create a universe without life is better than to create no universe at all. So it is imaginable that God should choose to do this. And it being imaginable that God should choose to do this, it is likewise

imaginable that it should be caused by God.

Section 6 How we know what is possible

We have already discussed one aspect of the problem of how we know what is possible in Section 5. In the case of global possibilities, such as there not being any life forms in the universe, it is not so easy to give an account of how we know the apparent possibility really to be such, though a rough-and-ready account in terms of imagining a causal story can be given. However, a similar problem will appear on any view of possibility.

What the present view does better at giving a natural account of than other views is our knowledge of local possibilities, such as that of Hitler never having existed. Possibility is nothing but capability, and we inductively know the capabilities of human beings. Thus, we know that Hitler's parents had the capability of remaining abstinent on the occasion when Hitler was conceived, and if we have good reason to believe, as we do (see Section 3.2.3, above), that Hitler could not have been conceived on another occasion, we thereby know that Hitler could have failed to exist. Likewise, I know that it is possible that I finish this sentence, because I know my capabilities and that finishing this sentence is among them. More interestingly, it is plausible that it is possible for there to be horse-like animals with a single horn because it is plausible that an evolutionary story could be sketched under which they would evolve. In fact, it seems that in general our best bet for proving that something is logically possible is to show that it is physically possible according to our best scientific theories, and this technique coheres very well with the present account.

Moreover, the Aristotelian-Leibnizian account is perhaps one's only hope for an account of knowledge of necessary truths that has a *causal* component. What the necessary truths are is grounded in the nature of the power of God, and this power is causally relevant to everything in this universe, and in particular to our thinking about necessary truths. Leibniz had an illuminationist epistemology of necessary truths according to Adams (1989, Section II.7.2.2); on the present view, such an epistemology actually has a causal component. It may not be necessary even to go all the way to an illuminationist epistemology to have an epistemology of necessary truths and modal claims compatible with the Leibnizian-Aristotelian account which epistemology has a causal component. Unlike in the case of Lewis's theory or theories that ground possibility in Platonic causally impotent things, there is some hope for the ontology of possibility and necessity to be causally relevant to the actual way by which we get to know of this ontology.

Section 7 The mystery and obscurity objections

Finally, one might challenge the whole account on the grounds of an unacceptable obscurity. After all, God is ontologically mysterious. He is an immaterial, necessary and concrete being, and all the concrete beings we normally meet with are material and contingent. Moreover, the concept of God is an obscure one. Bringing in God, this objection goes, to elucidate modality is explaining the somewhat obscure by the Obscurity of Obscurities Itself.

7.1 The ontological mystery

Yes, the God of the account is an immaterial, necessary and concrete being. There should be no unacceptable puzzlement over the first two parts of this. After all, on most accounts there must be beings that are both immaterial and necessary. For instance, it is difficult to see how the truthmakers of necessarily true mathematical claims could fail to be immaterial necessary beings. Moreover, there is good reason to believe in propositions, and these are immaterial and necessary. Being immaterial is no bar to being existent, and there is nothing *prima facie* absurd about a necessary being.

Of course Hume has argued that any being that can be imagined to exist can also be imagined not to exist. As a general principle this is already refuted by the mathematical cases, though it might be patched up by restriction to concrete cases. However, on our Aristotelian epistemology of modality, a state of affairs is only possible if there

actually is some causal factor that could have brought it about. To claim that the non-existence of some entity is imaginable is, then, to claim that something actual could have brought it about that this entity did not exist. But there is nothing absurd about the idea of an entity such that nothing could have brought about its non-existence. Hume's principle makes imaginability into possibility, whereas it is not enough to just imagine a thing in isolation, but one must imagine how it fits into the whole causal sequence. Of course if, like Hume, one does not believe causation is anything over and beyond constant conjunction, this reply will not impress Hume. But then the challenge is for Hume to come up with a satisfactory alternate account of modality, and none such is available as has been argued in Part III and Part IV.

Another Humean objection is to the very notion of a God as creator, following ideas of Grünbaum who has argued that God cannot cause the universe to exist, because causation involves statistical laws. However, the only reason to think that causation *as a concept* involves statistical laws is if one thinks that the kinds of causal influences that science studies are the only logically possible kind. And there is no reason for the assumption, unless one takes a scientific view on which the scientific usages of concepts are the only or primary ones. On the other hand, there is a *prima facie* possibility of personal causation, of the sort that many people believe to take place when a person intentionally brings about a certain effect.

More interesting is Grünbaum's objection that divine causation does not satisfy the Aristotelian practical syllogism, since there are no purposes for which the action is performed. Filling in the details, if there *were* such purposes, they would either be necessary or contingent. If they were necessary, then this universe would be necessary, which is absurd. But if they are contingent, then, since all contingent things have causes, they must have been brought about by a previous choice of God's, and the same problem reappears at that stage, thereby creating a vicious regress.

However, there is a regress-stopper, and that is a libertarian choice. In this kind of a choice, a person in the very same mental state can directly cause one of a number of alternate effects. The coherence of such an event can be supported by reflection on a situation where a person is choosing between two courses of actions, each indicated by different incommensurate value-systems: If I do *this*, I act immorally, and if I do *that*, I act against my best interest. One might say that as a necessary truth in cases like that there is some underlying reason one has which ensures that one makes the choice one does. However, there is no phenomenological reason to posit such a reason, and there is no sound reason not to take the phenomenology at face value as at least reflecting, if not the way things are, the way things coherently could be. It is true that we *do* sometimes say after the fact things like "My conscience was stronger than my selfishness", but it is also plausible phenomenologically to suppose that this is an *ex post facto* judgment: my conscience's "being stronger" than my selfishness may consist in nothing else than that I followed my conscience.

One might think that God would not be making a choice between different value systems, since being all-good and all-knowing he will make every choice in accordance with the objectively correct value system. That is true, but nonetheless God is choosing between different valuable things, often not commensurable ones. There is a certain kind of value to human beings; but there is also a certain kind of value, perhaps an incommensurable one, to a universe containing only spiritual beings. There is a value to Smith existing and there is a value to Jones existing, but it might be that the existence of the two is logically impossible (e.g., if the Kripkeans are right about the time of conception being an essential property, and if both are different possible persons who are not twins but who would be conceived at the same time by the same parents in their respective possible worlds) and the two values are incommensurable—one person need not be either more, or less, or equally valuable as another. Between *such* values, at least, God can make a libertarian choice: "I will pick *this* valuable thing." What is the reason for the choice? It is the intrinsic value of the thing, just as when I choose between two alternatives, each touted by a different value system, the reason I will finally give for my choice will be the value that I have settled on. This reason does not necessitate the choice, of course. But it is the reason. One might also ask: Why did you act on *this* value? The answer is the same: Because of its value. This is a regress-of-explanations ender, a complete

explanation.

Admittedly, we can say very little about the values that animated God's choice. But this ignorance is no more harmful to this theory of modality than the fact that Newton could say very little about the digits of the gravitational constant is harmful to his theory—indeed, we still can only say something about finitely many of these digits, while it has infinitely many. Every interesting theory reveals new unknowns, and this is only valuable.

7.2 Obscurity

Pressing further the objection that we do not know much about God, it does not seem like the concept of God is one that we are very clear about. After all, although we know enough to individuate God from all other possible beings, this is not much at all. What are divine ideas? What are God's powers and choices like? All these are invoked in the theory, but they are shrouded in obscurity.

However, they are not in complete obscurity. Analogy plays a crucial role here. The divine ideas are related to God in a way analogous to the way our ideas are related to us. Divine powers and choices are related to God in a way analogous to the way our powers and choices are to us—and we *do* have powers and *do* make free choices. These analogous claims are not much, and the analogies have their limits. But they do render the theory determinate enough to be open to discussion, and indeed more determinate on various points, e.g., the nature of propositions which is now seen to be that of divine ideas, than competing theories. Analogical reasoning in constructing theories is not to be brushed aside in general. For instance, as philosophers such as Sellars have argued, analogical reasoning plays a crucial role in scientific reasoning. The particles of the atomic theory of gases were seen as analogous to hard billiard balls. The analogy had its limitations, but gave some determinateness to the content of a theory whose ontology would otherwise be objectionably obscure.

Part VII. Final conclusions

Section 1 Cost-benefit arguments for the Aristotelian-Leibnizian ontology of possible worlds

1.1 Modality

The Aristotelian-Leibnizian (AL) model does give an account of the existent realities that are spoken of when we make modal claims. Moreover, it succeeds in doing this without making any implausible distinction between being-in-this-world and being-existent as Lewis does and without invoking any *objectionable* primitive modality as the ersatz views were seen to. In a precise sense, the possible worlds exist in the actual world: it is true at the actual world that possible worlds exist. But this is not paradoxical: The possible worlds are just divine ideas, one of which is actualized, and the actualized one contains within itself an account of all the other divine ideas. (Obviously, they must be infinite lest the regress prove vicious.)

Moreover, the account squares nicely with the intuition that attributions of the possibility of doing something should be claims precisely about the beings to whom this possibility is attributed, something which neither Lewisian nor ersatzist approaches allow for.

1.2 Explanation of what propositions are

As a non-modal bonus, the AL model provides us with a deeper insight into the nature of propositions, which it says are divine ideas. Admittedly, the nature of divine ideas is somewhat unclear, and so this clarification is only partial. However, we do have an analogical grasp of the nature of divine ideas based on our grasp of our own ideas, and this provides us with a better understanding of what propositions are than just considering them to be bare Platonic abstracta that represent realities in some mysterious way. Divine ideas represent realities analogously to the way that our ideas represent realities.

It is true that David Lewis's EMR also purports to give an account of the nature of propositions. Unfortunately, his reduction of propositions to collections of worlds is unacceptable whereas his "structured propositions" involve too much arbitrariness (see Section 4 of Part I). It might be argued that our present view also contains some arbitrariness. Why should we define "propositions" as *God's* ideas rather than the ideas of someone else? But there is good reason for this. For one, God is the only necessarily omniscient being, and presumably also the only being that necessarily has ideas of all realities. For another, the ground of the possibility of the content of those ideas that are of possibilities is the power of God: it is God's choice that is necessarily the first branching in the "history" (in quotes as God's choice might not be in time) of the cosmos, and it is appropriate the propositions be found at that level.

1.3 Not a completely new ontology

The AL model *does* involve a rich ontology containing God, an ontology strictly richer than that of the ersatzist. For, all the items in the ersatzist's menagerie exist on the AL model: there are propositions, i.e., divine ideas, and there are sentences in all possible languages, i.e., divine ideas of sentences in all possible languages. However, this richness is necessary to do justice to modality, as can be seen from the arguments against the ersatzist views in Part IV. Very briefly and schematically, an account of modality that involves merely propositions or sentences makes the ground of modality merely "descriptive": for the Parmenidean problem to be resolved, there needs to be an "executive power" by virtue of which the ersatz worlds are made *possible*, and this is God.

However rich the AL model's ontology may be, it is not revisionary in the way that EMR is. There is evidence for the ontology of the AL model, i.e., for the existence of God, that is prior to the AL model itself. Indeed, this ontology has historically been widely believed, and the historical success of this ontology makes it

plausible that there will be significantly fewer revisions to important areas of our thought such as ethics, induction and probability, whereas EMR's ontology entails many such revisions.

1.4 Connection with theistic arguments for the existence of God

If traditional theism is true, then God is the "ground of being", the first cause of the existence of all things outside himself, indeed the first cause of all contingent states of affairs. Moreover, good reasons can be given for why on the theistic view God is to be taken to be *necessarily* the first cause of all beings outside himself. First of all, if one brings God in as a first cause, as an explanation of all things other than himself, then to avoid Schopenhauer's "taxi cab" objection to the cosmological argument (Schopenhauer charged that the causal principle behind the cosmological argument was dismissed once the existence of God was proved, like a cab that is no longer needed once one is at the destination, and not applied to God himself) one must affirm that God is the explanation of his own existence, perhaps by there being a sound ontological argument, though possibly outside of our grasp, for his existence or by his existence being implicated by his essence. Moreover if God's actions are to provide the ultimate explanation of all contingent truths in *one* world, then Jerome Gellman (2000) has argued persuasively that the same must hold in all worlds and that God must necessarily be omnipotent. Thus, theism should hold, and many strands of it explicitly do, that God is *necessarily* the ultimate ground of all contingent being.

But if this is so, then traditional theism comes very close to the proposed Aristotelian-Leibnizian view. The only additional step that this view takes is that not only is it the case that God must be the first cause of the realization of any possibility, but that its being a possibility should be *analyzed* in terms of God. Theoretical simplicity and Ockham's razor are in favor of taking the two concepts "That which God can be the first cause of" and "That which is possible" which have the same extension and positing, as the Aristotelian-Leibnizian view does, that they are the same concept. Thus, if there is a God who is the ground of being, it is plausible to suppose him also to be the ground of possibility.

Consequently, all arguments for traditional theism also lend plausibility to the present view, and certainly, there are many such arguments (see, for instance, Swinburne, 1979). And, conversely, the present view's theoretical virtues in explicating modality lend plausibility to the view's truth and thus to traditional theism the central tenets of which the view entails.

Observe, too, how the AL account of possibility not only gives an account of what possibility is but also gives a partial explanation why what is actual is actual: namely, because of God's creative act.

Of course, similarly, many arguments against traditional theism are also arguments against the present view. One exception is the argument from evil, which need only be handled *only* if one wishes to supplement the basic AL model with a claim that the God in whom all possibility is grounded is perfectly good, something that I take to be the case but which is not a part of the AL model itself. And, aside from the argument from evil, there are few serious objections to theism. Substantiating the claim that none of the other putative objections to theism is a serious one is, of course, not a *completely* trivial task, but limitations of space do not permit it to be attempted here.

Section 2 Conclusions

We need possibility and probably also possible worlds. Without these, the collection of all truths would be seriously impoverished. Paradoxically, there are facts about the actual world that cannot be expressed without bringing in possibilities. However, this is going to be less paradoxical when we realize that the ontological ground of possibilities is the actual world, something that *must* be the case if we are to escape the force of the intuition that actuality and existence are the same whereas one cannot talk truly of what is not.

None of the extant serious accounts of possible worlds are fully satisfactory. This thesis very briefly sketches a model that takes from Aristotle the notion that potentialities are grounded in actual features of substances and from Leibniz the notion of possible worlds as the ideas of God, assuming the latter notion can be

made sense of. This model overcomes various fatal objections to the other theories, objections ranging from their lack of a solution to the Parmenidean problem to paradoxical conclusions such as that there is no such thing as inductive knowledge. Furthermore, the model gives a clarification of the nature of propositions. Since we need a theory of modality, and would like to have propositions and possible worlds, and the present theory is coherent and avoid egregious paradox, this itself is good reason to believe the Aristotelian-Leibnizian model to be true. As Leibniz said, “il semble que c’est quelque chose de considerable qu’une hypothese paroisse *possible*, quand toutes les autres ne le paroissent point, et ... il est extremement *probable* qu’une telle hypothese est la veritable.” (Gerhardt (1960–61), vol. III, p. 353). But, furthermore, the present theory receives additional evidential support from the various extant arguments for theism.

Lewis said that his theory was to be believed until one came along that gave the same benefits for a lower price. If one is optimistic about making sense of a faculty of intentionality in a non-natural being, this has indeed happened—though not *all* benefits are shared by the new theory, but only those that were not seen in Part II to be chimerical, like the alleged set-theoretic reduction of propositions.

Stepping back, we see a number of evaluations of contemporary theories that have been argued for along the way, showing that depending on how much we demand from our theory of possibility we can make-do with different ontologies. It is easiest to summarize these results with a table (see Table 1).

If all we want from our theory is:	Then we can:	At the price of:
Possible worlds.	Take a pure Platonic view of possible worlds as abstracta such as collections of propositions.	Not having an explanation of the intentionality of propositions or of the truthmakers of modal claims.
Possible worlds and something to say on the problem of the intentionality of propositions.	Accept the Leibnizian view of possible worlds as thinkings in the [146] mind of a God.	The primitive intentionality of a divine mind.
Truthmakers of modal claims.	Accept the Aristotelian view on which possibilities are grounded in causal powers and dispositions.	Not having global possibilities or possible worlds.
Truthmakers of modal claims and global possibilities.	Accept the Aristotelian view together with a necessary first cause without commitment to this first cause being a theistic one.	Not having possible worlds.
Everything: Truthmakers of modal claims, global possibilities, possible worlds, something to say on the problem of the intentionality of propositions, and whatever other theoretical benefits the existence of God provides.	Accept the Aristotelian-Leibnizian model.	The primitive intentionality of a divine mind and any other difficulties involved with the notion of God.

Table 1: Summary of Conclusions

APPENDIX

Appendix: Reinterpretation of functional systems: A sketch

For simplicity, consider a deterministic functional system, A , consisting of n interacting parts, each of which can be in one of at most k states, functioning from time 0 to time N , and functioning in a temporally discrete way so that the relevant states are all to be looked at integer values of the time variable, ranging from 0 to N . I assume the physically required transitions of the system are time-independent—if not, then just assume a clock is a part of the system.

If one is a functionalist, one can think of such a system as being a person (perhaps taken together with her environment and community, depending on what kind of story one tells) that is thinking of something. I will argue that one can map such a system onto the simple system, B , consisting of one photon moving through space relative to some fixed three-dimensional coordinate system (perhaps fixed by some fixed physical body that defines it). For simplicity assume a Euclidean space-time and start the system at time 0, with the photon constrained to be at the origin then.

To ward off a possible objection before proceeding further, note first that the parts of a functional system need not be implemented as physically distinct parts: all that matters is the pattern of causal interaction between the states of the different parts. When we perform an ideal computer simulation of a functional system, the *physical* parts of the simulating system need not map onto the physical parts of the simulated system in any relevant way. There would, for instance, be nothing wrong with emulating a system which uses, say, k binary memory modules, each capable of storing one of two values encoded as an electrical potential of 0 and 1 volts, respectively, with a hypothetical quaternary memory system containing half the number of memory modules, but with each capable of storing one of four values, encoded as an electrical potential of, say, -1 volts, 0 volts, 1 volts and 2 volts, respectively. We want to say that such an emulation would preserve functional interconnections. Otherwise, the basic insight of functionalism, namely multiple implementability of functional systems, is severely limited, since seemingly paradigmatic cases of functional equivalence would then fail to be cases of functional equivalence (e.g., the evident functional equivalence of a Z80 emulator running on my Pentium system with an actual Z80 system will be rather doubtful, since it is by no means obvious that what is implemented by k parts in one system will also be implemented by k parts in the other system).

We can represent the states of A as n -tuples (a_1, \dots, a_n) , where the a_i are integers between 1 and k , inclusive. Let A_1 be the set of physically possible states of A . For any state s in A_1 , let $e_0(s), e_1(s), \dots$ be the infinite sequence of states at integer values of time generated by starting A in state s and then letting it evolve in its normal way; note that $e_0(s) = s$. Now, what I shall call a pre-state of the photon in system B is something that can be represented as (\mathbf{x}, \mathbf{v}) where \mathbf{x} is a spatial position vector and \mathbf{v} is a velocity vector—of course only the direction of \mathbf{v} matters, since its magnitude must be equal to the speed of light c . Moreover, I shall restrict myself to those pre-states where in fact \mathbf{x} and \mathbf{v} point in the same direction, and count system B , for the functional purposes in question, as “broken” otherwise (the physical implementation of a functional system will of course have physically possible configurations in which the system is “broken”, and its states have no functional meaning—in a living being death is a paradigmatic case). Let B_0 be the set of all such pre-states.

Now, let \mathbf{f} be a function that assigns to each state s in A_1 a velocity vector of magnitude c . Let s_0 be some arbitrary fixed state in A_1 . We now define a function F from B_0 to S_0 as follows. Suppose first that (\mathbf{x}, \mathbf{v}) is such that \mathbf{x} and \mathbf{v} have the same direction (relative to the origin) and such that there exists a state s in A_1 with $\mathbf{f}(s) = \mathbf{v}$ (note that this can only hold for finitely many values of \mathbf{v} since A_1 is a finite set). Let i be the unique integer satisfying $i\ell|\mathbf{x}| < i+1$, where $|\mathbf{x}|$ is the distance \mathbf{x} from the origin. Then, put $F(\mathbf{x}, \mathbf{v}) = e_i(s)$. Consider now any pre-state (\mathbf{x}, \mathbf{v}) not satisfying the above conditions. Then, define $F(\mathbf{x}, \mathbf{v}) = e_i(s_0)$, where again i is the unique integer satisfying $i\ell|\mathbf{x}| < i+1$.

Now, define the equivalence relation \sim on B_0 by saying that $(\mathbf{x}, \mathbf{v}) \sim (\mathbf{x}^*, \mathbf{v}^*)$ if and only if $F(\mathbf{x}, \mathbf{v}) = F(\mathbf{x}^*, \mathbf{v}^*)$. We can then define the states of B to be the equivalence classes under the equivalence relation \sim . [147] Let B_1

be the set of such states. We can then define a map G of B_1 to A_1 by letting $G(a)=F(\mathbf{x}, \mathbf{v})$, whenever a is the equivalence class containing (\mathbf{x}, \mathbf{v}) . It is easy to observe that G is one-to-one and onto. Moreover, it preserves causality in the following sense. Let us restrict ourselves to discrete points in time. Then, if a is a state of B which causes a state b to occur one unit of time later, then $G(a)$ is a state of A which causes a state $G(b)$ to occur one unit of time later. The desired counterfactual relations obtain as well:

Were B in state a at time i , it would be in state b at time $i+1$
is true if and only if

Were A in state $G(a)$ at time i , it would be in state $G(b)$ at time $i+1$

If we wish to, we can define states of virtual parts of the system B , too. For, observe that A_1 is a set of n -tuples of integers. We can then say that the i th part of our system B now is in state j if and only if the i th member of the n -tuple $G(a)$ is equal to j , where a is the state that B is in now. Admittedly, the parts are only virtual, but that is not a problem for functional equivalence, as seen before.

Therefore, if functionalism is right, we can reinterpret the flight of a single photon as a thinking person.

And this is absurd. [\[148\]](#)

BIBLIOGRAPHY

Bibliography

- Adams, Robert M. "Theories of Actuality", *Noûs*, 8 (1974), pp. 211-31; reprinted in Loux, *The Possible and the Actual*.
- "Middle Knowledge and the Problem of Evil", *American Philosophical Quarterly*, 14 (1977), pp. 109-117; reprinted in Adams, *The Virtue of Faith* and Gale and Pruss, *The Existence of God*.
- "Primitive Thisness and Primitive Identity", *Journal of Philosophy*, 76 (1979), pp. 5-26.
- "Actualism and Thisness", *Synthese*, 49 (1981), pp. 3-42.
- *The Virtue of Faith and Other Essays in Philosophical Theology*. Oxford University Press, 1987.
- *Leibniz: Determinist, Theist, Idealist*. Oxford University Press, 1994.
- Alston, William. "The Inductive Argument from Evil and the Human Cognitive Condition", *Philosophical Perspectives*, 5 (1991), ed. by J. E. Tomberlin, pp. 29-68; to be reprinted in Gale and Pruss, *The Existence of God* (forthcoming).
- Ariew, Roger and Garber, Daniel. *G. W. Leibniz: Philosophical Essays*. Hackett, 1989.
- Aristotle. *De Interpretatione*.
- *Metaphysica*.
- *Physica*.
- Armstrong, D. M. *A Combinatorial Theory of Possibility*. Cambridge University Press, 1989.
- Augustine of Hippo. *The Confessions*. Trans. John K. Ryan, Doubleday, 1960.
- Barker, Stephen. "Counterfactuals, Probabilistic Counterfactuals and Causation", *Mind* 108 (1999), pp. 427-469.
- Bigelow, John. "Possible Worlds Foundations for Probability", *Journal of Philosophical Logic*, 5 (1976), pp. 299-320.
- Castañeda, H.-N. "Thinking and the Structure of the World", *Philosophia*, 4 (1974), pp. 3-40.
- "Fictional and Reality: Their Basic Connections", *Poetica*, 8 (1989), pp. 31-62.
- Chihara, Charles S. *The Worlds of Possibility: Modal Realism and the Semantics of Modal Logic*, Clarendon, 1998.
- Chisholm, Roderick. "Identity through Possible Worlds: Some Questions", *Noûs*, 1 (1967), pp. 1-8; reprinted in Loux, *The Possible and the Actual*.
- Craig, William Lane. "Defense of the *Kalm* Argument", in *Reasonable Faith: Christian Truth and Apologetics*, Crossway, 1994, pp. 91-100. Reprinted in Gale and Pruss (eds.), *The Existence of God* (forthcoming).
- Cresswell, M.J. "The World Is Everything That Is the Case", *Australasian Journal of Philosophy*, 50 (1972), pp. 1-13; reprinted in Loux, *The Possible and the Actual*.
- *Logics and Languages*. Methuen, 1973.
- Davis, Wayne A. "Indicative and Subjunctive Conditionals", *Philosophical Review*, 88 (1979), pp. 544-564.
- Dirac, P.A.M. *The Principles of Quantum Mechanics*, 4th edition, Oxford, 1987.
- Edginton, Dorothy. "On Conditionals", *Mind*, 104 (1995), pp. 235-329.
- Elga, Adam. "Self-locating Belief and the Sleeping Beauty Problem", *Analysis*, 60 (2000), pp. 143-147.
- Feidman, Fred. "Counterparts", *Journal of Philosophy*, 68 (1971), pp. 406-9.
- Forbes, Graeme. "Two Solutions to Chisholm's Paradox", *Philosophical Studies*, 46 (1984), pp. 171-87.
- Forrest, Peter. "Occam's Razor and Possible Worlds", *Monist*, 65 (1982), pp. 456-64.
- "Ways Worlds Could Be", *Australasian Journal of Philosophy*, 64 (1986).
- and Armstrong, D. M. "An Argument Against David Lewis' Theory of Possible Worlds", *Australasian Journal of Philosophy*, 62 (1984), pp. 164-168.
- French, Steven. "Quantum Physics and the Identity of Indiscernibles", *British Journal for the Philosophy of*

Science (1988), pp. 233-246.

Gale, Richard M. (ed.) *Blackwell Companion to Metaphysics*. Blackwell (forthcoming).

— *On the Nature and Existence of God*. Cambridge University Press, 1991.

— and Pruss, Alexander R. “A New Cosmological Argument”, *Religious Studies*, 34 (1999), pp. 461-476.

— and — (eds.) *The Existence of God*. International Research Library of Philosophy, Ashgate Publishing (forthcoming).

Gellman, Jerome. “Prospects for a Sound Stage 3 of Cosmological Arguments”, *Religious Studies*, 36 (2000), pp. 195–201; reprinted in Gale and Pruss, *The Existence of God*.

Gerhardt, C.I. *Die philosophischen Schriften von Gottfried Wilhelm Leibniz*. 7 vols. Olms Verlagsbuchhandlung, 1960–61.

— *Mathematische Schriften: G. W. Leibniz*. 7 vols. Olms Verlagsbuchhandlung, 1962.

Grim, Patrick. “There is No Set of All Truths”, *Analysis*, 46 (1986), pp. 186-191.

— “Worlds by Supervenience: Some Further Problems”, *Analysis*, 57 (1997), pp. 146-151.

Haack, Susan. “Lewis’ Ontological Slum”, *Review of Metaphysics*, 33 (1977), pp. 415-29.

Hare, R. M. *The Language of Morals*. Oxford University Press (1964).

Heller, Mark. “Five Layers of Interpretation for Possible Worlds”, *Philosophical Studies*, 90 (1998), pp. 205-214.

Hunter, Graeme, and Seager, William. “The Discreet Charm of Counterpart Theory”, *Analysis*, 41 (1983), pp. 73-6.

Jackson, Frank. “A Causal Theory of Counterfactuals”, *Australasian Journal of Philosophy*, 55 (1977), pp. 3-21.

Jeffrey, Richard C. *The Logic of Decision*. 2nd ed. University of Chicago Press (1983).

Jubien, Michael. “Could This Be Magic?” *Philosophical Review* (1991), pp. 249-267.

Kant, Immanuel. *Critique of Pure Reason*. Trans. by Norman Kemp Smith. St. Martin’s Press (1965).

— *Dritique of Practical Reason*. 3rd edition. Trans. by Lewis White Beck. Macmillan (1993).

Kirk, G.S., Raven, J.E. and Schofield, M. *The Presocratic Philosophers: A Critical History With a Selection of Texts*. 2nd edition. Cambridge University Press, 1983.

Klagge, James C. “Supervenience: Perspectives Versus Possible Worlds”, *Philosophical Quarterly*, 37 (1987), pp. 312-315.

Koons, Robert C. “A New Look at the Cosmological Argument”, *Philosophical Quarterly*, 34 (1997), pp. 171-192.

Kripke, Saul. *Naming and Necessity*. Harvard University Press (1980); previously printed (except for the preface) in *Semantics for Natural Language*, ed. by Donald Davidson and Gilbert Harman, Reidel, 1972.

Krips, H. “Irreducible Probabilities and Indeterminism”, *Journal of Philosophical Logic*, 18 (1989), 155-172.

Kuratowski, K. and Mostowski, A. *Set Theory: With an Introduction to Descriptive Set Theory*, New York: North-Holland, 1976.

Lappin, Shalom. “Why One Ought to Care About Identity Across Possible Worlds in Moral Judgments”, *Ratio*, 22 (1980), pp. 167-169.

Leftow, Brian. “A Leibnizian Cosmological Argument”, *Philosophical Studies*, 57 (1989), pp. 135-155.

Leslie, John. “The Prerequisites for Life in Our Universe”, *Truth*, 3 (1990), pp. 97–119.

— “A Neoplatonist’s Pantheism”, *The Monist*, 80 (1997), pp. 218-231.

— “The Divine Mind”, in: *The Good, the True and the Beautiful*, ed. by A. O’Hear, Cambridge University Press, 2000.

Lewis, David. “Anselm and Actuality”, *Noûs* 4 (1970), pp. 175-88; reprinted in Gale and Pruss, *The Existence of God* (forthcoming) and, with added postscripts, in Lewis, *Philosophical Papers*, volume 1

— *Counterfactuals*. Blackwell, 1973.

— “Counterfactual Dependence and Time’s Arrow”, *Noûs*, 1979a, pp. 455-476.

- “Attitudes *De Dicto* and *De Se*”, *Philosophical Review*, 88 (1979b), pp. 513-43; reprinted, with added postscripts, in Lewis, *Philosophical Papers*, volume 1.
- *Philosophical Papers*. 2 vols. Oxford University Press, 1983 and 1986.
- *On the Plurality of Worlds*. Basil Blackwell, 1986a.
- “Against Structural Universals”, *Australasian Journal of Philosophy*, 64 (1986b), 25-46.
- “Sleeping Beauty: Reply to Elga”, *Analysis* (forthcoming).
- Linsky, Bernard. “Truth Makers for Modal Propositions”, *Monist*, 77 (1994), pp. 192-206.
- Loux, Michael. *The Possible and the Actual: Readings in the Metaphysics of Modality*. Cornell University Press, 1979.
- Lycan, William G. “The Trouble with Possible Worlds”, in Loux, *The Possible and the Actual*.
- “Two—No, Three—Concepts of Possible Worlds”, *Proceedings of the Aristotelian Society*, 91 (1991), pp. 215-227.
- “The metaphysics of possibilia”, in Gale, *Blackwell Companion to Metaphysics*.
- *Modality and Meaning*, Kluwer, 1994.
- Mackie, Penelope. “Identity, Time, and Necessity”, *Proceedings of the Aristotelian Society*, 98 (1998), pp. 59-78.
- McCall, Storrs. “Counterfactuals Based on Real Possible Worlds”, *Noûs*, 18 (1984), pp. 463-478.
- Melia, Joseph. “Against Modalism”, *Philosophical Studies*, 68 (1992), pp. 35-56.
- Meyer, Robert K. “God exists!”, *Noûs*, 21 (1987), pp. 345-361.
- Monk, J.D. *Introduction to Set Theory*, New York: McGraw-Hill, 1969.
- Nolan, Daniel. “Recombination Unbound”, *Philosophical Studies*, 84 (1996), pp. 239-262.
- Parmenides. *Peri Phuseôs*. In: Kirk, Raven and Schofield, *The Presocratic Philosophers*.
- Parsons, “Prolegomenon to a Meinongian Semantics”, *Journal of Philosophy*, 71 (1974), pp. 561-581.
- *Nonexistent Objects*. Yale University Press, 1980.
- Place, Ullin T. “‘*De re*’ Modality without Possible Worlds”, *Acta Analytica* (1997), pp. 129-143.
- Plantinga, Alvin.
- “Transworld Identity or Worldbound Individuals?” in *Logic and Ontology*, ed. by Milton Munitz, New York University Press (1973); reprinted in Loux, *The Possible and the Actual*.
- *The Nature of Necessity*. Oxford University Press, 1974.
- “Actualism and Possible Worlds”, *Theoria*, 42 (1976), pp. 139-60; reprinted in Loux, *The Possible and the Actual*.
- “Two Concepts of Modality: Modal Realism and Modal Reductionism”, in *Philosophical Perspectives, 1: Metaphysics*, ed. by J.E. Tomberlin, Ridgeview Publishing (1987).
- *God and Other Minds: A Study in the Rational Justification of Belief in God*. Cornell University Press, 1990.
- Pruss, Alexander R. “Induction and Actuality”, manuscript.
- “The Cardinality Objection to David Lewis’s Modal Realism”, *Philosophical Studies*, 104 (2) (2001) 167-176.
- “Śaṅkara’s Principle and Two Ontomystical Arguments”, *International Journal for the Philosophy of Religion* (forthcoming).
- “The Actual and the Possible”, in Gale, *Blackwell Companion to Metaphysics* (forthcoming).
- “*De re* modality, Rawls and abortion”, manuscript.
- “David Lewis’s Counterfactual Arrow of Time”, manuscript.
- “Lying and Speaking your Interlocutor’s Language”, *The Thomist*, 63 (1999), 439-453.
- “Leibniz’s Approach to Individuation and Strawson’s Criticisms”, *Studia Leibnitiana*, 30 (1998), pp. 116-123.

- Putnam, Hilary. "Realism and reason", in *Meaning and the Moral Sciences*. Routledge and Kegan Paul, 1978.
- "Models and Reality", in *Realism and Reason*. Cambridge University Press, 1983, pp. 1- 25.
- Quine, W.V.O. *Word and Object*, M.I.T. Press, 1960.
- "Ontological Relativity", *Journal of Philosophy*, 65 (1968), pp. 185-212.
- Rescher, Nicholas. *Leibniz: An Introduction to his Philosophy*. Rowman and Littlefield, 1979.
- *G.W. Leibniz's Monadology: An Edition for Students*. University of Pittsburgh Press, 1991.
- "How Many Possible Worlds Are There?", *Philosophy and Phenomenological Research*, 59 (1999), pp. 403-420.
- and Brandom, Robert. *The Logic of Inconsistency: A Study in Non-Standard Possible-World Semantics And Ontology*, Rowman and Littlefield, 1979.
- Richards, Tom. "The Worlds of David Lewis", *Australasian Journal of Philosophy*, 53 (1975), pp. 105-118.
- Roper, Andrew. "Toward an Eliminative Reduction of Possible Worlds", *The Philosophical Quarterly*, 32 (1982), pp. 45-59.
- Roy, Tony. "Worlds and Modality", *Philosophical Review*, 102 (1993), pp. 335-362.
- Shalkowski, Scott A. "The Ontological Ground of the Alethic Modality", *Philosophical Review*, 103 (1994), pp. 669-688.
- Skyrms, Brian. "Possible Worlds, Physics and Metaphysics", *Philosophical Studies*, 30 (1976), pp. 323-32.
- Stalnaker, Robert. "A Theory of Conditionals", in *Studies in Logical Theory*, ed. by Nicholas Rescher, Blackwell, 1968.
- "Possible Worlds", *Noûs*, 10 (1976), pp. 65-75; reprinted in Loux, *The Possible and the Actual*.
- "Possible Worlds and Situations", *Journal of Philosophical Logic*, 15 (1986), pp. 109-123.
- Strawson, P. F. *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen, 1959.
- Swinburne, Richard. *The Existence of God*. Clarendon, 1979.
- "The Irreducibility of Causation", *Dialectica*, 51 (1997), pp. 79-91.
- Thomas Aquinas. *Summa Theologiae*.
- *Summa Contra Gentiles*.
- *Commentary on Aristotle's Metaphysics*.
- Turner, Donald A. "The Many-Universes Solution to the Problem of Evil", in Gale and Pruss, *The Existence of God*, ed. by Richard M. Gale and Alexander R. Pruss, Ashgate Publishing (forthcoming).
- van Fraassen, Bas C. "The Only Necessity is Verbal Necessity", *Journal of Philosophy*, 74 (1977), pp. 71-85.
- van Inwagen, Peter. "Indexicality and Actuality", *Philosophical Review*, 89 (1980), pp. 403-426.
- "Two Concepts of Possible Worlds", *Midwest Studies in Philosophy*, 11 (1986), pp. 185-213.

[1] Following Lewis's (1986a) practice, by a "unicorn" I shall simply mean an animal that looks much like the "unicorns" of our stories; I shall also assume that for ten generations the animal was born to one that looked like the "unicorn" of our stories, just in case somebody tells me that at some country fair she attended there was a "unicorn". I do not mean any specific natural kind of such animals, and hence avoid Kripkean objections.

[2] In conversation, Rescher has suggested that one can make do with talking about novels which specify some but not all aspects of a world rather than worlds here. The problem here (closely analogous to the problems Lewis [1986a, Section 3.2] raises for some linguistic theories) is with specifying what the implications of a novel are and with specifying which novels count as "possible". With regard to the problem of implications, since the novels specify only some aspects of the world, and since it is important for Rescher that the novels be ones that can be specified by us, much that the novel says will not so much be said as *entailed*. There is a problem, however, with specifying what a novel entails, since "entailment" is a quintessentially modal concept. Moreover, it might well be that the truth of what would happen were Hitler never to have been born depends on aspects of our world that we have no idea of at all (e.g., it might depend on some non-human, perhaps even non-material, beings that we

cannot even conceive of). A theory of counterfactuals should not prejudge this possibility. But when we talk in terms of Rescherian novels, we need to specify what aspects of our world are relevant and go into the novels. And this we might not be able to do. In fact, we might be radically wrong as to what aspects of the world should go into the novels, and the answers we get for counterfactual questions based on such novels would simply often be false—a falsity that a novel-based account cannot account for.

[3] Note that this library might even be a contingent being, in which case it would be contingent that there are possible worlds (a contingency which is useful since perhaps the language the library is written in will lack meaning in other worlds). This contingency would be represented on this crazy theory by saying that all the books other than the one describing the actual cosmos say somewhere: “And there is no Platonic library.” Observe how this involves one in no evident contradiction, though it leads to a denial of S5.

[4] It is impossible, even for an omniscient mind, to individuate impossible beings. How would one married bachelor differ from all others? Since a contradictory description entails all other descriptions, if x is a married bachelor, and y is any other individual, possible or not, then likewise x has the property of being identical with y . (The proof of this, discovered in the middle ages is: Since x is a married man, x has the property of being identical with y or married. But x is a bachelor, hence not married, and hence must be identical with y .) I do not in general know how to conceive of an impossible being.

[5] These are of course to be distinguished from the sentences “Cicero has the name ‘Cicero’” and “Cicero has the name ‘Tully.’” Unlike (3), “Cicero has the name ‘Cicero’” is not a necessary truth as Cicero might easily not have received that name.

[6] Loux (1979), p.16.

[7] Fr. 291. Throughout, I am following, with modifications, the translation of Kirk, Raven and Schofield (1990).

[8] Fr. 293.

[9] Alternately, one could follow David Lewis in rejecting the truthmaker theory as it is, and say instead that what makes a proposition true is there existing certain things and/or there not existing certain other things, and that the non-existence of something might be important, too. I will not discuss the Lewisian extension of the truthmaker theory further, because the same puzzle about, e.g., what existent thing makes it true that there could be unicorns remains if we allow that the proposition that there could be unicorns could be made true not just by the existence of something but also by a combination of one thing’s existence with another’s non-existence. (Certainly, the non-existence of unicorns does not make it true that unicorns are possible.)

[10] I am grateful to Robert Brandom for discussion of this issue.

[11] Fr. 296.

[12] *Ibid.*

[13] This third route is taken by Gale and Pruss (1999) in their argument for the existence of God based on the *weak* PSR which states that for every proposition p , it is *possible* that p

[14] *Ethics*, prop. XXIX.

[15] Actually, both Rescher and Leslie are willing to draw theological conclusions from their theories. It is for the best that there be a God, after all. Leslie (2000), though not Rescher, goes so far as to posit an infinite number of deities, since the more, the better. (This is of course problematic in light of the set theoretic fact that there is no largest infinity, as the same reasoning would say that it is always better to have a higher infinite cardinality of the set of deities than a lower infinite cardinality. In correspondence, Leslie invoked Cantor’s murky notion of an absolute infinity as an answer.)

[16] The best answer I know of is the combination weak theodicy championed by Alston (1991); see the discussion in Gale and Pruss, *The Existence of God*, Introduction, Section III.4.

[17] There is one version of the optimality principle that is capable of answering the problem of evil, and this is Donald Turner’s (forthcoming) view according to which all universes which have a favorable balance of good over bad are created by an all-good deity, the separate universes being unrelated spatio-temporally. The fact that our universe is non-optimal is not a problem for the theory as long as the balance of good over evil is favorable. And, overall, the world which is the mereological sum of all these universes is optimal. Rescher himself does not agree with this approach, since this large world consisting of many universes as a whole lacks the coherence and order of an optimal world. Some of the objections I shall levy against Lewis in Part III, below, can be retooled to work against Turner.

[18] Hippolytus, in: Kirk, Raven and Schofield (1990), p. 417, translation modified slightly.

[19] Gerhardt (1960–61), vol. III, p. 350.

[20] Gerhardt (1960–61), vol. III, p. 353.

[21] For instance, it may seem absurd to think the content of our beliefs, which content is of course a proposition, changes. If today I believe that it is now raining, I would not say tomorrow, even though it may be a quite sunny day, that I have since “changed my mind” and now “believe differently”.

[22] In fact, the universe could not lack all the things it has. For instance, it could not lack the truthmakers of mathematical propositions, which are presumably the same in all possible worlds.

[23] I am not denying S5 here by claiming that it might have been that there was only one possible *world*—though Lewis will see it this way. Rather, I am claiming that it surely might have been that there was only one *universe*, namely only one maximal spatio-temporally interconnected object. Lewis will take that as tantamount to a claim that there might have been only one world, but the intuitions behind my claim are intuitions about possibilities for physical reality—for what sorts of spatio-temporal objects might exist—and not for the space of possibilities.

[24] See, e.g., Swinburne (1979) for a survey of the evidence.

[25] Or, if one believes that propositions are intermediate between ideas and truthmakers, where is the truthmaker of the proposition expressed by the idea?

[26] This is a variant of an argument I heard from James V. Allen.

[27] This is parallel to the Sellarsian definition of an experience’s being an *of-X* type experience, namely that *f* is an *of-X* type experience if and only if were an *X* to be presented to one in the appropriate way, *f* would occur.

[28] Cf. my “*Śaṅkara’s Principle and Two Ontomystical Arguments*”.

[29] Lewis (1979a), p. 465.

[30] The reason I write “almost” is that there might be one kind of counterexample that some people’s common-sense might approve of: the idea of a divine intervention that is influenced by *future* prayers. But aside from the expressly miraculous, common sense does not admit of counterexamples to the direction of causation. Of course, there are difficult technical details in precisely formulating a statement of this intuition, especially when one is working in a context of relations between propositions. The future actions of the American people at the next presidential election *can* bring it about that Tipper *had* married the 43rd president of the United States. This is so because the description “the 43rd president of the United States” is temporally impure since who the referent of this description is depends on what happens in the future.

[31] Lewis (1979a), p. 465.

[32] E.g., Davis (1979).

[33] Kit Fine, review of David Lewis, *Counterfactuals* (Oxford: Blackwell, 1973), *Mind* 84 (1975), 451–458, p. 452, quoted in Lewis (1979a), p. 467.

[34] Lewis (1979a), p. 472.

[35] For instance, intuitively one might think that very close approximate similarity of particular fact over a much larger area (factor 4) outweighs the value of exact match throughout a small area. The world which is exactly like ours for all times in our future but whose past was *radically* different from the past of the actual world is surely further from our world than is a world which is the same as ours for all time except that the background radiation is everywhere and everywhen about $10^{-1,000,000,000}$ percent higher.

[36] Lewis (1979a), p. 468.

[37] *Ibid.*

[38] Actually, it is a little more complicated because the beam arrives in a more energetic state than it started on the moon. The reason for this is that in w_0 the atmosphere absorbed some of the energy of the moon-pointed beam, and so the beam arrived at the moon less energetic than it started at the laser. However, in w_0^* the reverse of this process happens: the beam is aligned in such a way that it picks up energy from the atmosphere along the way, this energy corresponding to that lost by the corresponding beam in w_0 .

[39] The reason it is *close* but perhaps not equal is that there may be different external effects in B_1^{**} and B_1 . We were assuming that we’re working with a billiard ball system that is approximately isolated, presumably likewise in our counterfactual worlds since they’re formed by merely small modifications of the actual world, and so the evolutions from E in these two worlds are going to be close.

[40] Indeed, libertarians take it that if determinism is true and if we have free will, then all our free actions are the results of

miracles in the brain effected by a non-physical self.

[41] We would just as intuitively judge that were Nixon to have pressed the button a nuclear holocaust would have ensued even if Nixon were forever locked in a chamber whose only causal connection with the outside world is a unidirectional wire that can set off the doomsday device and which chamber would disappear from existence entirely right after the potential button press, so that if Nixon pressed the button and yet a miracle prevented the signal from traveling outside to the outside world would have been no different from the case where Nixon did not press the button. Yet in such a world it would no longer be true on Lewis's analysis that were Nixon to press the button a nuclear holocaust would ensue, since the world in which a first miracle has Nixon press the button and a second small miracle prevents the signal from leaving would be closer to the actual world than a world in which just the first miracle occurs, since it would have *exact* spatio-temporal match in the future of the pressing of the button.

[42] If we are willing to confine our account to worlds that are indeterministic, on the grounds that counterfactuals in deterministic worlds are inescapably problematic, this exception can be dropped. If we keep it, we can either make it a restriction, or weight the value of match in a time-dependent way, with earlier event weighted much less than later ones.

[43] This account also differs from Lewis's in that it places exact match in the past at a higher priority than large scale nomic match in comparing worlds. In doing so, the counterfactual worlds to be considered may involve greater miracles, but one will eliminate the small degree of backtracking that Lewis admits his own account to have (see Section 3.1). We could instead make exact match in the past take second place to large scale nomic match, if we did not mind this small degree of backtracking, and did not want counterintuitive results such as that were Nixon to have pressed the button, a discontinuity in the spatial position of his finger would have occurred at the time of pressing it.

[44] The reason I say "all but one" is that if they *all* fainted, it would perhaps not be a *victory* for Napoleon over the English. But if one kept on fighting for a second or two, it would be a victory.

[45] Arguably, if the proposition that $1=1$ is a self-explanatory propositions, as seems likely, then the proposition that (($1=1$) and ravens are black) also explains that $1=1$.

[46] Of course, sometimes changing the order of conjoining may change the meaning of the component sentences, especially if these sentences involve pronouns. But then this is not really a case of the *same* sentences being conjoined, if the identity of sentences requires the identity of those contextual features needed to fix the meaning of its pronominal (and proverbial and prosential, etc.) terms. If one disagrees with this, then my claim will have to be restricted to sentences not having any pronominal (or proverbial or prosential, etc.) terms.

[47] Of course, one might say that the unstructured proposition $\{w_1, \dots, w_n\}$ (in the interesting cases with n infinite) is built-up as a disjunction of the propositions $\{w_1\}, \dots, \{w_n\}$, but this desperate and implausible (implausible as it would entail that the seemingly uncompound propositions that there exists an electron and that $1=1$ are in fact both infinitely disjunctive propositions) measure only delays the problem since the $\{w_i\}$ will be simple propositions.

[48] And allowing the regress will not help either since ultimately the structured propositions are built-up out of unstructured ones, and thus have to be basic. If the construction is infinitary nothing is gained here.

[49] A theist might attempt to argue that necessarily only God is all-good and only God is all-just, while being all-good and all-just are distinct basic properties. But this attempt is open to an objector who takes the scholastic view that in fact, God being simple, there is only one property that he has, which is in fact identical with himself.

[50] For instance, I myself have doubts about infinite complexity on *theistic* grounds in that I take each proposition to be *graspable* and yet one that is infinitely complex may not be graspable even to an omniscient intellect. But these grounds definitely will not be appealed to by Lewis!

[51] If not, we should work instead with the weakest symmetric transitive relation generated by spatio-temporal relatedness.

[52] Lewis (1970), pp. 184–185.

[53] Unless considerations of origin are essential to individuation, but what I say is true in a Lewisian setting.

[54] The main argument is, of course, the insistence that counterfactual properties apply to the same individual. But a point in space-time also has counterfactual properties (e.g., there being an electric charge-density r_1 at it, whereas in fact the electric charge-density at it is r_0).

[55] Plantinga (1987) and Lycan (1994 and forthcoming) have both criticized Lewis by saying that what happens in other Lewisian worlds has nothing to do with modality. This is no different from saying that what happens in other similar circumstances has nothing to do with causality or nomic modality, which is in danger of begging the question against the Humean. Moreover, one must be careful here. At least as read by Lycan (1994, p. 84), Plantinga claims that discovering facts about other concrete universes would not tell one anything about modality. Actually, *epistemically* it would provide *some* modal information. That someone very much like me becomes a biologist in some world is very good evidence for the claim that *I* could have become a biologist. And what better evidence can there be for the claim that there could be a unicorn than that there

is a unicorn? What discovering facts about other concrete universes would not do, I will argue, is provide *truthmakers* for modal claims.

[56] What a false proposition p is about presumably needs to be defined counterfactually, e.g., by saying that a false p is about x if and only if, were p true, p would be about x . Or one might want to relativize what entities a proposition is about to worlds. Thus, a proposition p is about x in w if p is true at w and it is true at w that p is about x . If a proposition is not actually true, then on this approach we cannot say what the proposition is about *simpliciter*, but can say what it is about in any given world in which it is true. This kind of relativization is necessary to give an account of disjunctive propositions which might be about different things in different worlds.

[57] *Per impossibile*, I believe.

[58] These examples are not sufficient to determine what we are to say about disjunctive cases such as “A horse exists or a dog exists”, or even “A horse exists”, but we do not need to determine the answer to this for the purposes of the following discussion. One solution would be to allow that a given proposition can have more than one truthmaker. It is, after all, reasonable to say that any horse is a truthmaker of the proposition that *a horse exists or a dog exists*, just as any dog is. If so, then we have two possible readings of the claim that p is about x if and only if x is involved in p 's truthmaker. We could say that p is about x if and only if x is involved in every truthmaker of p . Or we could say that p is about x if and only if x is involved in some truthmaker of p . The stronger reading is preferable if we do not want to say that the proposition that *a horse exists or a dog exists* is about Genghis Khan's third horse, and the spirit of the stricter definition fits better with the sense of “about” in the claim that the proposition *that I can jump* is about me.

[59] The argument in Section 10 as given is easy to adapt to a case where there are finitely many copies of each world *and* the number of copies does not differ from world to world. The argument there is a *reductio* of the possibility of there being objective probabilities not equal to 1/2. If the number of copies of each world were to depend on the world, there *could* be objective probabilities equal to numbers other than 1/2. However, these probabilities would always be rational numbers by an argument similar to the one given there. But is it not possible, as indeed present physics asserts to probably actually happen, that there be events whose objective probabilities are irrational numbers? Thus, the argument does indeed extend to cover the case of finitely many copies of each world. (But the extension of the variant argument in note 117, below, will not apply in such a case.)

[60] Note that although my arguments in Section 7 will presuppose the possibility of different infinite numbers of photons existing in the same place, they can be formulated in a way that is consistent with this interpretation, by talking of a single “ n -fold Photon”, with n infinite.

[61] E.g., the inductive argument in Section 9 will work as long as the induction-friendly worlds do not have more indiscernible copies than induction-unfriendly ones, and on this specific account of the number of copies of a world there is no reason to think they would.

[62] On some non-Lewisian theories, the actual world is not concrete either, but *corresponds* to our concrete universe. E.g., if a world is a maximal set of compossible propositions, a world w corresponds to our concrete universe if and only if all of its members are true. Such worlds are also ontologically on par with respect to their intrinsic properties, but one of these worlds is distinguished absolutely by being related in a representational way to the actual cosmos.

[63] The opponent of EMR may not be happy to define a universe as a maximal spatio-temporally related aggregate. She may prefer to simply define a universe as the aggregate of all there is. In non-Lewisian contexts, the reader is free to interpret “universe” in this way. Alternately, the following discussion can be read as predicated on the useful fiction, which Lewis alas thinks to be a fact, that it is not possible that there be in actuality two non-spatio-temporally related entities.

[64] Following Stalnaker (1976), Lycan (1994) distinguishes two claims about the indexicality of actuality. There is a merely semantical claim about how to understand sentences using the word “actual” and a more metaphysical claim that the actuality of the actual world is relational. It is the metaphysical claim that is of most interest in this thesis. Stalnaker had only claimed that only the semantical is independent of EMR. Lycan claims that the metaphysical one is, as well. For, Lycan (1994, p. 31) says that the Lewisian can after all say that one of the possible worlds is designated in a special way—and that is the world that is absolutely actual. Unfortunately, this fails because it is a contingent matter which world is thus designated, and EMR cannot account for this contingency as we have seen above. Curiously in the section of his text (pp. 30-34) where Lycan advertises the independence of EMR and the claim that actuality is indexical, there is only an argument that EMR and the denial of the indexicality thesis could be true, and no argument that the denial of EMR could go along with the indexicality thesis. As we have seen, the falsity of EMR entails that there is only one cosmos, and then the world which corresponds to that cosmos is actual. Besides, the various referents of an indexical expression are ontologically on par: the point of an indexical expression is to distinguish between them. Each referent of a tokening of “here” is equally real; likewise, if actuality is indexical, each referent of a tokening would be equally real. One might follow Stalnaker in claiming that the indexicality thesis is trivial, though. Assuming EMR is false, necessarily, the only tokenings that are real are actual ones. Hence, every tokening of “actual” occurs in the actual world, and so indeed “the actual world” refers always to the world in which it is tokened. However, this would be a very strange indexical, completely disanalogous to all other indexicals which in fact *are* tokened in various circumstances in which they have different referents. What is distinctive about an indexical is that to know what it refers to one must know the circumstances of utterance—knowing the referent of an indexical involves *a posteriori* knowledge. This is not the case with “the actual world” if EMR is known *a priori* to be false, as I think it can be known to be. For assuming I know *a priori* that

EMR is false, then I know *a priori* that every tokening of “the actual world” refers to the world which corresponds to the one existing cosmos. In fact, human beings do not develop indexical terms for entities that they think there is only one of. That is why “the Earth” was not an indexical term even when people thought that the Earth was the only planetary-type object (in the modern sense of “planetary”, of course). And should we find out that the Earth in fact *is* the only inhabited planet, we *still* could not trivially say that “the Earth” is an indexical term synonymous with “the planet we live on.”

[65]

Skyrms (1976) criticizes this kind of cost-benefit argument for EMR. Possible worlds are physical entities on Lewis’s account are physical entities. Thus only what *physical* theories require should count as evidence in favor of them (p. 326). However, physical theories have counterfactual implications, and thus presuppose modality, and if modality requires EMR, then physical theories require EMR. That something is water entails certain claims about how it would behave in counterfactual situations, and thus the theory that there is water has modal force. It might be objected that a distinction here has to be made between the statement of a law of nature, e.g., that all ravens are in fact black, and a statement that the law of nature is a law of nature, e.g., that it is a law of nature that all ravens are black. Scientific theories only state the laws of nature but do not state that these *are* laws of nature, while it is only the latter kind of statement that has modal force. But this kind of a view forgets that science as such not only seeks to make predictions but seeks to make explanations. And in order to make explanations it must not only assert the laws of nature but assert that these *are* laws of nature. That all ravens are black and Smitty is a raven would not be a scientific explanation if it were not a law of nature that all ravens are black. I take it, however, that a scientific explanation implicitly carries within itself the claim that the generalizations in it are nomic, and this claim is something that science has made. But the claim of nomicity of course has modal consequences.

[66]

Cf. Adams (1987), chapter 4.

[67]

However one explicates this notion. I myself hold that *all* lives of persons are of logical necessity worth living. Someone who does not share this conviction may wish to explicate the notion of “being worth living” in terms of a favorable balance of good things over bad things in life.

[68]

This issue has come up in Canada not long ago.

[69]

The rapist nonetheless is causally responsible for certain unhappinesses in her life, and this causal responsibility should be sufficient for liability.

[70]

See Section 3 of my “*De re* modality, Rawls and abortion” for a problem for Rawlsian ethics that is not removed by this or any other reasonable adjudication of what is to be veiled.

[71]

If y is a counterpart of x , and z is a counterpart of y , then x is an entity that could have been y , and y is an entity that could have been z , so that x is an entity that could have been an entity that is z , and hence by the *de re* implication above, x could have been z , so that z is a counterpart of x .

[72]

Lewis (1986a), p. 248.

[73]

Ibid., p. 249.

[74]

It does not rule out “The aggregate of Plato and his wisdom if Plato is wise, and Socrates otherwise.”

[75]

Note that it is difficult to see how a term that refers in *this* way could fail to be a rigid designator. The counterexample in footnote 74 is no longer available since the term considered there referred in some worlds to *two* entities: Plato and his wisdom. Nor could one just use “Plato if Plato is wise and a sparrow otherwise” as a counterexample, since in worlds where Plato is wise, the truthmaker of the claim that “Plato if Plato is wise and Socrates otherwise” involves not just Plato but also his wisdom.

[76]

One may be tempted in light of this to modify (21) to merely affirm that the existence of x entails the existence of no other entities than x . But that will not do. “Socrates” is certainly an intrinsic referrer, but the existence of Socrates entails the existence of many entities distinct from him, such as Socrates’ wisdom-or-ignorance or an instantiation of a certain genetic sequence. But these entities are not involved in the *truthmaker* of the claim that “Socrates exists”, even if their existence is entailed by this claim.

[77]

Of course, there is a sense of the word “actual” in English which carries connotations of presentness, and this differs from existence, but I shall scrupulously strive to avoid using “actual” in that sense.

[78]

Lewis (1986a), p. 99.

[79]

Lewis (1986a), pp. 99–100.

[80]

Gale (1996), pp. 191–192.

[81]

It seems intuitively obvious that it must be empty, but Lewis’s theory allows that two different things might have the same counterpart in another world. Thus, perhaps, there is some world where t and t_0 have the same counterpart. For instance, that might be true if that world has a time sequence that consists of merely one instant.

[82] See note 21 in Section 4.2 of Part I.

[83] This may however require Lewis to believe in indiscernible worlds, if one thinks it is logically possible that there are two indiscernible copies of our universe.

[84] It is not completely clear what these relations would mean if the ghosts were to be worldmates of material objects in a *relativistic* space-time of course. But perhaps for such cases there could still be a way that a given reference frame in the material space-time also mysteriously picks out simultaneity relations for the non-spatial spirits, in which case my discussion needs to be relativized to one a reference frame. Or maybe it will have to be said that in those worlds in which there are non-spatial spirits, temporal relations must be absolute.

[85] Ultimately, I believe that this is a *per impossibile* consideration. For, it will be essential to my account of modality in Part VI that there be a God who necessarily is the creator of all things that exist. But then all things that exist are causally related to God, and hence causally interconnected. Hence, the argument is ultimately *ad hominem*.

[86] I am, of course, in the end taking a Kantian line according to which the possibility of certain types of causal interactions is what renders temporal relations intelligible.

[87] This objection is described by Lewis (1986a, p. 123n6), according to whom it was given by D. C. Williams in a lecture at the University of Notre Dame in 1974.

[88] Gale notes that the principle must be carefully formulated, perhaps by saying that the description of the action done be temporally pure, in order to rule out counterexamples such as: that an ambulance medic helps the bank-robber who after his getaway has been in a car accident entails that something impermissible was done, namely that a bank was robbed.

[89] Appropriately described; see note 97.

[90] We do have a *prima facie* duty to help people become virtuous. But it is only a *prima facie* duty, subject to many defeaters, and in particular always defeated by a direct moral obligation applying to oneself.

[91] Lewis (1986a, p. 127) writes: “If modal realism makes a problem for anyone, it is for utilitarians”, but neither he nor I sees a disadvantage in this.

[92] It is not clear whether such knowledge is *possible*, so this might be a “*per impossibile*” conditional.

[93] This section owes much to a number of discussions with Richard M. Gale.

[94] I am assuming throughout the ensuing discussion that the worlds in question have only one counterpart of me.

[95] Observe that Lewis’s arguments against interworld causality do not apply to this case, because those arguments (Lewis, 1986a, p. 78) assume that one is talking of causing a proposition of the form “*p* is true at *w*” to be true where “*w*” is a *rigid* reference to a world. It is obvious that there is no interworld causality of this kind, since if a proposition *p* is true at *w* where “*w*” is rigid designator, then it is necessarily the case that (*p* is true at *w*).

[96] Here the assumption that the damage to the body is not permanent or fatal might come in, since one might think that there is a moral prohibition against suicide or maiming oneself whereas an act that violates a moral prohibition is not objectively heroic. However, in any case, if the damage is not permanent, the act appears to be morally permissible.

[97] The assumption that neither you nor the stranger have any ties to the countries served by HH and EHH may be relevant here. Some might think that if your family is among those that HH would save, you might be justified in suggesting the stranger give to HH instead, assuming he does not actually care about the exact number of people saved.

[98] Cf. note 97, above.

[99] Cf. Lewis (1979b).

[100] By “shares our world’s past” I mean, roughly (and I doubt this can be defined more than roughly), that any propositions about what was going on at a past time true in one world are true in the other. There is, of course, a difficulty in defining precisely what one means by “about what was going on at a past time.” One wants to exclude such propositions as “Tipper married the 43rd president of the United States” which, even if true, depends not just on the past but also on the future decisions of the American voters whether to elect Al Gore or not. I assume that the notion of “what was going on at a past time” has an intuitive meaning. Of course, if we were reductive physicalists, this would be easier to define: a continuant world would just be one that has the property of duplicating the spatio-temporal region of our past light cone.

[101] Only the relatively near future is relevant for “everyday” pragmatic considerations, and so I am focusing on the near future as *everyone*, including those who may say that we need only believe in inductive conclusions about the near future, will agree that we need to have reason to think that induction will continue to hold in the near future.

[102]

It might well be that D itself implies that it is a complete description of a world up to t_0 , in which case K will coincide with D . But I leave the issue open.

[103]

Specifically, in the case Lewis (forthcoming) is considering, Sleeping Beauty learns that it is now Monday. But the body of knowledge to which this is added includes the indexical claim that Sleeping Beauty is now awake. I am grateful to James Dreier for pointing out this paper by Lewis.

[104]

A haecceitist will disagree with this assumption, because worlds containing more individuals may have more qualitatively identical worlds corresponding to them. For instance, if a world w contains a finite number n of individuals, and if for each one of these n individuals, there are m possible individuals, where m is infinite, who are indiscernible from that individual, then there are at least m^n possible worlds indiscernible from w , and this number does vary with n . However, the arguments I give below can be adapted to handle such haecceitist concerns.

[105]

The unrestricted principle of recombination can be applied to duplicates of photons that have the property of (individually) dancing the polka, since it is clearly logically possible for one photon to dance the polka. And even if it isn't, then just have the photon do some other funny thing that is logically but not physically possible, and the argument will be unaffected.

[106]

In fact, a variant of Nolan's argument (see Section II.1, above) can be used to show that the collection of all irregular continuant worlds is so much larger than that of all regular continuants that it is too big to be a set, whereas the collection of continuant worlds is a set of some cardinality. This is left as an exercise to the reader.

[107]

Elga's story of Sleeping Beauty is an apparent counterexample to this principle: "Some researchers are going to put you to sleep. During the two days that your sleep will last, they will briefly wake you up either once or twice, depending on the toss of a fair coin (Heads: once; Tails: twice). After each waking, they will put you to back to sleep with a drug that makes you forget that waking. When you are first awakened, to what degree ought you believe that the outcome of the coin toss is Heads?" (Elga, 2000). Elga reasons that you should assign probability $1/3$ to the claim that the coin toss is Heads. After all, if in the epistemic situation in which you find yourself upon awakening you always said "Heads" you would be right exactly $1/3$ of the time. Certainly, at least, you should make bets according to this probability when woken up. However, the Impartiality of Reason would say that since you've learned only the indexical fact that you are awake *now*, you should not judge the probability any differently upon awakening than before the experiment, and obviously before the experiment the right assignment is $1/2$. However, as James Dreier (correspondence, 2001) has pointed out, one can reasonably hold that in fact the epistemic probabilities here diverge from betting odds, and that in fact the right probability to assign upon waking up is $1/2$. Dreier's argument for this is as essentially follows. Suppose that upon waking up you were told that before the whole experiment, you were offered a bet: you could pay \$10 and get \$11 if the coin was in fact Heads. It would then surely be rational for you upon waking up to wish that you had accepted this bet. But its being rational for you to wish that you had accepted this bet is incompatible with the belief that you are rationally assigning a $1/3$ probability to the claim that the outcome was Heads, since then your expected pay-off would be negative. Therefore, in fact, you would be rational to assign not $1/3$ probability but $1/2$ probability to this claim. Dreier's argument here illustrates the interesting fact that if betting odds are to go together with credences, we must be talking of bets which are known to be made only once per class of logically interrelated beliefs. A bet made before the experiment satisfies this proviso, but a bet made upon waking up does not, since the two wake-up times can be presumed to be identical and hence one does not know that the bet will be offered only once. And if one did, then it would be rational to bet as if the probability was indeed $1/2$. Alternately, one might argue that crucial to Elga's analysis of the case is the assignment of non-zero non-one probabilities to propositions such as that it is now Monday (the time of the first wake-up), whereas such propositions are either necessarily false or necessarily true, and for Bayesian consistency must thus be assigned either zero or one probabilities. For a different account, see Lewis (forthcoming).

[108]

An alternate, very controversial objection that I myself am partial to is that the idea of an uncaused event is absurd, and the idea of a non-deterministic non-personal cause is absurd. "Chance" only makes sense, on this objection, as a determination by a free (in the libertarian sense) agent. But the argument above can be reformulated in terms of choices of agents. Suppose Smith who is a very virtuous person and who is freely choosing whether to fulfill some very easy duty, failure to perform which would be a great evil. Since the duty is very easy, Smith is surely more likely than not to perform it. But the same paradox can be made to ensue for EMR as did in the case of the coin flip using this. It is not even necessary to suppose that the "more likely than not" is a numerical measure to make the argument work.

[109]

Anyway, for cardinality reasons we cannot hope that our languages will do the trick. After all, the number of sentences in our languages is countable, so the number of worlds constructed by means of our languages cannot exceed 2^{\aleph_0} , while there are a lot more worlds—at least c^c where c is the cardinality of the continuum, and, as argued in Section 7 of Part III, so many that there is no set of them.

[110]

This is how Lewis takes care of the cardinality objection (see note 118, above).

[111]

Pruss (1999).

[112]

Or, more precisely, equivalence classes of pairs (L,s) where L is a language and s is a sentence under a synonymy relation (since the same sentence might signify different things in different languages; one will, further, have to build into s all contextual

features).

[113]

See Section 7.3 of Part III, above.

[114]

Suppose, as many physicists think, that the ultimate laws of nature in the actual world are indeterministic. Consider, for instance, the law of nature according to which the electron coming out of the Stern-Gerlach experiment has spin-up with probability $\frac{1}{2}$ and spin-down with probability $\frac{1}{2}$. It is physically possible that in fact each time the Stern-Gerlach experiment was performed, the electron had spin-up, though this is highly improbable. Were this the case, then on Humean grounds it would no longer be the case that the electron coming out of the Stern-Gerlach experiment has spin-up with probability $\frac{1}{2}$, and so indeed the laws of nature would be different, in violation of (48). However, the plausibility of (48) would also make this a good argument against the Humean.

[115]

We shall see in Part VI and Part VII that a non-theistic Aristotelian who grounds possibility in the causal powers of prior things might deny this possibility though.

[116]

Mapping, not *function*, to be precise, because we are not dealing with sets but classes.

[117]

Cf. Gale and Pruss (1999).

[118]

This is the purpose for which they are used in *ibid.*, for instance.

[119]

This may be why Aristotle *both* holds essences to be normative *and* holds that they are things that are most definable, and it is the same reason why Plato thinks we can have a knowledge of the Forms, which are obviously normative being the states which other things strive to attain, that we cannot have of other things.

[120]

Rescher (1999), p. 409; italics as in original.

[121]

If one does not think that there is a necessarily existent Deity, then this also gives a complete description of a whole possible world. But even if there is a necessarily existent God, providing he is the God of traditional Christianity within whom there is no contingency whatsoever, the above gives a complete description of a whole possible world. For that the balls possess (1)–(4) entails then that they have the property of being created by that God, since it is necessarily the case that all concrete things other than God are created by God. Hence, arguably, even in that case the description given is implicitly sufficient for the whole world.

[122]

“On the Ultimate Origination of Things”, Ariew and Garber (1989), pp. 151–152.

[123]

I am grateful to Peter van Inwagen for forcing me to clarify this point.

[124]

Note that it is not ruled out that in some privileged cases the thinking might be its own object, as would be in the case of the Aristotelian theory according to which the knower’s understanding becomes the very thing that is thought. (Cf. *Metaphysics* L.9.)

[125]

Could one not maybe even *define* an abstract entity as being a single entity (i.e., not an aggregate of other entities) such that (a) reference to it is *intrinsic* (see Section 4.2.1.b of Part III) and (b) it is logically impossible for it to stand in causal relations? On this definition, the original Platonic Forms may not, however, have been abstract. For the Form of the Good plays an explanatory role in the creation of the universe by influencing the demiurge. If this influence is to be taken as causal, i.e., the demiurge learns of the Good causally, then the Form of the Good is not abstract in the above sense. Note that Plantinga in conversation has said that his Platonic entities *are* causally efficacious—but then Plantinga’s account of propositions is a Leibnizian one.

[126]

Moreover, if having intentionality were to require evolution, then one could make a transcendental synthetic *a priori* argument for the truth of evolution. This means that even before we discovered the physical evidence for evolution, we could have been justified in believing in evolution by arm-chair reasoning. This is surely highly implausible.

[127]

I am grateful to Robert Brandom for this objection.

[128]

See the rebuttal of Craig’s arguments in the introduction to Gale and Pruss, *The Existence of God*.

[129]

See note 43 for one important issue, however.

[130]

And of course Thomas himself does not stop with *et hoc dicimus Deum*: he goes on to give a number of arguments to show that the First Mover, First Cause, etc. has the attributes traditionally attributed to God.

[131]

For instance, that claim would be denied by someone who thought there were no necessary beings but who thought that necessarily something exists, e.g., because of the absurdity of there being nothing. (Cf. the Bergsonian anti-cosmological argument in Gale [1996, p. 273], though that argument, as Gale admitted in conversation, fails because its premise (41) will not be granted by a theist who insists that the truthmaker of a claim that no contingent being exists might involve some non-contingent being.)

[132]

One might think that this solution is indicated in Leibniz's words at the end of his necessary truths argument for the existence of God that through this God "those things which would otherwise be imaginary are realized, to use a barbaric but graphic expression" (Ariew and Garber, 1989, p. 152). However, this charitable interpretation of Leibniz would require Leibniz to have said "are *capable of being* realized", since not all the imaginary things are in fact created. The phrasing, at least on this translation, suggests that Leibniz is not talking of a realization of things that are possible but rather of the being-made-real through indwelling in God's mind of the necessary truths that propositions affirming possibility are.

[133]

See note 16.

[134]

See the discussion of middle knowledge in the introduction to Gale and Pruss (eds.), *The Existence of God*, as well as Adams (1977).

[135]

I am grateful to Richard M. Gale for the idea of introspection as part of the grasp of personal causation.

[136]

Of course there *are* general problems with ostensive definitions. If I define the notion of a causal power by pointing to a number of instances—e.g., my causal power in raising my arm, the giraffe's in bending its neck, the electron's in causing a photon emission in a CRT—it might still be unclear *which* common quality is sought for. After all, perhaps I am pointing out the disjunctive notion of *being one of these ostended instances*. Of course a similar issue will be a problem for any ostensive definitions, whereas it is a fact that we *do* succeed in making ostensive definitions, so the problem cannot be insurmountable. It seems likely that at least a part of a solution will in general be positing that natural kinds or basic ontological categories are sparse. Thus, when I point to a number of instances and make a successful ostensive definition, the quality of, e.g., being one of these instances (or some other contrived quality) will not, as a matter of fact, be a natural kind or a basic ontological category. If one manages to point to a number of instances of causal powers such that the narrowest basic ontological category that they all fall in is *causal power* then one will have accomplished the task of giving an ostensive definition of a causal power.

One might argue that Lewis could avail himself of a similar approach. I might point out a number of instances of things that stand in counterpart relations, and claim that there is only one "basic" or "privileged" relation of which these are cases. This would require a sparse theory of relations. Such a theory would be a theory of relations that cannot be reduced to a theory of set/class-theoretic relations between objects in different possible worlds (i.e., a relation being a subcollection of the collection of ordered pairs of objects existing in the same or different worlds), since the latter theory is far from being a sparse one. Therefore, it is not clear whether there is room in Lewis's ontology for such a theory. And I would suggest that if Lewis's account did have a theory of relations not reducible to these set/class-theoretic ones, then Lewis would have no principled way of objecting to the Platonic theories, since he would be presupposing a very similar ontology—universals (relations) that are not reducible to set/class-theoretic collections.

[137]

Cf. Swinburne (1997).

[138]

Observe the similarity between this argument and the one employed against Roy in Section 3.6 of Part IV.

[139]

See the discussion of the "n-fold Photon" in Section 3 of Part III.

[140]

Essentially, the argument is that any kind of pointing out, even counterfactual pointing out, raises the question as to why of two or more entities indiscernible in the actual world, one but not another should happen to have some property in some counterfactual world. In effect, this argument is applying the Principle of Sufficient Reason (PSR) in that counterfactual world. This needs the PSR to be a basic logical fact with metaphysical necessity, as indeed it is on Rescher's reading of Leibniz (N. Rescher, *Leibniz: An Introduction to His Philosophy*, Aldershot 1993 (reprint of 1979), p. 50). [The footnotes to this long quote are also from Pruss (1998).]

[141]

The only way this could possibly be done would be by saying, "Let me consider a random one of the entities satisfying C"—whatever that might mean (and it is likely that no meaning can be attached to this). But for the purposes of the argument, let us grant this possibility.

[142]

Actually, this argument only needs the logical form of the pre-established harmony, namely that each complete individual concept describes the whole universe. This logical form, as opposed to the thicker form that correlates consciously perceived phenomena with reality, holds in all possible worlds: "tout est lié dans chacun des Mondes possibles" (*La Théodicée*; GP [Gerhardt, 1960–61, Vol.] VI, 107, emphasis as in original)

[143]

If one could truly divide the monad into discernible pieces, then one would have the kind of counterfactual pointing out that is ruled out by the argument in note 149, above.

[144]

"neque enim amplius erit in rebus principium controversiam decidendi" (GP [Gerhardt, 1960–61, Vol.] II, 264).

[145]

Cf. Loux (1979, pp. 59-60).

[146]

This is the view that Plantinga has espoused in discussion.

[147]

Observe that in general in a functional system a functional state will be an equivalence class of physical states. There are

many physical states of a physical Z80 CPU that implement its being in some particular logical state—slightly different amounts of voltage in various parts of the CPU will qualify, for instance.

[148]

There are of course other arguments against functionalism. One that I find plausible is that the number of persons thinking a given thought is well-defined in a way in which the number of functional systems computing a given program is not. To see this, observe that there is no well-defined way of characterizing when there are two systems running the same program and when there is a single doubly redundant system running it.