2013

Metaphysical Dependence and Set Theory

John Wigglesworth

Graduate Center, City University of New York

METAPHYSICAL DEPENDENCE AND SET THEORY

by

JOHN WIGGLESWORTH

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2013

This manuscript has been read and accepted for the
Graduate Faculty in Philosophy in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Richard Mendelsohn

____________________ ____

Date Chair of Examining Committee

Iakovos Vasiliou

____________________ ____

Date Executive Officer

Graham Priest

____

Arnold Koslow

____

Melvin Fitting

____

Kit Fine

____

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

METAPHYSICAL DEPENDENCE AND SET THEORY

by

JOHN WIGGLESWORTH

Advisor: Professor Graham Priest

In this dissertation, I articulate and defend a counterfactual analysis of metaphysical dependence. It is natural to think that one thing x depends on another thing y iff had y not existed, then x wouldn't have existed either. But counterfactual analyses of metaphysical dependence are often rejected in the current literature. They are rejected because straightforward counterfactual analyses fail to accurately capture dependence relations between objects that exist necessarily, like mathematical objects. For example, it is taken as given that sets metaphysically depend on their members, while members do not metaphysically depend on the sets they belong to. The set $\{\emptyset\}$ metaphysically depends on $\emptyset$, while $\emptyset$ does not metaphysically depend on $\{\emptyset\}$. The dependence is asymmetric. But if counterfactuals are given a possible worlds analysis, as is standard, then the counterfactual approach to dependence will yield a symmetric dependence relation between these two sets. Because the counterfactual analysis fails to accurately capture dependence relations between sets and their members, most reject this approach to metaphysical dependence.

To generate the desired asymmetry, I argue that we should introduce impossible worlds into the framework for evaluating counterfactuals. I review independent reasons for admitting impossible worlds alongside possible worlds. Once we have impossible worlds at our disposal, we can consider worlds where, e.g., the empty set does not exist. I argue that in the worlds that are *ceteris paribus* like the actual world, where $\emptyset$ does not exist, $\{\emptyset\}$ does not exist either. And so, according to the counterfactual analysis of dependence, $\{\emptyset\}$ metaphysically depends on $\emptyset$, as desired. Conversely, however, there is no reason to think that every world that is *ceteris paribus* like the actual world, where $\{\emptyset\}$ does not exist, is such that $\emptyset$ does not exist either. And so $\emptyset$ does not metaphysically depend on $\{\emptyset\}$. After applying this extended counterfactual analysis to several set-theoretic cases, I show that it can be applied to account for dependence relations between other mathematical objects as well. I conclude by defending the counterfactual analysis, extended with impossible worlds, against several objections.

Acknowledgements

I am eternally grateful to my wife, Victoria. Her patience, encouragement, and understanding made the completion of this project possible. I also express my deepest thanks to Graham Priest, who challenged me to think in new ways about many ideas in this thesis and beyond. His guidance has made me a better philosopher. But I could not have accomplished anything without my parents. They have always believed in me, and they taught me to believe in myself.

There are many others who deserve thanks. My committee members, Arnie Koslow and Mel Fitting, devoted much of their valuable time to reading many drafts. This thesis would not be what it is without their insightful comments. I am grateful to Daniel Nolan and Jonathan Schaffer for their challenges to the main arguments in this thesis. I appreciate the constructive feedback and critique from audiences in New York, Melbourne, Bristol, and Nottingham, where parts of this work have been presented. And I must also thank Ricki Bliss for keeping me interested in metaphysics when I had my doubts. Lastly, I am forever indebted to Gary Matthews and Kris McDaniel, my first philosophy teachers, who inspired me to follow this path.

Contents

List of Figures

CHAPTER 1

Introduction

Dependence has become something of a “hot topic” in metaphysics. Some see it as a central concept in the study of reality. A concept as important, or arguably more important than the concept of existence. For once we know what exists, it is how those things relate to one another that tells us what reality is like. And it is thought that one of the fundamental ways in which things relate to one another is in virtue of metaphysical dependence. Metaphysical dependence gives structure to the world.

One explanation as to why metaphysical dependence is so important is its ubiquity. If we look, we can find metaphysical dependence everywhere. Properties depend on objects; wholes depend on parts; holes depend on hosts; moral facts depend on non-moral facts; modal facts depend on non-modal facts; legal facts depend on non-legal facts, the world depends on mind; everything depends on God. God depends on Herself. And so on. I do not claim that these dependence relations actually hold. They are examples of dependence claims that can be made and have been made.

One explanation as to why metaphysical dependence is ubiquitous is that it may just be an umbrella term, used to capture many different concepts. There is a plurality of relations, each of which could plausibly be labeled a dependence relation: supervenience, composition, constitution, determination, causation, truth-making, grounding, priority, holding “in virtue of”, explanation, entailment, and arguably set-membership.

Of course, these are not all terms for the same relation. Rather, it may be that the phrase ‘metaphysical dependence’ is ambiguous between them, or between the elements of some subset, or some superset, of them. The goal of this project is to single out one plausible understanding of metaphysical dependence, and show that under this particular conception of dependence, sets metaphysically depend on their members. And not vice versa. At least, not usually. Under this conception, in general, things that are members of sets do not metaphysically depend on the sets that they are members of. When it comes to sets and their members, metaphysical dependence is in most cases asymmetric.

The claim that sets metaphysically depend on their members may strike you as odd. It is widely accepted that some concrete objects metaphysically depend on others. But why think the same holds for sets? Sets are mathematical objects. And mathematical objects are said to exist necessarily, if they exist at all. They could not have failed to exist. In what sense, then, could they depend on anything else? Their existence is, in effect, guaranteed. Nothing else is needed. Certainly, additional complications arise when one tries to articulate dependence relations between objects that exist necessarily. But the fact that it is complicated or odd does not justify ignoring the thought. It is the thought that we wish to explore. And what we show is that, at least in the set-theoretic case, the claim that some mathematical objects metaphysically depend on others is true. And we give a framework within which we can understand these dependence claims.

That being said, on a general level, we wish to remain neutral with respect to the ontological status of mathematical objects. The claim we defend is to some degree a hypothetical one: If mathematical objects exist, then they enter into certain dependence relations. For example, if sets exist, then they depend on their members. We take as an assumption, then, that purely mathematical objects like numbers and sets exist, and that they exist necessarily. Indeed, it is for the most part standard to think that if mathematical objects exist, then they exist necessarily. But we do not try to defend these claims here; we simply try to articulate a conception of metaphysical dependence that would hold between certain mathematical objects, given that they exist.

On these assumptions, then, we argue that there is a conception of metaphysical dependence that holds between sets and their members. We do not argue that this is the *only* conception, or the *one and true* conception of metaphysical dependence. We admit that there are many conceptions of dependence. For instance, the dependence that holds between a set and its members may differ from the dependence that holds between a whole and its parts. In fact we think these are two very different notions of dependence. We only claim that the conception of dependence articulated here is a plausible one, and that it is the conception that captures the dependence that holds between sets and their members.

One immediate question is: Why do this? Why go to great lengths to articulate a notion of metaphysical dependence that holds between sets and their members? I offer three reasons. One reason, as noted above, is to make sense of claims of metaphysical dependence in the context of set theory, claims of dependence that people actually make. In this respect we have to be a little bit careful, because the phrase "set theory" itself may be ambiguous, or at least imprecise. Some may take talk of set theory at a general or informal level. At this level, we usually think of sets of physical objects, like lions and tigers. Here we might be thinking of sets as extensions of predicates, or perhaps as being identical with properties. We may also think of sets of possible worlds, in which case we may be thinking of sets as being identical with propositions. The idea is that sets are simply collections of objects.

There is also the more mathematical Set Theory, where the sets to be examined are usually taken to be pure mathematical objects. Here we are looking at sets of numbers perhaps, or more likely sets of other sets. On the mathematical approach, Set Theory is understood as a collection of axioms that govern the membership relation, and which describe a universe of mathematical objects called sets. The mathematical theory of sets comprises those axioms and what logically follows from them. This theory tells us what is true in the universe of sets. One might even take this universe of sets to serve as a foundation for the entirety of mathematics. We argue that under both the informal and formal approaches to set theory, one can find claims to the effect that sets metaphysically depend on their members. It is claims like these that we wish to make sense of.

A second reason to articulate a notion of set-theoretic dependence is that the set-theoretic cases often cause problems for straightforward, plausible, intuitive analyses of metaphysical dependence. We will look at two of these analyses throughout the course of the dissertation, called the modal analysis and the counterfactual analysis. The basic naive intuition behind both of these kinds of analyses is that one thing depends on the other when you can't have the one without the other. However, each analysis tries to capture this intuition in a slightly different way. Both of these analyses can easily account for dependence relations that hold between concrete objects. But they struggle to capture dependence relations between abstract objects, like pure sets, or even between concrete objects and the impure sets that contain them, where these sets might be seen as partly concrete and partly abstract. What makes things more problematic is that these analyses actually do work for some set-theoretic cases. That the modal and counterfactual analyses work for some set-theoretic cases suggests that we are not dealing with two different conceptions of metaphysical dependence, one conception that applies to concrete objects, and a different conception that applies to sets. Because they work in so many cases, there is a motivation to preserve something like the modal or counterfactual analyses, and figure out how to make them work for the problematic set-theoretic cases, thus giving us a notion of dependence that unifies the concrete and abstract cases. That is exactly what we intend to do.

Given that we are trying to apply modal and counterfactual conceptions of dependence to sets, some of which are purely mathematical objects, a third motivation emerges. Purely mathematical objects are often thought to exist necessarily, if they exist at all. The necessary existence of mathematical objects makes it difficult to apply modal concepts, like necessity, possibility, and impossibility, to these objects. In effect, these modal concepts are trivialized in the context of necessarily existing objects. Something is necessary with respect to these objects if and only if it is true. Something is impossible if and only if it is false. Nothing is merely possible or contingent with respect to these objects. Modal notions appear to lose their force, and perhaps even their meaning, when applied to mathematical objects. We cannot say anything very interesting about what is possible and what is necessary with respect to mathematical objects. By articulating a modal or counterfactual understanding of metaphysical dependence that can accurately, and non-trivially, apply to pure mathematical sets, we give sense to a fuller notion of necessity, possibility, and impossibility, which can then apply in a fruitful way to mathematical objects.

Given these motivations, we see this dissertation as making a contribution to several areas of philosophy, including metaphysics, the philosophy of mathematics, and the emerging area of the philosophy of set theory. Allow me to briefly describe the dialectic of the dissertation, following which I give a more detailed outline of each chapter.

1. The Plan

The dialectic is as follows. We start with the idea of metaphysical dependence. Recognizing that there are many different notions of dependence, we hone in on one: the idea that one thing depends on another if, and only if, you cannot have the one without the other. In this characterization, the word "cannot" suggests a modal component; the word "have" suggests an existential component. And so we have a notion of metaphysical dependence in terms of necessary conditions on existence: one thing depends on another if it is necessary, for the one thing to exist, that the other exist as well. Call this the modal/existential conception of dependence.

We argue that this modal/existential approach to dependence has historical precedent. We give textual evidence that celebrated philosophers in the western tradition appeal to precisely this notion of metaphysical dependence. In essence, this is the first stage of the dialectic. The second stage argues that there are reasons for thinking that sets metaphysically depend on their members. And in most cases, this dependence is asymmetric. Furthermore, the sense in which sets depend on their members is captured precisely by the modal/existential conception. What is problematic is that contemporary treatments of modality, in terms of possible worlds, do not allow us to accurately capture this dependence in an obvious way.

Moving to the third stage of the dialectic, we take two steps to resolve this problem. The first is a shift from a straightforward modal analysis of metaphysical dependence to a more nuanced counterfactual account of metaphysical dependence. The second is to extend our treatment of modality to include impossible worlds. To justify this extension, we argue in favor of the existence of impossible worlds.

The argument in favor of impossible worlds has two components. The first shows that adding impossible worlds to mainstream theories of possible worlds does not cause any problems. The second presents positive arguments in favor of impossible worlds. After arguing for the existence of impossible worlds, the fourth stage of the dialectic presents the counterfactual analysis of metaphysical dependence, extended with impossible worlds. We apply the analysis to a variety of set-theoretic cases. We show that the extended counterfactual analysis accurately captures the dependence relations between sets and their members. We also argue that this analysis can be extended to capture dependence relations between other mathematical objects. Having presented the counterfactual analysis in detail, showing that it accurately captures dependence relations where other accounts fail, we then consider and reply to objections.

That is the basic plan. We now describe this plan in more detail, summarizing each chapter of the dissertation. Chapter 2 begins with a selective history of metaphysical dependence. The purpose of this historical survey is to show that the modal/existential conception of dependence is prevalent in the writings of many important figures in the history of western philosophy. Of course, we cannot be comprehensive. So we pick three major figures, examining selections from Aristotle, Descartes, and Husserl. This selection temporally spans almost the entirety of western philosophy, from the classical period to the beginning of the twentieth century. And while all three figures are western philosophers, their approaches to philosophy are quite different. We argue that all three have a modal/existential notion of dependence. After this selective historical tour of the concept of metaphysical dependence, Chapter 2 then looks in more detail at the present discussion of dependence, arguing that the modal/existential conception of dependence is one of many different conceptions of dependence in the current literature. By taking this historical approach, from Aristotle to the present, we show that the modal/existential conception of dependence is a perfectly legitimate understanding of metaphysical dependence. Chapter 2 concludes with further discussion on the particular conception of metaphysical dependence that we focus on in this project. We look at the structural properties that metaphysical dependence is assumed to have, and whether or not the modal/existential conception has these properties. We also compare the modal/existential conception of dependence with other conceptions of metaphysical dependence.

Chapter 3 is divided into two parts. The first part explores reasons why one might defend the claim that sets metaphysically depend on their members. We look at two arguments. The first shows that the fact that sets depend on their members follows directly from our conception of set. The second shows that the claim that sets metaphysically depend on their members is implicit in the view, common in the foundations of mathematics, that the cumulative hierarchy of sets exhausts the set-theoretic universe. According to this view, the cumulative hierarchy, as captured by the iterative conception of set, is *the* universe of sets.

The second part of Chapter 3 examines several ways to formulate a notion of metaphysical dependence in modal terms. We show that a standard, straightforward modal analysis, as well as a more nuanced counterfactual analysis of metaphysical dependence both fail to accurately capture the relation of dependence that holds between sets and their members. We do not, as some have done, take this failure to imply that sets do not depend on their members. Rather, we infer from these results that these analyses do not faithfully represent the modal/existential conception of dependence. The primary goal of this project is to articulate a notion of metaphysical dependence that accurately characterizes the dependence between sets and their members. Chapter 3 ends by taking the first steps to extend these analyses to try to avoid these failures. What does the most work in this development is the appeal to impossible worlds.

Chapters 4 and 5 explore impossible worlds in some depth. Chapter 4 begins with a general discussion of how to characterize the difference between possible and impossible worlds. We then show that one can add impossible worlds to mainstream theories of possible worlds without too many problems. We look at both realist and anti-realist theories of worlds. On the realist side, we examine Lewis's extreme concrete realism, linguistic ersatzism and a version of ersatzism based on states of affairs, combinatorialism, and even a hybrid realist theory that combines Lewisian possible worlds with set-theoretic constructions. We also look at anti-realist theories of worlds, like fictionalism and noneism. Impossible worlds can easily be added to all of these theories, both realist and anti-realist. Furthermore, if you accept many of the standard arguments in favor of the existence of possible worlds, there are equally plausible analogous arguments in favor of impossible worlds that you should probably accept as well. Chapter 5 looks at several positive arguments in favor of the existence of impossible worlds. Many of these arguments are pragmatic, appealing to their use in developing other logical and philosophical theories. Impossible worlds are essential to giving the truth conditions for a relevant conditional, as well as for giving a semantics for the logics of knowledge and belief. And they can be used to give straightforward, plausible philosophical accounts of propositions (as sets of worlds), properties (as extensions across worlds), and counterfactual conditionals with impossible antecedents. More generally, we suggest that many philosophers make implicit appeal to impossible worlds simply in virtue of the kinds of philosophical debate that they engage in. Many philosophical positions are believed to be necessarily true, if true at all. It follows that, when two philosophers genuinely disagree, at least one of them is arguing from a position that is necessarily false, i.e., impossible. But without impossible worlds, the kind of reasoning that goes on in these situations is trivialized. For this reason, as responsible philosophers, we should all believe in impossible worlds. Chapter 5 also looks at, and rejects, a general worry about impossible worlds.

Chapter 6 takes us back to the concept of metaphysical dependence. Having argued for the existence of impossible worlds, we propose an alternative counterfactual analysis of dependence that deploys these worlds. We show that this analysis accurately captures the dependence relation between sets and their members. According to this conception, sets metaphysically depend on their members, and in most cases members do not metaphysically depend on the sets that they belong to. We show that this analysis succeeds in many cases. It succeeds with both pure mathematical sets and impure sets with urelements. It succeeds with both finite and infinite sets. It can even be applied to non-well-founded sets. We then explore the structural properties of this relation of metaphysical dependence. We concede that it is not guaranteed to have the structural properties (like asymmetry and well-foundedness) that are normally desired of a dependence relation.[1] For this reason, we call the conception of dependence presented here a *minimal* conception of metaphysical dependence. We maintain that it is not obvious that we should assume that metaphysical dependence has these structural properties. In general, we should try not to assume very much at all. Rather, we must *argue* that metaphysical dependence has these properties. Producing these arguments is not the project of this dissertation. We are simply interested in articulating this minimal conception of metaphysical dependence, and showing that it can produce the result that sets depend on their members, while in most cases members do not depend on the sets they belong to. We leave to others the task of constructing arguments to the effect that dependence should have any additional structural properties. But we concede that dependence could have these properties.

[1]We note that even if the dependence relation is not inherently symmetric, it can still be that the dependence between sets and their members is not symmetric. That is, the fact that a relation is not asymmetric does not make it symmetric.

And we show how, if one were so inclined, one could add conditions to the minimal relation of metaphysical dependence to ensure that the dependence relation has those structural properties. We conclude Chapter 6 by investigating the possibility of extending the counterfactual analysis of metaphysical dependence to account for some of the dependence claims made by the mathematical structuralist.

Chapter 7 concludes the dissertation by considering and replying to several objections to the counterfactual analysis of dependence, extended with impossible worlds. The objections are divided into three categories. The first considers general objections to high-level assumptions that we make at various points. These include the assumption that modal notions should be given a worlds analysis, the assumption that counterfactuals should be understood in terms of worlds, and even the assumption that non-actual worlds exist at all. We mention these objections, with very brief replies, mainly to set them aside. We also touch on the fact that counterfactuals are sensitive to context, and we describe our approach to context sensitivity. Objections of the second kind take the form of potential alternatives to our counterfactual analysis of dependence, extended with impossible worlds. We consider two alternatives and give reasons for rejecting them. Objections of the third kind take the form of specific worries about some of the consequences of the counterfactual analysis extended with impossible worlds. We take these to be the strongest objections to the view presented here, and we reply to them. We conclude the dissertation with a general discussion about the worries some may have about impossibility and how to think about impossible situations. To ease these worries, we show that there is a perfectly legitimate way to reason about these situations. The fact that they are impossible should not scare us.

CHAPTER 2

Metaphysical Dependence, Then and Now

In this chapter, we argue that the modal/existential conception of metaphysical dependence is a perfectly legitimate understanding of dependence. To do this, we first propose what we take to be a plausible interpretation of the modal/existential conception. We then show that this conception plays a role in the writings of several influential figures in the history of western philosophy. Specifically, we look at the writings of Aristotle, Descartes, and Husserl. Following this historical treatment, we discuss the place of metaphysical dependence in the twentieth and twenty first centuries. While in the past, the notion of dependence was used quite a bit, it is not until recently that the relation, or better the relations, of dependence have become objects of study in their own right.

We compare different conceptions of metaphysical dependence in the latter half of this chapter. At the moment, it would be good to focus on the particular conception of dependence that we are interested in. This conception of dependence tries to capture some kind of connection between the existence of one thing and the existence of another. Understood this way, it is a connection between objects. Later, however, we will see that it can be understood as a relation between states of affairs. As a connection between objects, however, it can be expressed, perhaps naively, with the following phrase: one object metaphysically depends on another if, and only if, you can't have the one without the other.

Understood in this naive way, metaphysical dependence involves notions of necessity and existence. For this reason we refer to this conception as the modal/existential conception of metaphysical dependence. It yields what we may call the naive analysis of metaphysical dependence.

NAIVE ANALYSIS. *x metaphysically depends on y if, and only if, it is impossible that x exists and y does not.*

Though we call this an "analysis" of metaphysical dependence, we take it primarily to give necessary and sufficient conditions for when one thing metaphysically depends on another, where the relevant conception of metaphysical dependence is the modal/existential conception.

According to the naive analysis, claims of metaphysical dependence are equivalent to claims about what is impossible. Alternatively, they may be formulated as claims about what is necessary, as something is impossible if and only if, necessarily, it does not occur. The twentieth century has witnessed major advances in the logical and philosophical study of modal concepts like possibility and necessity. It is standard to interpret modal claims to be claims about possible worlds. These include claims about what is possible, impossible, or necessary, claims about what could occur, what couldn't occur, or what must occur. If something could occur, if it is possible, then there is some possible world where it does occur. If something couldn't occur, if it is not possible, then there is no possible world where it does occur. If something must occur, if it is necessary, then it occurs in every possible world.

Under a possible-worlds interpretation, to say that you can't have one thing without the other means there is no possible world where the one exists but the other does not. Combining the naive analysis with a possible-worlds interpretation of modality gives us the modal analysis of metaphysical dependence.

MODAL ANALYSIS. *x metaphysically depends on y if, and only if, there is no possible world w such that x exists in w and y does not exist in w.*

The modal analysis is a more precise statement of our original naive intuition. The modal analysis proposes a determinate condition for dependence, while the naive conception just tries to capture our vague intuitions. In some sense they say the same thing. It's just that the modal analysis does it better. Despite differences in precision, they should stand and fall together.

But they don't. Somewhere along the path from the naive intuition to the modal analysis, something gets lost. For, according to the naive intuition, sets depend on their members because you can't have the set without each of the members. But members usually do not depend on the sets they belong to, because you could have the members without the set. The modal analysis does not preserve these intuitions.

The modal analysis works perfectly well for concrete objects. Consider a statue and the lump of clay that constitutes the statue. It seems natural to say that the statue metaphysically depends on the lump of clay, and not vice versa. The modal analysis confirms this dependence claim. Consider the following modal claims.

- There is no possible world w such that the statue exists at w and the lump does not exist at w.
- There is no possible world w such that the lump exists at w and the statue does not exist at w.

Presumably the first claim is true and the second claim is false. Certainly the second claim is false, as you could squash the clay, destroying the statue without destroying the lump. One could try to argue that the first claim is false on the basis that the same statue could have been made with a different lump of clay. But that isn't quite right. Certainly a very similar statue could have been made with a different lump of clay. But it would not be the same statue. For the second statue could have been made, using the second lump of clay, even if the original statue existed. In this situation, we would have two distinct statues. I take this to be one of the main points of those who claim that the statue and the lump have different modal properties. So the first claim is true. Effectively, you can't have the statue without the lump, but you can have the lump without the statue. The modal analysis, therefore, preserves our intuitions that the statue depends on the lump, but not vice versa.

But the modal analysis does not accurately capture dependence relations between sets and their members. It fails to capture these dependence relations because sets and their members exist in all of the same possible worlds. More precisely, for any set $S = \{s_1, s_2, \ldots\}$, necessarily, S exists iff all of $s_1, s_2, \ldots$ exist. On the modal analysis, necessary co-existence of a set with its member entails that sets depend on their members *and* that members depend on their sets. In these cases, according to the modal analysis, the dependence is symmetric.

It may be that there are cases of symmetric dependence, but the set-theoretic cases are not obviously of this kind. Which is why it seems natural to say that you can't have a set without its members, though you could have the members without the set. So the naive intuition and the modal analysis come apart in the set-theoretic cases.

I suppose one could bite the bullet and accept the symmetric dependence of a set on its members. Sets are not your ordinary everyday concrete objects. They are abstract. And abstract objects are not always well behaved. A case might be made for the symmetric dependence between a set and its members. I do not endorse this symmetry in the straightforward set-theoretic case.[1] But it is an option one could pursue.

The modal analysis has a bit more to answer for though. There are some sets that are taken to exist necessarily, like the empty set $\emptyset$ (the set with no members), or its singleton $\{\emptyset\}$ (the set whose sole member is the empty set). These sets are pure sets — sets whose members are all sets, and their members are all sets, and their members are all sets, and so on. These are the sets that mathematics deals with, and as they are completely abstract objects, they are taken to exist necessarily (if they exist at all). Necessarily existing objects pose a particular problem, because the modal analysis will entail that any arbitrary object metaphysically depends on a necessary existent. If, for example, the empty set exists necessarily, then it couldn't have failed to exist. So, for any object you choose, whether its you or me or Socrates, there is no possible world where that object exists and the empty set doesn't. For there is no possible world where the empty set does not exist! When necessary existents are involved, the modal analysis trivializes the dependence relation. Every object depends on the empty set. In fact, every object depends on any necessary existent.

We explore the problems of the modal analysis in more detail in the next chapter. Despite its problems, the modal analysis is the straightforward way to make a very natural understanding of metaphysical dependence philosophically precise. And it is an understanding of dependence that has historical precedence. We argue that something like the modal/existential conception of metaphysical dependence can be found in the writings of several influential philosophers throughout the history of western philosophy. We look at three philosophers, representing three different philosophical periods. The philosophers we examine are Aristotle, Descartes, and Husserl. Following the historical discussion, we look at metaphysical dependence in the twentieth and twenty first centuries.

[1]Though see Chapter 6, Section 2.6 on non-well-founded sets.

1. Then: a selective history of metaphysical dependence

We begin with Aristotle. Aristotle has a general metaphysical picture that divides the world into different ontological categories, different categories of being. There are the categories of substance, of quality, of quantity, of place, among others. In total there are ten categories, and they combine in various ways to produce facts. For example, the substance *Socrates* combines with the place *in the market* to form the metaphysical fact that Socrates is in the market.

Of these ten categories, one of them — substance — plays a fundamental role. Within this category, there are two kinds of substance: individual substances and universal substances. Individual substances include particular things like you and me as individual humans; universal substances include more general things like humanity. In Aristotle's language, individual substances are primary. They lay the foundation upon which the rest of reality is built. "Thus all the other things are either said of the primary substances as subjects or in them as subjects. So if the primary substances did not exist it would be impossible for any of the other things to exist" (*Categories*, 2b4-7). Primary substances, the individual substances like you and me, hold up the rest of the world. Without them, there would be no other categories. There would be no quality or quantity. There would be no location. There would be no humanity. For something to be primary is for it to be such that nothing could exist without it.

Aristotle clearly has the conception of a necessary connection between the existence of things. Nothing could exist without primary substances. In the *Metaphysics* Aristotle connects the idea of primary substance with the idea of independence. For Aristotle, to be primary is to be independent, in that things that are primary can exist without anything else. "Now there are several senses in which a thing is said to be primary; but substance is primary in every sense — in formula, in order of knowledge, in time. For of the other categories none can exist independently, but only substance." (*Metaphysics*, 1028a32-4). In other words, primary substances are the only things that are independent. That is, if something is independent, then it is a primary substance. The other categories are dependent. The other categories depend on primary substances for their existence. They cannot exist without primary substance. Indeed, the whole of reality cannot exist without primary substance.

In his discussion of primary substance, Aristotle also appeals to a notion of priority. "Some things then are called prior and posterior ... in respect of nature and substance, i.e. those which can be without other things, while the others cannot be without *them*, — a distinction which Plato used. If we consider the various senses of 'being', firstly the subject is prior (so that substance is prior) (*Metaphysics*, 1019a1-5). I interpret this passage as confirming that (primary) substance is prior with respect to being. And I understand "being" to be existence. Accordingly, we have that if something is prior, then it can be without other things. And if something is primary substance, then it is prior with respect to being or existence. Combining these, we have that if something is primary substance, then it can be without other things. Putting this together with the results, from the previous paragraph, that if something is independent, then it is primary substance, we have by transitivity that if something is independent, then it can be without other things.

This connection between independence and "being without other things" approaches our naive intuitions about a modal/existential conception of metaphysical dependence. It is by no means a perfect match. For one thing it only expresses a necessary condition for independence. But we have shown that Aristotle endorses some connection between the dependence of one thing on another and the question of whether one thing can exist without the other.

For Aristotle, that substance is independent is practically definitional. Substance just is what is independent and what everything else depends on. But this is no innocent, nominal definition. This is a *real* definition, with serious metaphysical consequences. Aristotle's definition of substance plays a crucial role in his general metaphysical view, because it has consequences that contradict the views of Aristotle's teacher, Plato. According to Plato, universals (i.e., the Forms) are independent; they can exist independently from anything else. But for Aristotle substance is the only thing that is independent, and everything else depends on substance. So, if the Forms exist at all, they must depend on substance, thus contradicting Plato's claim that they are independent.[2]

Well after the ancient period, metaphysical dependence continues to play an important role in the study of metaphysics.[3] Consider Descartes' dualist picture of the world. The dualist picture is

---

[2]See Shields (2007), p. 156.

[3]In the two millennia between Aristotle and Descartes, hints of the modal/existential conception of dependence can be found in the writings of other influential western philosophers. Consider, e.g., Aquinas (1968): "Matter, then, cannot

a metaphysical picture, motivated by Descartes' belief that his mind is independent from his body. In the second *Meditation*, Descartes develops a distinction between the body and what he calls the 'I', which he eventually identifies with the mind, the thinking substance. As long as Descartes is thinking, this 'I' must exist. But it is entirely possible for the 'I' to exist without the body. "Am I not so bound up with a body and with senses that I cannot exist without them?" (1985b, p. 16).[4] The rhetorical nature of Descartes' question implies that he believes the mind can exist without the body. He takes this fact about what is possible to imply that the mind does not depend on the body for its existence. This implication suggests that Descartes is working with a general principle: If *x* can exist without *y*, then *x* does not depend on *y*. This general principle amounts to one half of the naive conception of dependence.

We can see that Descartes endorses the other half of the naive conception by looking at his earlier *Discourse on Method*, Part Four, where the conceptual connection between dependence and possibility is more fully expressed.

> From this I knew I was a substance whose whole essence or nature is simply to think, and which does not require any place, or depend on any material thing, in order to exist. Accordingly this 'I' — that is, the soul by which I am what I am — is entirely distinct from the body, and indeed is easier to know than the body, and would not fail to be whatever it is, even if the body did not exist (1985a, p. 127).

Starting with the claim that Descartes' 'I' does not depend on any body, it follows that the 'I' can exist without any body. This suggests another general principle, which is the converse of what we saw from the *Meditations*: if *x* does not depend on *y*, then *x* can exist without *y*. Putting these two general principles together, we have the naive intuition: *x* depends on *y* iff *x* can't exist without *y*.

At the end of the second *Meditation*, Descartes confirms that this 'I' is the mind, the thinking substance. Descartes position is that the mind is independent of the body. Most philosophers today would disagree with this position, arguing that there is some kind of dependence between the mind

---

exist without some form, but there can be form without matter: form as such does not depend on matter" (1968, p. 53).

[4]The Veitch translation of 1901 uses the word "dependent" in place of "bound up with". "Am I so dependent on the body and the senses that without these I cannot exist?" (2008, p. 79). I quote the Cottingham translation in the main text, as it is the standard translation. In this context, being "bound up" with something does seem to imply some kind of dependence.

and the body. This dependence is often stated in the form of a supervenience thesis, where the mental (or mental properties) supervenes on the physical (or physical properties). We consider the connection between dependence and supervenience in Section 2 of this chapter.

Explicit discussion of the dependence relation in more contemporary philosophy begins with Husserl. His *Logical Investigations* (1970) is the first comprehensive investigation of metaphysical dependence produced in the 20th century. While Husserl's work is mostly ignored in the study of western analytic metaphysics, we are lucky to have available some commentary from prominent western analytic metaphysicists. Thanks to the efforts of Kit Fine (1995b) and Peter Simons (1982), we have a better understanding of Husserl's extensive treatment of topics that are particularly relevant to current analytic metaphysics, including metaphysical dependence and mereology. As we see in Husserl's third *Investigation*, the concepts of part, whole, and dependence serve as the basis for Husserl's ontology. As such, they fortify his general phenomenological approach. The very first sentence of the third *Investigation* reads, "The difference between 'abstract' and 'concrete' contents, which is plainly the same as Stumpf's distinction between *dependent* (non-independent) and *independent* contents, is most important for all phenomenological investigations" (p. 3).[5] These basic ontological concepts, and Husserl's views on them, provide a foundation for his entire philosophical system and method.

Broadly construed, phenomenology is the study of the way in which things appear to us in conscious experience. This includes the representation of a structure in which things appear to be related to each other in various ways. For Husserl, the most fundamental of these relations are the part-whole relation and the relation of dependence. The third *Investigation* takes up the task of exploring these relations. In developing his theory of these basic relations, Husserl takes an axiomatic approach (§14, pp. 25 – 7). But underlying this approach seems to be a basic understanding of what it means for one thing to depend on another. This basic understanding is the naive analysis of dependence.

Given his phenomenological approach, Husserl discusses metaphysical dependence between mental contents. Mental contents belong to conscious experiences (acts of thinking, perceiving,

[5]Unless otherwise noted, all emphases are in the original passages.

desiring, etc.). Some mental contents depend on others; some mental contents do not, and so are independent. In his discussion of dependent and independent contents, Husserl explicitly endorses something very much like the naive analysis of dependence.

> Occasionally one hears the difference between independent and non-independent contents expressed in the attractive formula: Independent contents (part-contents) could be presented by themselves, non-independent contents only noticed by themselves, not presented by themselves. ... What [the expression 'presented by itself'] plainly means is that it is possible to present the object as something *existing by itself*, as *independently there* in the face of all other objects. A thing or piece of a thing can be presented by itself — this means it would be what it is even if everything outside it were annihilated. If we form a presentation of it, we are not necessarily referred to something else, included in which, or attached to which, or associated with which, it has being, or on whose mercy it depends, as it were, for its existence. We can imagine it as existing by itself alone, and beyond it nothing else (§6, pp. 10 – 11).

An independent content is a content that could exist in our conscious experience without any other mental contents. A non-independent (in other words, dependent) content is a content that could not exist without other mental contents. The existence of a dependent content requires the existence of the other contents on which it depends.

Though Husserl's understanding of dependence is phrased in terms of contents, he endorses the same distinction between independent and non-independent objects, with a parallel understanding of what this distinction means.

> We need only say 'object' and 'partial object', instead of 'content' and 'partial content' — the term 'content' we regard as the narrower term, the one restricted to the sphere of phenomenology — to achieve an *objective distinction* freed from all relation to interpretative acts and to any phenomenological content that might be interpreted. *No reference back to consciousness is therefore needed, no reference to differences in the 'mode of presentation'*, to determine the difference between 'abstract' and 'concrete' which is here in question (§5, p. 10).

For our purposes, very little turns on the question of whether dependence holds between objects or contents, or both. In both cases, dependence is a relation between kinds of things that exist. For this reason, Husserl appears to be working with something like the naive analysis, as he articulates metaphysical dependence in terms of what is possible or not with respect to the existence of these things.

There may be a worry that, given Husserl's emphasis on mental content, it is an open question as to exactly what notion of possibility is meant when he claims that an independent content may exist or be presented by itself. It is well recognized that our ability to present a content by itself does not entail that it is possible for the object of that content (if there is one) to exist by itself. For Husserl, this is not a worry. "What cannot be thought, cannot be, what cannot be, cannot be thought. ... Wherever therefore the word 'can' occurs in conjunction with the pregnant use of 'think', there is a reference, not to a subjective necessity, i.e. to the subjective incapacity-to-represent-things-otherwise, but to the objectively-ideal necessity of an inability-to-be-otherwise" (§§6 – 7, pp. 11 – 12). Husserl may well be mistaken in thinking that imaginability or conceivability entails possibility. But given the fact that he does equate these ideas, the relevant possibility in his understanding of dependence is genuine (metaphysical) possibility.

The concept of metaphysical dependence is an old one. It dates back at least to Aristotle, and arguably back to Plato, as he believed that the Forms existed independently from their instances. And it may even go back further. Well after the ancient period, we see it in Descartes, as well as in Husserl. And it is surely present in the writings of many western philosophers in between. The concept of metaphysical dependence also extends beyond the borders of classical western philosophy. Metaphysical dependence is central to some traditions of eastern philosophy.[6]

It is natural to posit some connection between the metaphysical dependence of one thing on another with the question of whether the one can exist without the other. I have tried to show that several influential western philosophers have endorsed such a connection. I am sure one could find the connection made in other prominent philosophers, both east and west, both past and present, and no doubt in the future as well. At this point, we should examine the contemporary literature on metaphysical dependence. We will see that a lot of work has been done, especially over the last 20 – 30 years, to articulate this and other related notions.

[6] See, for example, the writings of Nāgārjuna (Garfield 1995).

2. Now: metaphysical dependence on the contemporary scene

Contemporary metaphysics has a genuine interest in dependence. Trends in twentieth century metaphysics have focused on questions of existence, almost to the exclusion of everything else. What kinds of things exist? Do numbers exist? Do properties exist? It often goes unrecognized that questions of existence cannot easily be separated from questions of dependence. If something exists, then so do all of the other things upon which it depends. And given that something exists, the next relevant metaphysical question may well ask whether or not it is fundamental, which is a question about dependence: Does it or does it not depend on anything else?

It is also revealing to notice that many traditional metaphysical questions can be, and have been, phrased in terms of dependence. Does the external world depend on the mental? Does the mind depend on the body, or the brain? Does the whole depend on its parts? Do moral properties depend on non-moral, or natural properties? Do modal properties depend on non-modal, or actual properties?

It is likely that different conceptions of dependence are in play when we ask, and try to answer, questions like these. There are many relations that can be, and have been, understood as dependence relations. In Chapter 1 we briefly looked at various relations that might be considered dependence relations: supervenience, composition, constitution, determination, causation, truth-making, grounding, priority, holding "in virtue of", explanation, entailment, and arguably set-membership. For simplicity, we call these "dependence-like" relations.

What we have tried to do is single out one relation that is "dependence-like". That relation expresses a connection between the existence of one thing and the existence of another. The connection is a modal connection: one thing depends on another iff it cannot exist without the other. And so we refer to this relation as one that captures the modal/existential conception of metaphysical dependence.

To make the modal/existential conception of metaphysical dependence more clear, we have given what we call an "analysis" of this dependence-like relation.

MODAL ANALYSIS. *x metaphysically depends on y if, and only if, there is no possible world w such that x exists in w and y does not exist in w.*

While we call this an analysis, the intention is only to give necessary and sufficient truth conditions for claims that this particular dependence-like relation holds between a pair of objects.

We do not claim that this analysis gives necessary and sufficient truth conditions for other dependence-like relations. For example, the modal analysis does not give truth conditions for supervenience claims, or entailment claims. Though each of these dependence-like relations can be characterized as a dependence relation, they can differ from each other in important ways. Our claim is only that there is a particular conception of metaphysical dependence, the modal/existential conception, that expresses a genuine dependence-like relation, and that we can give accurate necessary and sufficient truth conditions for when this relation holds. Furthermore, we do not think that the modal/existential conception of dependence can be identified with, or is even equivalent to, any of the other dependence-like relations listed above.

In support of our claim that the modal/existential conception of dependence expresses a genuine dependence-like relation, we can point to discussions in the literature that make reference to this conception. Besides the historical references discussed already, the modal/existential conception of dependence can be found in many places in the contemporary literature. Consider, for example, Peter Simons (1987), speaking of the concept of "ontological independence": x is ontologically independent of y if x can exist without y, if there is a possible world in which x exists and y does not" (p. 301). Not only does Simons endorse the naive intuition, he endorses the more philosophically precise modal analysis of dependence as well. Kit Fine (1995a) also examines (a logically equivalent formulation of) the modal/existential conception of dependence. "One thing x will depend upon another y just in case it is necessary that y exists if x exists (or in the symbolism of modal logic, $\Box(Ex \to Ey)$)" (p. 270). Taking the box to be a quantifier over possible worlds, as is standard, this is precisely the modal analysis of dependence that we have articulated in our efforts to give a precise formulation of the modal/existential conception of dependence. Fine ultimately rejects the modal analysis as insufficient to accurately capture dependence relations between certain objects, sets and their members in particular. We ultimately reject that modal analysis for the same reason. But the fact that it is ultimately inadequate does not mean that it was not a natural place to start. And the fact that this particular analysis of the modal/existential conception of dependence fails does not imply that no successful analysis can be given in terms of existence and modality.

To be sure, one could find other references as well. But what the modal/existential conception has going for it is that it is intuitively plausible. Before the process of rigorous philosophical analysis begins, the idea that one thing depends on another if, and only if, the one can't exist without the other is a very natural and widespread idea. The modal/existential conception is therefore a plausible candidate for being a dependence-like relation.

That being said, it does differ from other dependence-like relations in several important ways. In fact, each of these dependence-like relations differs from the others in important ways. The ways in which these relations can differ may be divided into two kinds. They may differ with respect to their relata, the things or kinds of things that they relate. And they may differ with respect to the structural properties of the relation, e.g., whether they are reflexive, transitive, etc. To get a handle on the ways in which these relations may differ, we look at one familiar dependence-like relation — supervenience — in some depth. We follow this with a discussion of some other dependence-like relations — grounding, priority, and fundamentality — which play a prominent role in contemporary discussions of dependence, but which may not be as familiar.

Let's start with supervenience. Supervenience is an important dependence-like relation. Unlike the modal/existential conception of dependence, which expresses a relation between objects, supervenience is most commonly taken to be a relation between sets of properties. There are many different versions of supervenience, but the general idea is captured by the following conditions. Given two sets of properties, *A*-properties and *B*-properties, we have:

SUPERVENIENCE. *A-properties supervene on B-properties if, and only if, there cannot be a difference in A-properties without a difference in B-properties* (McGlaughlin and Bennett 2008).

We note that on this general conception, supervenience claims are modal claims, as they are equivalent to claims about what *cannot* happen.

Supervenience has had a prominent role in the philosophy of mind. Specifically, it has been used to capture the idea that the mental depends on the physical, as is often claimed by physicalists about the mind. The idea is that mental properties supervene on physical properties, in that there cannot be a difference in mental properties without there being a difference in physical properties. Consider, for example, Donald Davidson's position.

> Although the position I describe denies there are psychophysical laws, it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respects, or that an object cannot alter in some mental respect without altering in some physical respect" (1970, p. 214).

Davidson explicitly endorses the supervenience of the mental on the physical as a dependence-like relation. The supervenience claim expresses, or at least implies, that some kind of dependence holds: if the mental supervenes on the physical, then the mental depends on the physical.

That supervenience is a dependence-like relation is also explicit in more contemporary formulations of the supervenience thesis by Jaegwon Kim (2000, 2007).

> We must conclude then that mind-body supervenience itself is not an *explanatory theory*; it merely states a pattern of property covariation between the mental and the physical and points to the existence of a dependency relation between the two. Yet supervenience is silent on the nature of the dependence relation that might explain why the mental supervenes on the physical. Another way of putting the point would be this: supervenience is not a *type* of dependence relation — it is not a relation that can be placed alongside causal dependence, reductive dependence, mereological dependence, dependence grounded in definability or entailment, and the like. Rather, any of these dependence relations can generate the required covariation of properties and thereby qualify as a supervenience relation (2000 p. 14).

Kim argues that the supervenience of the mental on the physical suggests that the mental depends on the physical in some way, as it may be that the covariation of properties that supervenience requires is generated by the holding of some other dependence relation.

As we noted above, supervenience claims are equivalent to modal claims, claims about what cannot happen. In the context of physicalism about the mind, supervenience claims like those made by Davidson and Kim may entail that a further relation of metaphysical dependence holds, where the relevant conception of dependence is the modal/existential conception.

> Mental properties *supervene* on physical properties, in that necessarily, for any mental property M , if anything has M at time t, there exists a physical base (or subvenient) property P such that it has P at t, and necessarily anything that has P at a time has M at that time. ... That is, every mental property has a physical base that guarantees its instantiation. Moreover, without such a physical base, a mental property cannot be instantiated (Ibid., pp. 9 – 10).

If we suppose that mental properties supervene on physical properties, then it is impossible for there to exist some x which has property M without there existing another property P such that x has P. Here we have metaphysical dependence, according to the modal/existential conception, between properties M and P.

As we mentioned before, supervenience is not a relation between objects. It is a relation between sets of properties. In this way it differs from the modal/existential conception of metaphysical dependence. One could argue that these two dependence-like relations differ in other ways as well. Specifically, one may think that they differ with respect to their structural properties. Supervenience is not symmetric: there may be some sets of properties that supervene on each other, though this certainly does not hold for every pair. For example, the surface area and volume of a sphere supervene on each other. Many take metaphysical dependence to be asymmetric: there is no pair of objects such that each object depends on the other. Supervenience is also reflexive: everything supervenes on itself. This follows in a straightforward way from the above characterization of Generic Supervenience: as there cannot be a difference in A-properties without a difference in A-properties, it follows that A-properties supervene on A-properties. If you think metaphysical dependence is asymmetric, then irreflexivity follows: nothing depends on itself. Metaphysical dependence is usually characterized as being asymmetric and irreflexive, but we will take up the question of whether these structural properties accurately characterize metaphysical dependence in Chapter 6.

We can see differences and similarities along these lines when we compare other dependence-like relations. Supervenience and determination hold between properties. Composition, constitution, set-membership, and perhaps priority hold between objects. Causation, explanation, and holding "in virtue of" hold between facts. Entailment holds between propositions or sentences. Truth-making and grounding cut across metaphysical categories. Supervenience, composition, and entailment have reflexive and symmetric cases. Constitution, set-formation, priority, explanation, holding "in virtue of", truth-making, grounding, and arguably causation are asymmetric, and therefore irreflexive. Though this is not an exhaustive comparison of these relations, we can see that they differ in many ways. Arguably they are all different kinds of dependence relations. And there are probably other dependence-like relations that I have not included, which differ from all of these in various ways. There is nothing wrong with using the word "dependence" as an umbrella term to refer to this family of concepts. Each one is a variety of dependence.

Some of these relations, however, have played a particularly prominent role in contemporary discussions of dependence. These are grounding and priority. Additionally, the notion of fundamentality, which is not straightforwardly a relation, has also played a significant part. We consider in some depth the connections between the modal/existential conception of metaphysical dependence and these other notions. However, as the discussion of metaphysical dependence has only recently begun to flourish, these concepts are not yet sufficiently well-defined or delineated, which leads to some degree of confusion in the literature. Some of these concepts are often run together. Different people mean different things when they deploy these concepts, and they are not sufficiently clear about how they are using them. By examining the connections between the modal/existential conception of dependence on one hand, and grounding, priority, and fundamentality on the other, the goal is to untangle this jumble of concepts.

To start, compare dependence with grounding. You might take dependence simply to be the converse relation of grounding: x metaphysically depends on y if, and only if, y grounds x. But grounding does seem like a slightly different concept. For example, it seems true that a whole is grounded (at least in part) in each of its parts. But it is not obvious that a whole metaphysically depends, in a modal/existential way, on each of its parts. It is not the case that, necessarily, if a particular part did not exist, then the whole would not exist. It is possible that my thumb is chopped off and annihilated, yet I still exist. Conversely, I would argue that a person metaphysically depends on his or her parents. Certainly it's true that, necessarily, if your parents do not exist (at any time), then you do not exist (at any time). The modal/existential conception of dependence would imply that you are therefore dependent on your parents. But it seems at the very least awkward to say that an individual is grounded in his or her parents.[7]

There may be cases where x depends on y and y grounds x. Perhaps this happens in the set-theoretic cases. The man Socrates grounds the set {Socrates}. Furthermore, {Socrates} metaphysically depends on Socrates, at least on the naive conception, because you can't have {Socrates} without Socrates. I am actually skeptical about applying the terminology of ground to the set-theoretic cases, as it still is not clear to me what exactly the grounding relation consists in. But however it comes to be understood, there may be cases where grounding and dependence match up. I just do not think that they match up in every case.

Grounding might be given a slightly different interpretation such that to say x grounds y means that x is *the ground* of y. On this interpretation, x is fundamental: there is nothing that grounds x. We can imagine a chain of objects from x to y, such that each object in the chain bears some relation to the next.

$$x < o_1 < o_2 < \ldots < o_n < y$$

In this example, x, and only x, is the ground of each of $o_1, o_2, \ldots, y$. Then we have some other relation, symbolized as $<$, that holds between adjacent members in the chain. We might call this relation *priority*. Each object is metaphysically prior to every object that "comes after" it in the chain. It seems that priority understood in this way is simply equivalent to our original conception of grounding. However, if we reserve the term grounding for the relation that holds between two objects when one is the fundamental ground of the other, then we could define grounding in terms of priority as follows: x grounds y iff x is prior to y, and nothing is prior to x.

We have used the term 'fundamental' to express a two-place grounding relation between objects, where one is the ground of another. But fundamentality can be used as a one-place predicate, describing a property of something, rather than how one thing relates to another. Understood this way, fundamentality can be defined in terms of priority as well: x is absolutely fundamental iff

[7]Perhaps this is a controversial case, as the dependence here may be more properly described as causal rather than metaphysical. However, the case does satisfy the truth conditions according to the modal analysis.

nothing is prior to x. Alternatively, we may define absolute fundamentality in terms of metaphysical dependence as well: x is absolutely fundamental iff x does not metaphysically depend on anything. Given that dependence and priority are distinct, we must take care to specify which conception we have in mind when deploying the notion of absolute fundamentality. Either way, if we represent the metaphysical structure of the world as a partial ordering, ordered either by the priority relation or the dependence relation, then absolutely fundamental objects are the *minimal* objects.

To make things more confusing, we might introduce a relative sense of fundamentality, where one thing is more fundamental than another. Relative fundamentality is a two place relation, though it differs from the relation according to which one thing fundamentally grounds another. Under this conception, one thing may be more fundamental than another, though neither need be *fundamental* in any absolute sense. With a conception of relative fundamentality in hand, we could then use it to define a notion of absolute fundamentality: x is absolutely fundamental iff there is nothing that is more (relatively) fundamental than x. A relation of relative fundamentality may hold between two objects even when neither object is absolutely fundamental. For example, one might say that an atom is more fundamental than a brick, while maintaining that neither is absolutely fundamental, as there are things (protons, electrons, etc.) that are more fundamental than both atoms and bricks.

Is there a connection between relative fundamentality and dependence? You might argue that x metaphysically depends on y iff y is more fundamental than x. But there are two issues with this analysis. First, you then need to give some account of what it means for one thing to be more fundamental than another. This is no straightforward task. Second, it is not clear that any account of relative fundamentality will be quite right as an analysis of metaphysical dependence. For it seems that an atom in my left thumb is more fundamental than a brick in the Empire State Building, but few would say that the Empire State Building metaphysically depends on an atom in my left thumb. The Empire State Building does not depend on any part of me.

If there is a coherent notion of relative fundamentality, where we say that one thing is more fundamental than another, the notion suggests that reality has a hierarchical structure, where this hierarchy is divided into levels, measured in degrees of fundamentality. At the bottom (if there is a bottom) we have *the fundamental*, upon which everything else depends. The rest of reality can then be ordered in some way according to relations of relative fundamentality. We have been speaking of these as relations between particular objects: this brick, that atom, the Empire State Building. But you might see relative fundamentality as arising from relations between kinds of things. A particular electron is more fundamental than a particular atom because every electron is more fundamental than every atom. Metaphysical dependence, on the other hand, can be taken as a relation between particulars. A particular statue depends on a particular lump of clay, but not because every statue depends on every lump of clay. This may explain why the two notions come apart, as with the atom and the Empire State Building.

Many think that absolutely fundamental objects, if there are any, are located in the very small, the "atoms". Not the real, biochemical things that we call atoms, but the simples, the things without parts. The motivation here is that absolute fundamentality, and relative fundamentality, track something like the notion of proper parthood: x is absolutely fundamental iff x has no proper parts; x is more fundamental than y iff x is a proper part of y. Though this doesn't quite seem right. For one thing, it's not clear that my hand is more fundamental than I am, even though my hand is a proper part of me. Furthermore, the atom in my left thumb does seem more fundamental than the Empire State Building, though the atom is not a proper part of the building. So the connection between fundamentality and parthood is not as obvious as it may at first seem. But the misguided intuition that there is an obvious connection has been motivation for locating the fundamental in the very small. On this picture, the really small things are more fundamental than the mid-sized things; the mid-sized things are more fundamental than the somewhat big things; the somewhat big things are more fundamental than the really big things; and the really big things are more fundamental than the totality of things, the cosmos. That's a pretty common picture of the metaphysical structure of the world. There are some, however, who do not endorse this picture. Jonathan Schaffer (2010) argues that fundamentality tracks part-hood in the opposite direction. For Schaffer, there is just one fundamental thing: the cosmos, the entire universe. Proper parts of the cosmos metaphysically depend on, and are grounded in, the cosmos.

I'm not quite sure if fundamentality should track the proper parthood relation in any direction. But I do agree that objects are the kind of thing that can be fundamental, and that objects are the kind of thing such that one can be more fundamental than another. And I have been talking as if metaphysical dependence is also a relation between objects. On the modal/existential conception of metaphysical dependence, such talk makes sense. We say that one object depends on another because you can't have the one object without the other. It is impossible for the one object to exist without the other. There is no possible world where the one object exists and the other does not.

But I think that metaphysical dependence is more properly understood as a relation between states of affairs. In fact, I hold the stronger position that metaphysical dependence holds between facts — states of affairs that actually obtain. We can interpret the metaphysical dependence of one object on another in terms of facts about the existence of those objects. Talking about two objects x and y, we say that "x metaphysically depends on y" when the fact $< x$ exists $>$ metaphysically depends on the fact $< y$ exists $>$.

Talk of metaphysical dependence in terms of facts allows us to extend our conception of metaphysical dependence. For example, we can say that the fact that the statue is solid metaphysically depends on the fact that the lump of clay is solid. And the modal analysis of metaphysical dependence can be modified in a natural way so as to relate facts instead of objects.

MODAL ANALYSIS (FACTS). *The fact* $< x >$ *metaphysically depends on the fact* $< y >$ *if, and only if, there is no possible world* w *such that* $< x >$ *obtains at* w *and* $< y >$ *does not obtain at* w*.*

This version of the modal analysis can accommodate claims that one object metaphysically depends on another. We can then say that, e.g., the statue metaphysically depends on the lump because there is no possible world w such that $<$ the statue exists $>$ obtains at w and $<$ the lump exists $>$ does not obtain at W.

Taking facts to be the relata of the relation of metaphysical dependence allows us to evaluate claims that dependence seemingly holds between relata of different ontological kinds. For example, one might argue that a fact metaphysically depends on the objects that are involved in that fact. The fact that $<$ Socrates is in the market $>$ depends on Socrates. And this dependence holds because there is no possible world *w* such that $<$ Socrates is in the market $>$ obtains at *w* and $<$ Socrates exists $>$ does not obtain at *w*.

Or one might argue that the truth of a statement metaphysically depends on the objects that the statement is about.

> [I]f there is a man, the statement whereby we say that there is a man is true, and reciprocally — since if the statement whereby we say that there is a man is true, there is a man. And whereas the true statement is in no way the cause of the actual thing's existence, the actual thing does seem in some way the cause of the statement's being true: it is because the actual thing exists or does not that the statement is called true or false (Aristotle, *Categories*, 14b17-23)

In other words, truth depends on being. Aristotle labels this dependence as causal, but it could be evaluated as a claim of metaphysical dependence.[8] The truth of the statement depends on the man. Suppose that the man in question is Socrates. The truth of the statement 'Socrates exists' depends on Socrates because there is no possible world *w* such that $<$ 'Socrates exists' is true $>$ obtains at *w* and $<$ Socrates exists $>$ does not obtain at *w*.

Taking metaphysical dependence to hold between states of affairs can handle other cases as well. For instance, it is sometimes said that particularized properties, or tropes, metaphysically depend on the things that instantiate them. A particular apple's redness metaphysically depends on the apple. For simplicity, let's give the apple a name — Apple — and the trope a name — Red. This dependence holds because there is no possible world *w* such that $<$ Apple instantiates Red $>$ obtains at *w* and $<$ Apple exists $>$ does not obtain at *w*.

Though I take the relation of metaphysical dependence to hold between facts, I will continue to speak of it as a relation that holds between objects. I will do this because almost all of the dependence claims that we evaluate in the remainder of the dissertation are claims that the existence of one object depends on the existence of another. So we will appeal to the modal analysis as originally stated.

MODAL ANALYSIS. *x metaphysically depends on y if, and only if, there is no possible world w such that x exists in w and y does not exist in w.*

[8]The kind of causation Aristotle has in mind is most likely not efficient causation. The relationship between man and statement is not one that brings about some change or alteration.

As we will see, the modal analysis is unable to accurately capture the asymmetric dependence that holds between sets and their members. So we will be considering alternative analyses. These alternatives will also be stated as if the dependence relation holds between objects. Translating the conditions on dependence relations between objects into conditions on dependence relations between states of affairs can be done in these cases as we have done here.

The preceding discussion has hopefully made plausible the view that dependence is worthy of serious metaphysical discussion. We have tried to show that the modal/existential conception is an intuitive start to understanding one kind of metaphysical dependence. And we have tried to distinguish the modal/existential conception of metaphysical dependence from other dependence-like relations. From here on, when we refer to the relation of metaphysical dependence, we are referring to metaphysical dependence according to the modal/existential conception. We have also argued that metaphysical dependence should be understood as a relation that holds between facts, i.e., states of affairs that actually obtain, though we will proceed, for simplicity, to talk of it as a relation between objects. We have started to examine a potential analysis of the modal/existential conception of metaphysical dependence. On this analysis, called the modal analysis, claims about the dependence of one thing on another are to be evaluated in terms of what possible worlds exist and what things do or do not co-exist in those worlds. In the next chapter, we will look at one specific case of metaphysical dependence, the dependence of a set upon its members. We will argue that there is reason to think that sets do in fact metaphysically depend on their members. And we will show that the modal analysis does not provide plausible truth conditions for dependence in the set-theoretic cases.

CHAPTER 3

The Modal Analysis of Dependence

The modal analysis gives truth conditions for the modal/existential conception of metaphysical dependence. We want truth conditions that confirm our intuition that sets depend on their members. We would also like these conditions to show that, in most cases, this dependence is asymmetric. The modal analysis does not do this. For this reason, the modal analysis is unsatisfactory. The second part of this chapter shows in detail the many ways in which the modal analysis fails to capture our intuitions about set-theoretic dependence. Before that, though, we explore reasons for thinking some relation of dependence holds between a set and its members. First, we examine a general argument that sets depend on their members based on the idea that sets are simply collections of objects. Then we show that the idea that sets depend on their members has been used to justify the claim that the cumulative hierarchy of sets exhausts the set-theoretic universe. This is not so much an argument that sets depend on their members. Rather, it is evidence that some people working in set theory actually think a relation of dependence holds in the set-theoretic cases.

1. Set-Theoretic Dependence

The question of whether sets depend on their members is of general metaphysical importance, as discussed in Chapter 2. But it is also of specific importance to the philosophy of mathematics. Set theory is supposed to provide a foundation for all of mathematics. A strong endorsement of this position can be found in Kunen (2009): "[a]ll abstract mathematical concepts are set-theoretic. All concrete mathematical objects are specific sets" (p. 14).[1] To be sure, Kunen's view is extreme. But most mathematicians accept a strong connection between set theory and the rest of mathematics. Many would agree that every mathematical object can be represented as a set, and that every mathematical statement can be translated into a statement about sets. Set theory is therefore the study of the nature and extent of the mathematical universe. Moreover it is seen to be a virtue when

[1] See also Mayberry (2000).

it is clear how to interpret a mathematical proof in the language of set theory. Doing so gives more support for the correctness of the proof. Furthermore, set theory, as the study of properties of and relations between sets, is its own branch of mathematical inquiry, independent of the foundational role of the universe of sets. What dependence relations hold between these sets will then tell us something interesting about the nature of some of the objects that mathematics studies.

We claim that a relation of metaphysical dependence holds between a set and its members. A set depends on each one of its members for its existence. For any s and S, if $s \in S$, then S metaphysically depends on s. Note that the converse implication does not automatically hold. It is not automatically the case that if S depends on s, then $s \in S$. This is easiest to see by noting that the dependence relation is transitive, while set-membership is not. The set-membership relation, therefore, is not, or is not equivalent to, a dependence relation.

In discussions of metaphysical dependence, many assume, or at least grant the intuition, that sets depend on their members. They usually do so without any argument.[2]

Here is a quick argument that sets metaphysically depend on their members: you can't have the set without the members. Sets *need* their members to exist. We have a naive understanding of metaphysical dependence, according to the modal/existential conception, where one thing depends on another iff you can't have the one without the other. Consider any set $S = \{s_1, s_2, \ldots\}$. For any s_i, S metaphysically depends on s_i because you can't have S without s_i. So we have that a set metaphysically depends on each and every one of its members.

Here is a quick argument that objects do not metaphysically depend on the sets that they are members of: you can have the members without the set. Members do not need the sets that they are members of. Given a set $S = \{s_1, s_2, \ldots\}$, you can have any s_i without S. Even stronger, you might even think that you can have all of the s_i together without S. So we have that objects do not metaphysically depend on the sets that they are members of.

But the second argument is not quite as smooth as the first. One may accept the first argument as relatively straightforward, while being suspicious of the second. The second argument prompts the question: What sense of "can" is at play here? The word indicates that we are dealing with

[2]See Correia (2005) Fine (1995a), Lowe (2005), Schaffer (2010), Schnieder (2007), Trogdon (forthcoming), among others.

a notion of possibility. But there are many kinds of possibility: physical, metaphysical, logical, epistemic, ethical. There may be a sense of "can" such that, it cannot be that once you have all the members you don't have the set as well. From this perspective, once you have the members, you get the set for free.

The hesitation to accept the argument most likely comes from the fact that it makes use of a naive conception of dependence. The idea of having one thing without another has not been made sufficiently precise. We make the naive conception of dependence more precise in Section 2 of this chapter.

1.1. What sets are. In the meantime we examine an independent argument in favor of the metaphysical dependence of sets on their members. This argument starts with our basic conception of what a set is. A set is, very simply, any collection of well-defined objects into a whole. One of our first conceptions of set comes from Cantor's *Grundlagen*:

> In general, by a 'manifold' or 'set' I understand every multiplicity which can be thought of as one, i.e., every aggregate of determinate elements which can be united into a whole by some law. (Cantor 1883, p. 85).

Generally, we collect together objects that have something in common. Sets are often defined in terms of a property or properties that all of its members satisfy. "Intuitively, a set is a collection of all elements that satisfy a certain given property" (Jech 2006, p. 3). A naive approach to sets takes any property to define a set, endorsing an unrestricted version of the comprehension axiom.

UNRESTRICTED COMPREHENSION. *There exists a set y, such that for all x, x is a member of y if, and only if, x has property Φ.* $\exists y \forall x[x \in y \leftrightarrow \Phi(x)]$

Unrestricted Comprehension is an axiom schema, so that there is one such axiom for every property (or every formula) Φ. Unrestricted Comprehension produces an inconsistent set theory, due to Russell's paradox, among others. The standard Zermelo-Fraenkel set theory (ZF) uses a separation axiom instead.

SEPARATION. *For any set z, there exists a set y, such that for all x, x is a member of y if, and only if, x is a member of z and x has property Φ.* $\forall z \exists y \forall x[x \in y \leftrightarrow x \in z \wedge \Phi(x)]$

Separation is also a schema, with one axiom for each formula Φ. Whether you endorse the naive approach with Unrestricted Comprehension or the modern axiomatic approach with Separation, both axioms capture the idea that sets are collections, bringing together objects, that are already present, having some property in common.

Sets collect multiple objects into a whole; they collect a many into a one. Given any objects, there is another single object — the collection, or set, of those objects. But in order to have any collection of objects, whether they are objects of thought or otherwise, one needs to have the objects to collect. One cannot collect things that do not exist. In this sense, the objects are prior to, one might say necessary for, the existence of the set.

Conversely, sets are not in the same way necessary for, nor prior to, their members. Understood as a collection, a one from many, a set is something over and above its members. Conceptually, we can have the objects without the set, but we cannot have the set without the objects. But that's just what the modal/existential conception requires for metaphysical dependence: one thing depends on another when you can't have the one without the other. So it follows from what sets are, as collections of objects, that sets metaphysically depend on their members in this modal/existential way.

We have argued that, in order for something to be collected into a set, it must exist. You might question this claim, however. Both the naive and the axiomatic approach to sets say that any property, or any formula, Φ can be used to define a set. If our language has modal operators, then we could consider the set of all x such that "Possibly, $\Phi(x)$" holds. For example, you might consider the set of all possible lions, or the set of all possible unicorns. These sets will have as members objects that do not exist.

I think the appropriate response here is as follows. If you do believe that there are non-existent objects, then you may feel free to collect those objects into sets as you wish. However, any set that has non-existent objects will itself be non-existent. All we claim is that if a set exists, it's members must exist as well, regardless of what other existential status objects can have.

The claim, then, is that sets metaphysically depend on their members in virtue of what sets are. Sets are collections of objects. Furthermore, sets are extensional — they are *mere* collections of objects. It doesn't matter how those objects are related to one another. For example, they need not be ordered in any particular way. That sets are extensional is captured, in set theory, by the Axiom of Extensionality.

EXTENSIONALITY. *For all sets x and y, if it's true that for all z, z is a member of x if, and only if, z is a member of y, then x is identical to y.* $\forall x \forall y[\forall z(z \in x \leftrightarrow z \in y) \rightarrow x = y]$

Extensionality imposes necessary and sufficient conditions for when two sets are identical: they must have exactly the same members. Identity conditions tell us a lot about what things are, as it tells us what makes two of these things the same. One can derive from Extensionality an argument in favor of the claim that sets metaphysically depend on their members, according to the modal/existential conception of dependence. In effect, this is a more detailed version of the argument given at the beginning of this section.

The Axiom of Extensionality is an axiom of formal, mathematical Set Theory. The idea behind axiomatic set theory is that we take the membership relation to be primitive, and use axioms to determine how it works. The standard axiomatization is Zermelo-Fraenkel set theory with the Axiom of Choice (ZFC). Extensionality is one of the axioms of ZFC. But Extensionality is not your average axiom. It differs from the other axioms of ZFC because it is definitional in nature. It gives the identity conditions for sets, and so it describes one of the basic properties of the kind of objects set theory deals with. It tells us something fundamental about the objects of set theory.

> The axiom of extensionality enjoys a special epistemological status shared by none of the other axioms of ZF. Were someone to deny another of the axioms of ZF, we would be rather more inclined to suppose, on the basis of his denial alone, that he believed that axiom false than we would if he denied the axiom of extensionality. ... [I]f someone were to say, 'there are distinct sets with the same members,' he would thereby justify us in thinking his usage nonstandard far more than someone who asserted the denial of some other axiom. ... That the concepts of *set* and *being a member of* obey the axiom of extensionality is a far more central feature of our use of them than is the fact that they obey any other axiom. A theory that denied, or even failed to affirm, some of the other axioms of ZF might be called a set theory, albeit a deviant or fragmentary one. But a theory that did not affirm that the objects with which it dealt were identical if they had the same members would only by charity be called a theory of *sets* alone. (Boolos 1971, pp. 27 – 8).[3]

[3] See also Burgess (2004).

One can deny the other axioms of set theory and still claim to be talking about sets. But to deny Extensionality is to change the subject. That Extensionality is definitional is further evidenced by its inclusion in the majority of alternative axiomatizations of set theory, including von Neumann-Bernays-Gödel set theory, Quine's *New Foundations*, and in the contemporary approaches to naive (or paraconsistent) set theory (see Priest 2006, Ch. 10, and Weber 2009, 2010). Other set theories, like Potter's ZU (1990, 2004), do not include Extensionality as an axiom because it follows easily from other definitions.

The Axiom of Extensionality suggests that some kind of dependence holds between a set and its members. Two sets are identical if and only if they have exactly the same members. These identity conditions imply that the identity of a set depends, in some sense, on its members. If a set had different members, it would be a different set.

But this identity dependence is not the same kind of dependence that the modal/existential conception of dependence tries to capture. In the case of sets, however, one can argue from identity dependence to modal/existential dependence. Consider a set S with members $s_1, s_2, \ldots$. The set S is the set that it is because it has each s_i as one of its members. If any particular member s_i of S failed to exist, then S could not have s_i as a member. S would thus lose its identity. It would no longer be S; it would be some other set. So S's identity depends on the existence of its members. But in this situation nothing else could be the set S either, as nothing could include s_i as a member, given that s_i does not exist. So the set S would not exist at all. A set's existence, therefore, depends on the existence of its members. You can't have the set without the members.

There is another potential argument in this area. You might think that sets metaphysically depend on their members because sets have their members as parts. That sets have their members as parts does not follow automatically from our basic conception of sets as collections of objects. And perhaps we don't usually think of members as being parts of the sets that contain them. If we are to think of sets having parts at all, it is more common in the contemporary literature to take the (nonempty) subsets of a set to be its parts.[4] But Kit Fine (2010) argues for a pluralist approach to parthood, such that there are many conceptions of part, one of which matches up with

[4] See Lewis (1991).

the set-membership relation. “[The pluralist] may well think that the way in which a pint of milk is part of a quart is different from the way in which the letter (type) ‘c’ is part of the word (type) ‘cat’ and different from the way in which a member is part of a set” (p. 561). Indeed, if we are going to be pluralists about the notion of part, then we can say that members are parts of sets on one conception and subsets are parts of sets on another.

If we do take members to be parts of sets, then one can apply an argument that wholes metaphysically depend on their parts to show that sets metaphysically depend on their members. It is not clear, however, that wholes actually do metaphysically depend on their parts, especially under the modal/existential conception of dependence. In fact, we briefly argued against this view in Chapter 2. My thumb is a part of me. But it's not the case that you can't have me without my thumb. I do not, therefore, metaphysically depend on my thumb.

However, you may be a mereological essentialist, in which case parts are essential to their wholes.[5] If you are a mereological essentialist, then you can't have a whole without each of its parts. Taking members to be parts of sets, you have that sets metaphysical depend on their members. Even if you are not a mereological essentialist, you may still think that sets have their parts essentially. This is reasonable, given the extensional nature of sets. If you took Aristotle out of the set {Plato , Aristotle}, and put in Socrates, you'd have a different set {Plato , Socrates}. If you take sets to have their parts essentially, and you take their parts to be their members, then you can't have a set without each of its members. On the modal/existential conception of dependence, you have that sets depend on their members. This may seem a trivial result, as we normally understand essence modally. But one need not do so.[6] And with a non-modal understanding of essence, it hopefully still follows that if one thing is essential to a second thing, then you can't have the second without the first. In which case, you've got metaphysical dependence.

1.2. The Iterative Conception of Set. According to the previous section, that sets metaphysically depend on their members follows from our basic conception of what sets are. I think this is

[5]See Chisholm (1973, 1976).

[6]Kit Fine (1994), for example, takes a non-modal approach to essence.

probably one of the strongest arguments in favor of the dependence claim. But I would like to consider another reason why one might think that sets metaphysically depend on their members. This discussion does not start with the conception of what sets *are*. Rather, it starts with the conception of what sets *there are*.

Here's a conception one could have of what sets there are: For any property Φ, there is a set of all and only those things that have Φ. We've seen this before. It is a version of Unrestricted Comprehension, and it is one of the axioms of (a straightforward version of) naive set theory. Interestingly, in connection with the previous section, the only other axiom is Extensionality. These are the two axioms of naive set theory — Extensionality to tell us what sets are, and Unrestricted Comprehension to tell us what sets there are. Unrestricted Comprehension tells us what sets exist. Those who endorse this principle take it to characterize the set-theoretic universe.

Not many people endorse Unrestricted Comprehension. At least not anymore. Because it is well known that Unrestricted Comprehension leads to paradox. So while naive set theory is a natural approach to set theory, the inconsistency of Unrestricted Comprehension has led many to drop that axiom. The axioms that replace it are based on a slightly different conception of what sets there are. This is called the iterative conception, and it captures a picture of the set-theoretic universe that many people working in set theory accept.[7]

The iterative conception is often traced to Zermelo (1930), and a version of it is explicit in Gödel (1947). It is the concept of set "according to which a set is anything obtainable from the integers (or some other well-defined objects) by iterated application of the operation 'set of' " (p. 180). Hao Wang (1974, p. 187) even argues that this conception was what Cantor had in mind when he first developed set theory.

The iterative conception results in a picture of the set-theoretic universe being built up in stages. To build anything, one needs something to start with. We start with the empty set $\emptyset$, the set with no members. Once we have the empty set, use it to build more sets. For example, given the empty set $\emptyset$, we can take its singleton $\{\emptyset\}$. Now we have two sets. Use these two sets to build more sets,

[7]See Boolos (1971, 1989).

$\{\{\emptyset\}\}$ and $\{\emptyset, \{\emptyset\}\}$. And off we go. This process does not end. Through successive applications of building operations, we can construct the universe of sets.

More precisely, the universe of sets can be divided into levels V_α. Accordingly, the first level is the empty set $\emptyset$, and each successive level consist of the power set — the set of all subsets — of the previous level. This is iterated along the ordinals. At limit ordinals, we simply collect up every set that is generated at lower levels.

$$\begin{aligned} V_0 &= \emptyset \\ V_{\alpha+1} &= \mathscr{P}(V_\alpha) \\ V_\lambda &= \textstyle\bigcup_{\alpha<\lambda} V_\alpha \text{ if } \lambda \text{ is a limit ordinal} \end{aligned}$$

This universe V of sets, which consists of all of the levels generated in this way, is often called the cumulative hierarchy. It differs from the universe of sets that results from the naive conception in that, as far as we know, there are no sets that generate inconsistencies. On the iterative conception, the universe is a hierarchy of sets, with the empty set serving as its foundation. Without the empty set, the hierarchy cannot "get off the ground". Without the empty set, there would be no hierarchy.

This picture of the set-theoretic universe implies a dependence relation between sets and their members. The dependence relation gives some structure to the set-theoretic universe. Members are prior to the sets that contain them. They must be available before the sets that contain them can be constructed. Sets are built out of their members. Of course the building metaphor is only a metaphor, and in some respects a poor one. For instance, things are usually built in space and over time. Sets are not built in this sense. Members are metaphysically prior, not temporally prior to the sets that contain them. The dependence relation that this picture suggests is one of metaphysical dependence under a modal/existential conception: you can't have the set without the members.

The claim that the iterative conception of a set is based on some notion of metaphysical dependence is made explicit in Potter (2004). "[T]here is a fundamental relation of presupposition, priority, or ... *dependence* between collections" (p. 36, emphasis in original). But Potter wonders about exactly what this dependence relation amount to. It is not temporal priority. Nor is it, according to him, a relation that can be captured in terms of our normal understanding of necessity, for basically the same reasons that Fine (1995a) gives, and which we explore in more detail in the second half of this chapter. We subsequently argue, in Chapter 6, that there is a coherent way to understand this dependence between a set and its members.

It is therefore the widely accepted iterative conception of set, with its built-in notion of priority, that seems to bring with it the claim that sets metaphysically depend on their members. The most straightforward understanding of priority is that it is a dependence relation. And this conception of priority at least entails a modal/existential conception of dependence: you can't have a set without its members. Indeed, this conception of the set-theoretic universe might be seen to entail a particularly strong form of set-theoretic dependence, stronger than the modal/existential conception.

Within the cumulative hierarchy, certain kinds of sets — self-membered sets, pairs of sets that are members of each other, sets with infinitely descending chains of membership — do not exist. From this fact about the set-theoretic universe and the claim that sets metaphysically depend on their members, one might try to argue that the dependence relation has certain structural properties: asymmetry (and thus irreflexivity) and well-foundedness. But just because the membership relation has these properties in the cumulative hierarchy does not entail that the dependence relation must have these properties as well, even when restricted to the context of sets.

The mistaken view that set-theoretic dependence must have these structural properties is exacerbated by the claim, endorsed by some, that set-theoretic dependence is what justifies some of the standard axioms of set theory. Once we have the iterative conception of set, the next step is to axiomatize it. The standard axiomatization of set theory is Zermelo-Fraenkel set theory, with the Axiom of Choice (ZFC). Most of the axioms, like Pairing and Union and Power Set tell us, given that some sets already exist, what others set we can add.

PAIRING. *For all sets x and y, there exists a set z such that for all u, u is a member of z if, and only if, u = x or u = y.* $\forall x \forall y \exists z \forall u (u \in z \leftrightarrow u = x \vee u = y)$

UNION. *For all sets x, there exists a set z such that for all u, u is a member of z if, and only if, there exists a y such that y is a member of x and u is a member of y.*
$\forall x \exists z \forall u [u \in z \leftrightarrow \exists y (y \in x \wedge u \in y)]$

POWER SET. *For all sets x, there exists a set z such that for all u, u is a member of z if, and only if u is a subset of x.* $\forall x \exists z \forall u(u \in z \leftrightarrow u \subseteq x)$

All of these axioms say that given some sets exist, we can add more sets that satisfy certain conditions.[8] They are justified by the view, made explicit by the iterative conception, that once you have some sets, you can build more sets out of them. But there are two axioms that do not do this. The first is Extensionality, which does not say anything about what sets exist. Extensionality simply gives, for all of the sets that do exist, the conditions under which any two of them are identical. The justification for Extensionality, as we have seen, is that it tells us what kind of objects we are dealing with. The other axiom is the Axiom of Regularity, which is also known as the Axiom of Foundation.

FOUNDATION. *Every non-empty set x contains a member y such that x and y are disjoint sets.* $\forall x[x \neq \emptyset \rightarrow \exists y(y \in x \wedge y \cap x = \emptyset)]$

Foundation also tells us what sets exist. But it is not a building principle; it is a restrictive principle. Foundation does not say, given the existence of some sets, we can derive the existence of some more sets. Foundation says that certain kinds of sets cannot exist in the set-theoretic universe. As a consequence of Foundation it follows that there are no sets that are members of themselves, that there are no pairs of sets that are members of each other, and that there are no sets that have infinitely descending chains of membership.

The primary justification for Foundation is that it follows from the iterative conception of set. Foundation is required to accurately represent the cumulative hierarchy of sets as exhausting the set-theoretic universe. Without Foundation, it is consistent with the remaining axioms that there exist sets that are not in the cumulative hierarchy, i.e., not generated at any level according to the iterative conception. These are the non-well-founded sets — the self-membered sets, the pairs of sets that are members of each other, the sets that have infinitely descending chains of membership.[9]

[8]The Axioms of Choice and Replacement and the Axiom Schema of Separation do this as well. The Axiom of Infinity is an even stronger set existence axiom.
[9]See Aczel (1988) and Barwise & Moss (1996).

The iterative conception rules these sets out. And so there should be an axiom that rules these sets out. The Axiom of Foundation does just this.

Notice that Foundation is not justified on mathematical grounds. Set theory does not need the axiom in order to serve as a foundation of mathematics. "Foundation is the one axiom unnecessary for the recasting of mathematics in set-theoretic terms" (Kanamori 2009, p. XV). It may be that Foundation makes doing set theory easier. Some proofs may be shorter when they make use of Foundation. Some definitions that Foundation allows may be more convenient. But the axiom is not essential. One can do set theory perfectly well without Foundation.

But some have suggested that Foundation is justified on the grounds that it follows from the dependence of a set on its members. As noted by Charles Parsons "the priority of the elements of a set to the set ... is reflected in the theory itself by the fact that membership is a well-founded relation" (1990, p. 336). Recall that priority cannot be understood as temporal priority. If we take the priority claim to be a dependence claim, according to which one thing depends on another iff you can't have the one without the other, then it is not clear that priority justifies the strength of the Axiom of Foundation. Foundation rules out non-well-founded sets, but we will show in Chapter 6 that a plausible understanding of the modal/existential conception of dependence does not. The strength of Foundation is therefore not fully justified solely by the fact that sets metaphysically depend on their members.

Some have tried to argue that one need not, and perhaps should not posit a dependence relation between sets and their members in order to endorse the iterative conception of set. Luca Incurvati (forthcoming) argues instead for what he calls the minimalist conception of set. Incurvati's targets are those who try to use the idea that sets metaphysically depend on their members to argue in favor of the view that the cumulative hierarchy exhausts the universe of sets. He claims that, because we have no plausible analysis of metaphysical dependence, we must take it as primitive. But if we take it as primitive, then we have no reason to believe it is well-founded, or irreflexive. There is no reason to believe that it has any of the properties we normally take it to have. And so a primitive dependence relation cannot serve to explain why the universe of sets is exhausted by the cumulative hierarchy.

As a primitive dependence relation is not explanatory, we should dispense with it. Instead, according to Incurvati, we should be minimalist about sets in the following respect.

> According to the minimalist account, the content of the iterative conception is exhausted by saying that it is the conception of set according to which sets are the objects that occur at one level or another of the cumulative hierarchy. The analogy here is with the case of the natural numbers: it is plausible to regard our conception of the natural numbers as exhausted by saying that numbers are the objects that occur at one point or another in the natural number series. Thus, just as on this view our conception of the natural number structure can be conveyed by saying that the natural numbers are 0, and its successor, and *its* successor, and *ITS* successor, and so on, on the minimalist account the iterative conception of the universe of (pure) sets can be conveyed by saying that the sets are the empty set and the set containing the empty set, sets of those, sets of *those*, sets of *THOSE*, and so on, and then there is an infinite set, and sets of its elements, sets of those, sets of *those*, sets of *THOSE*, and so on, as far as possible; similarly for the case when individuals are admitted (forthcoming, p. 14).

This is essentially the account of the iterative conception given by Boolos (1989). The point is that the universe of sets is arranged in a particular way. We have the empty set $\emptyset$, and the singleton of the empty set $\{\emptyset\}$, and sets of those, and sets of all *those*, etc. Any talk of building, or priority, or dependence, is simply a metaphor for the hierarchical structure of the universe of sets.

There are several things to say about Incurvati's proposal. The first is that it cannot be an argument in favor of the view that the cumulative hierarchy is the universe of sets. If anything, it *assumes* that the set-theoretic universe is exhausted by the cumulative hierarchy, and gives us a way to describe it. Incurvati admits this, and gives a different argument in favor of this conclusion. The argument is that the minimalist conception is the most satisfactory according to certain virtues that we think a conception of set should have. Unfortunately, he does not say what those virtues are.

But what is more problematic is that Incurvati seems to think that we have some pre-theoretical intuition about the structure and extent of the universe of sets. We must, if we are to believe that the minimalist conception accurately describes it. He uses the natural number structure as an analogy. But the analogy does not quite fit. It is reasonable to think that we do have a pretty good idea of the structure of the natural numbers. We have something that we call the "standard model" of the

Peano Axioms. It is far from clear that we have a good idea of the full extent of the universe of sets. How, then, can we know that the minimalist conception accurately describes this universe?

But even on the minimalist conception, you could still argue that there is some kind of dependence going on. For given the minimalist conception, it still seems true that, for example, you can't have the singleton of the empty set $\{\emptyset\}$ without the empty set $\emptyset$. And it this conception of dependence, the modal/existential conception, that we argue holds between sets and their members. If you can't have sets without there members, as it seems you can't on the minimalist conception, then, in an important sense of dependence, sets metaphysically depend on their members.

I would like to conclude this section by returning to the Axiom of Foundation. I have argued against Incurvati that it is not clear that we have a full grasp of the universe of sets, and so we cannot know if the minimalist conception accurately captures the extent of this universe. Some would argue that we already know that anything like the minimalist conception, or the iterative conception, does not capture the full extent of the set-theoretic universe. The idea here is that the cumulative hierarchy only allows for well-founded sets. But we should allow non-well-founded sets as well. There are a couple of reasons for this. The first is that we know that ZFC set theory is equi-consistent with another theory that gets rid of Foundation and adds an Anti-Foundation Axiom. The theory ZFC^- is ZFC without the Axiom of Foundation. We then take ZFC^- and add an axiom, the Anti-Foundation Axiom or AFA, that allows for non-well-founded sets.[10] It has been shown that ZFC is consistent if, and only if, $\mathsf{ZFC}^- + \mathsf{AFA}$ is consistent. In fact, each theory is interpretable in the other. The second reason to believe in non-well-founded sets is that they are useful in modeling non-well-founded phenomena.[11]

So it may be that there are non-well-founded sets after all — self-membered sets, pairs of sets that are members of each other, sets with infinitely descending chains of membership. Can we have set-theoretic dependence in the context of non-well-founded sets? We could say that non-well-founded sets do not metaphysically depend on their members, but well-founded sets do. One might say this because one believes that the dependence relation must be irreflexive, asymmetric, and well-founded. I take it that the person who says this has good arguments to believe that the

[10]See Aczel (1988).
[11]See Barwise & Moss (1996).

dependence relation has these structural properties. If she does, then she can deny that non-well-founded sets depend on their members, and yet she can still maintain that well-founded sets do.

In Chapter 6, we will explore in more depth an alternative. There, we will argue for a minimalist conception of metaphysical dependence, based on the modal/existential conception, that allows for non-well-founded sets to depend on their members.

As set theory is often taken to be the foundation of mathematics, whether or not a set metaphysically depends on its members, and how we are to understand this claim, are important philosophical questions. We have argued here that there is good reason to think that sets depend on their members, as it follows from our basic conception of what a set is. We have also argued that the claim that sets metaphysically depend on their members follows from the generally accepted iterative conception of set.

Given that sets do metaphysically depend on their members, we desire a precise philosophical analysis of this dependence relation. We have already begun to look at candidates, namely, the modal analysis presented in Chapter 2. In the remainder of this chapter, we show why the modal analysis, and a similar analysis based on counterfactuals, both fail to accurately capture the asymmetric dependence between sets and their members. After showing that these analyses fail, we discuss why they fail, and take the first steps to developing a satisfactory analysis of set-theoretic dependence that maintains the modal/existential flavor.

2. The Modal Analysis and How it Fails

The modal analysis is a philosophically precise interpretation of the naive intuition that one thing metaphysically depends on another when you can't have the one without the other. We start by looking at three logically equivalent versions of the modal analysis.

MODAL ANALYSIS, V1. *x metaphysically depends on y if, and only if, there is no possible world w such that x exists in w and y does not exist in w.*

MODAL ANALYSIS, V2. *x metaphysically depends on y if, and only if, in every possible world w, if x exists in w, then y exists in w.*

MODAL ANALYSIS, V3. *x metaphysically depends on y if, and only if, in every possible world w, if y doesn't exist in w, then x doesn't exist in.*

On the modal analysis, dependence claims can be taken as straightforward claims about what possible worlds exist in the space of possibility.

There are several benefits to the modal analysis. The first benefit, as we have mentioned several times already, is that it maintains the historical intuition that metaphysical dependence has something to do with modality, with what is and what is not possible, and with existence. The modal analysis makes this connection very strong by giving truth conditions for metaphysical dependence in precisely these terms.

A second, perhaps more metaphysically and theoretically important benefit, is that it gives us a bit of conceptual economy. We have effectively reduced the concept of metaphysical dependence to the concept of existence. Claims about metaphysical dependence are simply claims about which possible worlds do or do not exist. This reduction is a two-step process. The naive analysis reduces metaphysical dependence to modality: dependence claims are claims about what can or cannot happen. The modal analysis then reduces claims about what can or cannot happen to claims about the existence of possible worlds.

The full reduction is successful only if one takes talk of possible worlds literally. If, for example, you take talk of possible worlds to be engaging in some kind of useful fiction, while denying that they really exist, then the second step of the reduction does not go through. But even in that case, we still have some degree of conceptual economy, as we can still allow the first step of the reduction from metaphysical dependence to modality. How one then understands modality may determine whether a modal analysis of metaphysical dependence succeeds in capturing dependence relations between sets.

Lastly, it is worth pointing out that if one is a realist about possible worlds, then the full reduction does not commit one to any particular ontology of possible worlds. We may take possible worlds to be concrete like the actual world, or we may take them to be abstract. The modal analysis does not rely on any particular view of possible worlds. We will discuss the relevance of the metaphysics of modality in the next chapter. But the modal analysis is neutral with respect to any assumptions about the nature of possible worlds, which can also be seen as a benefit.

Unfortunately, these benefits are outweighed by the fact that the modal analysis fails to accurately capture dependence relations in certain cases. It shouldn't be too hard to see what goes wrong. There are cases where one thing does not depend on another, even though there are no possible worlds where the one exists and the other does not. In other words, for some cases, the modal analysis does not give conditions that are sufficient. These cases often occur when either (1) at least one object in the relation exists necessarily, or (2) the objects bearing the relation exist in exactly the same possible worlds. These cases occur most often when one or both of the relata are abstract objects, like sets.

To illustrate this problem we look at simple sets of concrete objects. Consider the singleton of Socrates — the set whose sole member is Socrates. It is natural to think that the set {Socrates} metaphysically depends on the man Socrates. Remember, sets collect their members together. If a particular member doesn't exist, then it cannot be collected. The modal analysis says {Socrates} depends on Socrates because of some fact about the space of possible worlds: there is no possible world where {Socrates} exists but Socrates does not. So {Socrates} depends for its existence on Socrates.

It is also natural to suppose that while a set metaphysically depends on its members, members generally do not metaphysically depend on the sets they belong to. The dependence relation, at least in these cases, is asymmetric. Socrates the man does not depend on {Socrates} the set for its existence. Surely you can have the man without the set. But set-theoretic principles say, if you have the man, then you have the set. It follows that there is no possible world where Socrates exists but {Socrates} doesn't. According to the modal analysis, this fact about the space of possible worlds implies that Socrates metaphysically depends on {Socrates}. The modal analysis says that the dependence relation between a set and its members is symmetric, contrary to our intuitions about how the dependence relation works in these cases.

Symmetric dependence generalizes to any two things that exist in all of the same possible worlds. Any two necessary existents exist in all of the same possible worlds, because they both exist in every possible world. So the same kind of failure of the modal analysis will occur in cases where we are considering two things that exist necessarily, like pure sets. Take for example the empty set, $\emptyset$, and its singleton, $\{\emptyset\}$. There is no possible world where $\{\emptyset\}$ exists but $\emptyset$ doesn't, if only because there is no possible world where $\emptyset$ fails to exist. The modal analysis says that this fact about the space of possible worlds implies that $\{\emptyset\}$ depends on $\emptyset$, as it should. There is, however, also no possible world where $\emptyset$ exists but $\{\emptyset\}$ doesn't, because there is no possible world where $\{\emptyset\}$ fails to exist. The modal analysis says that this fact about the space of possible worlds implies that $\emptyset$ depends on $\{\emptyset\}$. And so again, it follows from the modal analysis that the metaphysical dependence between a set and its members, even when those sets are pure and exist necessarily, is symmetric.

In addition to the symmetric dependence between sets and their members, it also follows from the modal analysis that anything whatsoever metaphysically depends on objects that exist necessarily. Consider Socrates and the empty set. Certainly Socrates does not depend on $\emptyset$ for his existence. But there is no possible world where Socrates exists and $\emptyset$ does not, as $\emptyset$ exists in every possible world. It follows from the modal analysis that, contrary to our intuitions about dependence, Socrates the man metaphysically depends on the empty set. As Socrates was chosen arbitrarily in this example, the modal analysis implies that any object you choose metaphysically depends on the empty set. And given that the empty set was chosen somewhat arbitrarily as well (any necessarily existing object will do), the modal analysis implies that any object you choose metaphysically depends on every necessarily existing object.

The modal analysis is therefore problematic for anyone who believes in things that exist necessarily, whether they are pure sets or numbers or other abstract objects.

This is sometimes referred to as a problem of hyper-intensionality. There is no intensional distinction between things that exist in all of the same possible worlds. Possible worlds do not discriminate between Socrates and {Socrates}, or between $\emptyset$ and $\{\emptyset\}$. But the relation of metaphysical dependence does, as in both cases, one depends on the other, and not vice versa. So the dependence relation is fixed by a hyper-intensional distinction between two objects.

The problem of hyper-intensionality affects other plausible analyses that are motivated by the modal/existential conception of dependence. Consider a more nuanced counterfactual analysis of dependence.

COUNTERFACTUAL ANALYSIS. *x metaphysically depends on y if, and only if, had y not existed, then x wouldn't have existed either.*

It is standard to take a possible-worlds analysis of counterfactuals, such that a counterfactual "Had *A* occurred, then *B* would have occurred" is true if, and only if, every possible world that makes *A* true, and which satisfies certain further conditions, also makes *B* true. How those further conditions are specified is a matter of debate. It is standard to specify the conditions in terms of a notion of similarity between worlds. But one can also specify the conditions in other ways, for example, using *ceteris paribus* clauses. We explore these options in more depth in Chapter 6.

For now it doesn't matter. Because all of the cases that we have considered will encounter the same problems under the counterfactual analysis as they did under the modal analysis. Consider $\emptyset$ and $\{\emptyset\}$. For any extra conditions you choose, there will be no worlds that satisfy those conditions and make it that $\emptyset$ does not exist, because there are no worlds where $\emptyset$ does not exist. So $\{\emptyset\}$ will trivially depend on $\emptyset$. But symmetrically $\emptyset$ will trivially depend on $\{\emptyset\}$. And any object you choose will trivially depend on any necessarily existing object.

We can derive symmetric dependence between Socrates and {Socrates} under the counterfactual analysis as well. For any conditions you choose, consider the possible worlds that satisfy those conditions and which make it that Socrates does not exist. These worlds must also make it the case that {Socrates} doesn't exist either, for otherwise they would be impossible. Conversely, for any conditions you choose, consider the possible worlds that satisfy those conditions and which make it that {Socrates} does not exist. These worlds must also make it the case that Socrates doesn't exist either. For if Socrates existed but {Socrates} did not, these worlds would violate certain necessarily true set existence principles, which say that if you have the man, then you have the set as well. And so the worlds would be impossible, as they make a necessary truth false.

As you can see, sets cause significant problems for straightforward, natural analyses of metaphysical dependence that are motivated by the modal/existential conception. These analyses all deliver the counter-intuitive results that dependence between a set and its member is symmetric, and that everything depends on each necessarily existing object. These problems have persuaded many to abandon the project of analyzing metaphysical dependence in modal terms. It is our contention that abandoning these projects is premature. An account of dependence can be given, and it can be given in a way that preserves our modal intuitions about dependence.

3. An Alternative Modal Analysis

Neither the modal analysis nor the counterfactual analysis can accurately capture dependence relations, because they both say that some things depend on others when they really don't. But these analyses do have something going for them. They are appealing because they try to capture our intuitions that there is a conception of metaphysical dependence that is inherently modal. It's just that these analyses do a poor job of capturing these intuitions. That these particular analyses fail does not require us to reject every analysis of metaphysical dependence that is modal at heart.

What generates problems for the modal and counterfactual analyses is the possible worlds interpretation of modality. We simply do not have enough worlds to account for dependence relations between things that exist in all of the same possible worlds, including things that exist in every possible world. We need worlds where members exist without the sets that they belong to. And we need worlds where necessary existents fail to exist. But there are no such worlds. That is, there are no such *possible* worlds.

If there are no possible worlds that have what we need, then we must go beyond what is possible. We must turn to impossible worlds. The fact that a world is possible imposes a restriction on what can happen in that world. We should remove this restriction. Removing the limitation of possibility invites the impossible into our domain. This is good because there are impossible worlds where things exist without their singletons, and where necessary objects fail to exist. We can use these impossible worlds to fill in the gaps of the modal analysis.

The straightforward way to incorporate impossible worlds into the modal analysis is to allow the modal operators to range over all worlds, both possible and impossible.[12] By expanding our domain, we get what I call the revised modal analysis:

[12]We assume that those categories are both exclusive and exhaustive.

REVISED MODAL ANALYSIS. *x metaphysically depends on y if, and only if, there is no possible or impossible world, w, such that x exists in w and y does not exist in w.*

Unfortunately, this simple modification will not yield asymmetric dependence between sets and their members.

When we restrict the modal analysis to possible worlds, we get symmetric dependence between a set and its members. Allowing impossible worlds will avoid this particular symmetry, but it will also create a new problem. Consider the empty set and its singleton. There are surely impossible worlds where the empty set exists and its singleton does not. They are worlds where certain set existence principles, which say if you have something then you have its singleton as well, fail to hold. Because these principles are thought to hold in every possible world, any worlds where they fail to hold are impossible. As there are now worlds where $\emptyset$ exists but $\{\emptyset\}$ doesn't, the revised modal analysis says that $\emptyset$ does not depend on $\{\emptyset\}$, thus avoiding the major problem of the modal analysis.

But impossible worlds can be chaotic. There are impossible worlds where anything goes. And so there will be impossible worlds where the singleton of the empty set exists and the empty set does not. Given the existence of these worlds, the revised modal analysis says that $\{\emptyset\}$ does not depend on $\emptyset$. We have gone from symmetric dependence between a set and its members to no dependence at all.

This problem generalizes to any pair of distinct objects x and y. Even if x and y exist together in every possible world, there are nevertheless some impossible worlds where x exists but y doesn't, and some impossible worlds where y exists but x doesn't. In fact, the revised modal analysis implies that no two objects enter into dependence relations! The problem is that there are just too many impossible worlds. Considering them all at once is too much. So we need to find, for each instance of the dependence relation, some plausible way to restrict the impossible worlds under consideration. In each case, we need *some* impossible worlds but not *all* of them. Which impossible worlds we include will vary from case to case.

In Chapter 6, we introduce a variation on the counterfactual analysis that deploys impossible worlds in a manageable way. In doing so, we are able to look at a specific range of appropriate worlds that are relevant to the particular instance of dependence under examination. As impossible worlds are crucial to the analysis of metaphysical dependence that we endorse, we examine them in detail in Chapters 4 and 5.

It may seem that appealing to impossible worlds is an extreme response to the problems that sets cause for the modal and counterfactual analyses. Chapters 4 and 5 present some considerations that should ease this worry. But it is worth asking whether we should work so hard to preserve our intuitions that dependence is a modal notion? Perhaps some other, non-modal analysis would do better, without the added cost of impossible worlds. Another option is to take the relation of metaphysical dependence as primitive. Let's discuss these alternatives.

We consider two non-modal analyses of metaphysical dependence: one from Kit Fine (1995a), and one from Fabrice Correia (2005). Fine rejects the modal analysis of metaphysical dependence, specifically because the modal analysis gets it wrong in cases that involve sets and cases that involve necessary existents.[13] Fine instead proposes an analysis of metaphysical dependence in terms of the essence or nature or identity of an object.

> For we may take *x* to depend upon *y* if *y* is a constituent of a proposition that is true in virtue of the identity of *x*, or, alternatively, if *y* is a constituent of an essential property of *x* (1995a, p. 275).

For Fine, looking at the space of possibilities is the wrong way to understand dependence. To establish a dependence relation between two objects, one should examine their essential properties.

Fine's view is perhaps a natural one. The fact that one thing depends on another could be seen to be related to the essential properties of one or both of those things. But the primary motivation for Fine's view is the failure of analyses that are motivated by the modal/existential conception, like the modal and counterfactual analyses discussed in the previous section. If one can show that there is a successful analysis that maintains the intuitions of the modal/existential conception, it is then no longer necessary to introduce a new concept, especially one as poorly understood as essence.[14] Fine does not take the standard line on essence, where essential properties are articulated in terms

[13]What I call metaphysical dependence, Fine calls ontological dependence.

[14]Correia (2005) expresses a stronger version of this worry, claiming that one could be a complete skeptic about the notion of essence, and yet accept some notion of metaphysical dependence. This suggests that metaphysical dependence should not be analyzed in terms of essence. See p. 52.

of possible worlds: x has F essentially iff x has F in every possible world (in which x exists). For Fine, an essential property of an object is one that must occur in the definition of that object. Definitions take the form of propositions that are true in virtue of the identity of the object.[15] But here is where the analysis ends. Importantly, we are not told how we are to understand the relation of a proposition being true in virtue of the identity of an object. It is intended to be a relation between a proposition and an object, but the relation is left unanalyzed (Fine 1995a, p. 273).

Fine rejects the possible worlds account of essence for reasons similar to his rejection of the modal analysis. It is true that in every possible world in which Socrates exists, he has the property of being a member of {Socrates}. On this notion of essence, Socrates is essentially a member of his singleton, which seems wrong. Socrates is essentially a man; he is not essentially a member of any set. Concerning questions of essence and of dependence, Socrates has some undesirable connection with {Socrates}, either being essentially a member of {Socrates}, or being dependent on {Socrates}. There is reason to think, however, that if some alternative analysis can accurately capture dependence relations in modal terms, then it may be possible, and preferable, to capture essence in modal terms as well. Perhaps one can tell a plausible story about essences in terms of possible and impossible worlds.[16] But even if this kind of story does not work out for essences, the success of a new analysis of dependence, like the counterfactual analysis extended with impossible worlds that we present in Chapter 6, resolves any need to rely on the notion of an essential property, or the unanalyzed relation of being true in virtue of, in order to articulate metaphysical dependence.[17]

---

[15] See Fine (1994).

[16] Some work in this area has been done. Fabrice Correia (2007) offers an account of essence in modal terms by distinguishing between global and local possibility. Locally possible worlds allow for worlds where Socrates exists, but {Socrates} does not. Such worlds are locally, though not globally, possible. There are, however, no locally (or globally) possible worlds where {Socrates} exists but Socrates does not. Correia is thus able to achieve an asymmetry between Socrates and {Socrates}, and derives an account based on this asymmetry that yields the result that {Socrates} essentially contains Socrates, but Socrates is not essentially a member of {Socrates}. See also Fine's (2007) response to Correia. In order to achieve his result, Correia appeals to a notion of certain facts "being about" certain objects. As we do not appeal to such a notion to achieve an asymmetry between Socrates and {Socrates}, our accounts are quite different.

[17] E. J. Lowe (1998) offers an account of metaphysical dependence that also relies on the notion of essence. But he doesn't explain his notion of essence until Lowe (2008), where something's essence is to be understood as *what the thing is*, or *what it is to be that thing*. Unfortunately this understanding of essence is hardly illuminating.

Correia (2005) offers an alternative understanding of metaphysical dependence in terms of the notion of grounding, where grounding is a relation between facts.[18] Though Correia takes grounding as a primitive notion, he describes the relation as an explanatory one. The fact that {Socrates} exists is grounded in the fact that Socrates exists because the existence of Socrates explains the existence of {Socrates}, and not the other way around. Correia believes that the explanatory nature of the grounding relation is best elucidated by use of the term "in virtue of". Some fact A holds in virtue of the fact that B, the fact that $C, \ldots$. A fact may be either partially or completely grounded in other facts.

Correia then introduces the idea of one thing being *based on* another. "An object x is based on an object y when the fact that x exists is partly grounded in some fact about $y, \ldots$ when some feature of y helps explain the existence of x" (2005, p. 66). Correia then uses the idea of base to define metaphysically dependence:

CORREIA (2005). *x metaphysically depends on $y \stackrel{def}{=}$ necessarily, if x exists then x is based on y.*

That is, x depends on y means that in every possible world in which x exists, the fact that x exists is partly explained by some fact about y.

What is worrying about Correia's account of metaphysical dependence is that it makes explicit appeal to an idea that is notoriously hard to articulate: *explanation*. Correia recognizes this difficulty. He says, "Given that nowadays the vocabulary of explanation is so marred by non-objective connotations, it is preferable to avoid using it and prefer 'x exists in virtue of the fact that y exists' or 'y's existing makes x exist' to 'the existence of y explains the existence of x' " (p. 53). But changing the language in this way does not make the idea any more clear. How exactly should 'in virtue of' be understood? That is not to say that one cannot explain explanation. But Correia offers us no clues as to how we should understand this notion. Correia uses explanation as a panacea to solve all of the difficulties of metaphysical dependence. But without an account of explanation, this solution falls flat.

[18]What I call metaphysical dependence, Correia calls existential dependence.

Why then, must metaphysical dependence have any analysis? If every plausible analysis runs into problems, one might be tempted to give up and simply take the notion of metaphysical dependence as primitive — a concept that we generally understand, and which we might use to define other concepts, but which cannot be given a precise philosophical analysis.[19]

But our failure to articulate a natural account of metaphysical dependence, one which accurately captures dependence relations in all cases, does not imply that such an account cannot be given. In some respects, taking a concept to be primitive should be a last resort. It means we have tried everything we can think of, and nothing works. All of the candidate analyses of dependence have problems. Taking dependence as primitive implies these problems are insurmountable. My claim is that we have not exhausted all of our options.

Alternative, non-modal analyses trade in the notion of dependence for notions far more elusive. And taking dependence as primitive is not entirely satisfactory. But there are also positive reasons for endorsing a modal analysis. First, it works in almost all of the appropriate cases. It works in cases that involve concrete objects. And it works in many cases that involve abstract objects, or at least objects that are partially abstract, like impure sets. There are plausible set-theoretic examples where both the modal and the counterfactual analysis do give the proper, asymmetric dependence relations. Consider the pair set {Socrates, Plato}. Sets are taken to metaphysically depend on each of their members. So this set depends on Socrates, and it depends on Plato. The modal analysis, just with possible worlds, confirms this. Consider the following claims about the space of possible worlds.

> *There is no possible world w such that* $\{Socrates, Plato\}$ *exists in w and Socrates does not exist in w.*
>
> *There is no possible world w such that* $\{Socrates, Plato\}$ *exists in w and Plato does not exist in w.*

Both seem true. The set {Socrates, Plato} must be related to both Socrates and Plato in virtue of the inverse membership relation. But if either one doesn't exist, then nothing can bear this relationship to *both* Socrates and Plato. So nothing can be the set {Socrates, Plato}. So that set doesn't exist.

[19] Ross Cameron (2008) and Gideon Rosen (2010), for example, take metaphysical dependence as primitive.

One could give a similar argument using the counterfactual analysis, again with possible worlds only.

But consider the converse dependence relation. We do not want to say that either Socrates or Plato metaphysically depends on the set {Socrates, Plato}. The modal analysis, just with possible worlds, confirms this as well.

> *There is no possible world w such that Socrates exists in w and* $\{Socrates, Plato\}$ *does not exist in w.*
>
> *There is no possible world w such that Plato exists in w and* $\{Socrates, Plato\}$ *does not exist in w.*

Both seem false. In the first case, there could be a world where the pair set fails to exist because Plato doesn't exist, even though Socrates does. And analogously for the second conditional. Again, similar results hold under the counterfactual analysis.

Problems arise for the modal and counterfactual analyses only when the relata of the dependence relation exist in all of the same possible worlds. This happens in the case of singletons and their members, and when the relata are necessarily existing objects. But there are some set-theoretic cases, in addition to all of the concrete cases, where the modal and counterfactual analyses work perfectly well. So there is a positive reason to see what adjustments we can make to these analyses in order to get them to work in the problematic set-theoretic cases. The adjustment that we endorse is to appeal to impossible worlds in addition to possible worlds. This will not save the modal analysis, but it will give us a counterfactual analysis that can accurately capture the asymmetric dependence relations between sets and their members.

CHAPTER 4

Impossible Worlds

David Lewis once said, "had I not been such a commonsensical chap, I might be defending not only a plurality of possible worlds, but also a plurality of impossible worlds, whereof you speak truly by contradicting yourself" (1986, p. 1). Impossible worlds are certainly not as popular in the philosophical community as their possible counterparts. Some are still hesitant to embrace or even to discuss them. For impossible worlds are worlds where what couldn't possibly happen happens. Just as possible worlds are different ways the world could be, impossible worlds are different ways the world couldn't be.

Just as there are many ways in which something may be possible — physically possible, epistemically possible, ethically possible, metaphysically possible, logically possible — there are many ways in which something may be impossible. A physical impossibility is something that could not happen given the actual laws of physics. A logical impossibility is something that could not happen given the actual laws of logic.

What do I mean by the *actual* laws of logic? Well, the actual world is governed by logical laws. These laws tell us what inferences are valid and what inferences are not valid. We generally take laws of logic to sanction inferences that are valid under something like first-order classical logic. For example, we generally take the inference from $\neg\neg A$ to A to be a valid inference, for any first-order sentence A. Accordingly, there is a logical law that tells us that, for any A, if $\neg\neg A$ is true, then A is true. I take it that this is a logical law at the actual world. That means that, for any A, if $\neg\neg A$ is true at the actual world, then A is true at the actual world.

If we are to allow for logically impossible worlds, then the qualification "at the actual world" is essential. For there may be worlds w such that, for some A, $\neg\neg A$ is true at w and it is not the case that A is true at w. From the perspective of the actual world (assuming the actual world is classical), w is a logically impossible world. It is logically impossible because it does not have the same logical laws. Namely, the actual logical law that says *if* $\neg\neg A$ *is true, then* A *is true* is not a law at w. Of course, the world w may be governed by other logical laws. What we have said so far is consistent with w being governed by the laws of intuitionistic logic. The important point is that different worlds can have different laws of logic, just like different worlds can have different laws of physics. We can speak of the actual world's logical laws; we can speak of world w's logical laws. These are the laws of logic that hold at these particular worlds.

Given this connection between laws and possibility, we have a way to distinguish between logically possible and logically impossible worlds. Take the set L_α of logical laws of the actual world. We then have the following principle:

$$\textit{World } w \textit{ is logically possible iff } L_w = L_\alpha.$$

But we should be a bit more careful, as logical possibility should now be understood in a relative sense. Worlds are logically possible or impossible from the perspective of other worlds. From the perspective of the actual world, the world that does not sanction the inference from $\neg\neg A$ to A is logically impossible. And from the perspective of that world, the actual world is logically impossible, because it does sanction that inference. Generalizing, the space of worlds is divided into equivalence classes of worlds based on their logical laws.

There is much more to be said about logical laws, and we will return to them shortly. But we should recognize that the connection between laws and possibility is certainly evident in the literature on physical possibility. Consider Tim Maudlin (2007).

> A physicist who accepts Einstein's General Theory of Relativity will also believe that it is physically possible for a universe to be closed (to collapse in a Big Crunch) and possible for a universe to be open (continue expanding forever). This is especially evident since we don't yet know whether our own universe is open or closed, so empirical data are still needed to determine which possibility obtains. But even if the issue were settled, the laws of gravitation, as we understand them, admit both possibilities. Anyone who accepts Einstein's laws of gravitation, or Newton's laws of motion and gravitation, must admit the physical possibility of a solar system with six planets even if no such system actually exists. If one believes that the laws of nature governing some sort of event, say a coin flip, are irreducibly probabilistic (and that the outcomes of flips are independent of one another) then one must admit it to be physically possible for any sequence of heads and tails to result from a series of flips.

> I take these sorts of inference to be manifest in the actual practice of science and to be intuitively compelling. Any account of the nature of physical laws should account for them (pp. 7 – 8).

It is the laws of physics that determine what is physically possible. That is, it is the actual laws of physics that determine what is physically possible from the perspective of the actual world. According to these laws, for example, the entropy of isolated systems never decreases, and the speed of light is always constant. But there may be worlds that are governed according to other laws of physics. And in some of these worlds, perhaps there are isolated systems where entropy decreases, or perhaps the speed of light is not constant, or perhaps some other law of physics is violated. These worlds are physically impossible because their laws of physics do not match up with the physical laws of the actual world.

We can use the actual laws of physics to carve up the space of worlds into two groups: those that are physically possible and those that are physically impossible, from the perspective of the actual world. Take the set P_α of physical laws of the actual world. We then have that from the perspective of the actual world:

World w is physically possible iff $P_w = P_\alpha$.

This is easily generalized to any pairs of worlds to express their relative possibility with respect to one another. In some ways, the laws act to define an accessibility relation R on the set of worlds: given two worlds w_1, w_2 we have that $w_1 R w_2$ iff $P_{w_1} = P_{w_2}$. Clearly, this relation is an equivalence relation, being reflexive, symmetric, and transitive.

The fact that the physical laws divide the space of worlds into two groups, does not require that we know what the physical laws are. In fact, it is clear that we do not know what all of the laws of physics are. But we take for granted that there are laws of physics, whatever they may be. We then use these laws to pick out a class of worlds that we call the physically possible worlds.

Notice also that we do not arrive at these laws by looking to the physically possible worlds and asking what holds true at all of them. We take the laws to be prior, at least epistemically, and then define the physically possible worlds in terms of the laws. It is our contention that roughly the same model applies to both logical possibility and metaphysical possibility. Our main concern is metaphysical possibility, which we will discuss shortly, as it is metaphysical possibility, or rather metaphysical impossibility, that is most relevant to the counterfactual analysis of metaphysical dependence.

In each of these kinds of possibility — physical, metaphysical, logical – all we really require is that the possible worlds are the worlds with exactly the same laws. But we mentioned in passing that the laws seem to have an epistemic priority. We do not look at all of the, e.g., physically possible worlds, see what they have in common, and say that those are the physical laws. Rather, we take it that, at least in principle, we can discover the actual laws of physics just by looking at the actual world.[1] Once we discover the laws, we use them to distinguish between the physically possible and the physically impossible. Similarly for metaphysical and logical possibility.

We have physicists and philosophers of science who try to discover the laws of physics. Whatever they turn out to be, it seems reasonable to assume that they will be sentences of a certain kind, in the language of physics. Most of them will probably be universally quantified statements. And as we will see below, we take the metaphysical laws to be sentences as well. I would like to take a short digression to consider the laws of logic and show that they cannot simply be sentences of one kind or another.

1. Logical Laws

There are many familiar examples of laws of logic. These laws are usually captured by schemata.

EXCLUDED MIDDLE. $A \vee \neg A$

NON-CONTRADICTION. $\neg(A \wedge \neg A)$

IDENTITY. $A \rightarrow A$

It is thought that these schemata hold for any sentence A. Any sentence that has one of these forms must, by logical necessity, be true. Presumably there are other logical laws as well. Indeed, it is natural to take any instance of any logically valid formula to be a logical law. These instances will be sentences in a language. But these sentences do not exhaust the logical laws.

[1]We don't really have much choice.

Remember that we have been arguing for a connection between logical laws and logical possibility. From the perspective of the actual world, we have that:

$$\textit{World } w \textit{ is logically possible iff } L_w = L_\alpha.$$

But, assuming classical logic is the correct logic, suppose that we take the logical laws to be all instances of formulas that are valid under classical first-order logic **FOL**. These include sentences of the form $\neg(A \wedge \neg A)$, $A \ \vee \ \neg A$, and $A \rightarrow A$. Take the set of (all instances of) formulas that are valid under **FOL**. It turns out that this set matches precisely the set of (all instances of) formulas that are valid under the first-order logic of paradox **LP** (Priest 1979). That means a world governed by the laws of **LP** would have the same logical laws as a world governed by **FOL**, like the actual world. Given this match of logical laws, any world governed by the laws of **LP** would be logically possible from the perspective of the actual world, or any world governed by the laws of **FOL**. But a world governed by the laws of **LP** would allow some sentences to be both true and false, and surely that is not possible under the laws of **FOL**.

The issue, then, is to find a suitable characterization of

$$\varphi \text{ is a logical law at world } w.$$

The qualification 'at world w' is necessary, as some worlds will have different logical laws. What I propose, though it is just a suggestion, is to try to derive the logical laws of a world w from what is true at w. As we have just seen, we cannot focus only on instances of logically valid formulas. Instead we should appeal to the idea of a valid inference. This works best if we use multiple conclusion inferences. But to get the idea, we start with single conclusion inferences.

Single conclusion inferences say that if every member of a set Σ of sentences holds at w, then a sentence B holds at w. Given that $\Sigma = \{A_1, A_2, \ldots\}$, we can represent this inference as $A_1, A_2, \ldots \Rightarrow B$. For example, if w obeys the laws of classical logic, then it should be a valid inference of w that for any formula A, if $\neg\neg A$ holds at w, then A holds at w as well. Supposing that the actual world obeys the laws of classical logic, the inference $\neg\neg A \Rightarrow A$ should be valid at the actual world. The idea is that the inference from $\neg\neg A$ to A is valid because there is some logical law that says so. Even better, we could say that those valid inferences *are* the logical laws.

But which inferences count as logical laws? A tempting, though incorrect, account appeals to classical semantic entailment. It is a logical law that if $\neg\neg A$ holds at w, then A holds at w, because $\neg\neg A$ semantically entails A. And $\neg\neg A$ semantically entails A because every classical first-order model that satisfies $\neg\neg A$ also satisfies A.

But in characterizing the logical laws of w in terms of semantic entailment, we have not made any specific reference to what is true at w. So there is nothing about the worlds that would prevent this characterization of logical law from holding at every world. The point, though, is to give a characterization of logical law such that some different worlds have different logical laws. We want this because the claim is that what differentiates logically possible worlds from logically impossible worlds is that logically impossible worlds have different logical laws.

Instead of focusing on semantic entailment, we keep the idea of a logical inference and employ the idea of truth-preservation under uniform substitution.[2] We want the logical laws to consist of inferences. For the moment we are working with single conclusion inferences, so we take inferences to be ordered pairs $\langle\Sigma, B\rangle$, such that Σ is a set of sentences and B is a single sentence. We take the logical laws of a world w to be a set L_w of inferences (ordered pairs). The goal is to figure out exactly what ordered pairs belong in L_w.

The inferences in L_w should be those that preserve truth under uniform substitution of the non-logical signs of our language. Suppose our language is a first-order language with identity. We take the logical signs to be our usual connectives, the existential and universal quantifiers, and the identity sign. The non-logical signs include predicate signs and constant signs, where these are understood to be two different categories of non-logical signs. If we allow polyadic predicates, then predicates of different adicity are in different categories. A uniform substitution instance of a sentence is a sentence that is produced by uniformly replacing some of the non-logical signs of the original sentence with new non-logical signs of the same category. Uniformly substituting new signs for some of the non-logical signs in all of the sentences of a given inference $\langle\Sigma, B\rangle$ yields a new inference $\langle\Gamma, C\rangle$. Call $\langle\Gamma, C\rangle$ a uniform substitution instance of $\langle\Sigma, B\rangle$.

[2] The account offered here is inspired by Tarski's account of logical consequence as given in his 1936.

The inference $\langle\Sigma, B\rangle$ is a logical law of w iff every uniform substitution instance $\langle\Gamma, C\rangle$ of $\langle\Sigma, B\rangle$ is such that if every member of Γ is true at w, then C is true at w. In other words, the inference $\langle\Sigma, B\rangle$ is a logical law iff every one of its uniform substitution instances preserves truth at w.[3] The set L_w of logical laws of a world contains just those inferences whose uniform substitution instances preserve truth at w. The set L_α of logical laws of the actual world contains just those inferences whose uniform substitution instances preserve truth at the actual world. We then have that

$$\textit{World } w \textit{ is logically possible iff } L_w = L_\alpha.$$

Given this conception of logical law, we should make one final adjustment. There may be some worlds that obey the laws of a many-valued logic. Classical logic has two truth values: **t** for true and **f** for false. Other logics have more. The logic **LP**, for example, has three: **t**, **f**, and a third value **b** (which is usually interpreted as **both** true and false). Some worlds may be governed by the laws of **LP**. And **LP** allows for valid inferences between sentences that involve the third value **b**.

To accommodate these kinds of worlds, our characterization of logical law can focus, not on sentences that are true at a given world, but sentences that take what is called a designated value. Valid inferences are inferences that preserve designated values. The designated values of, e.g., **FOL** consist solely of **t**. The designated values of **LP** are **t** and the third value **b**.

The question of what counts as a designated value at a world is therefore crucial. Designated values are supposed to capture a more generic notion of truth. If the logic of a world is **LP**, the designated values of that world are **t** = 'true only' and **b** = 'both true and false', because any sentence that has one of these values is at least true.

The logical laws of a world should therefore be those inferences $\langle\Sigma, B\rangle$ where every uniform substitution instance $\langle\Gamma, C\rangle$ of $\langle\Sigma, B\rangle$ preserves designated values. That is, for every uniform substitution instance $\langle\Gamma, C\rangle$ of $\langle\Sigma, B\rangle$, if every member of Γ takes a designated value at w, then C takes a designated value at w as well.

However, an approach to logical law based on single conclusion inferences is not entirely satisfactory. It is not entirely satisfactory because it is not sufficient to rule out worlds that we

[3]For purposes of this paper, I take truth preservation to be the property that an inference $\langle\Sigma, B\rangle$ has when either A is true or one of Σ is not true.

would consider to be impossible. One can come up with worlds that should be logically impossible, but which obey all of the logical laws as we have characterized them here.

The objection is analogous to the original objection given earlier against the idea that logical laws are simply instances of classically valid formulas. The objection there was that there are different logics, like **FOL** and **LP**, that validate exactly the same formulas. Well, in fact, there are different logics that validate exactly the same single conclusion inferences as well. Two logics like this are **FOL** and the logic that extends the three-valued logic $\mathbf{K}_3$ with a supervaluational semantics.

Supervaluational semantics take logics that allow for non-classical truth values and use supervaluations to define logical consequence.[4] This is usually done for a three-valued logic. Take the three-valued logic $\mathbf{K}_3$, whose truth-values are **t**, **f**, and the third value **n** (which is usually interpreted as **neither** true nor false). Let v be a $\mathbf{K}_3$ valuation, and let $v \leq v'$ mean that v' is a classical valuation that agrees with v everywhere except where the valuation of a propositional variable p is the third value **n**. Where $v(p) = \mathbf{n}$, $v'(p)$ is either **t** or **f**. Truth values for the rest of the formulas under v' are then given recursively in the usual way. We can define the supervaluation v^+ of v for every formula A.

$$v^+(A) = \mathbf{t} \text{ iff for all } v' \text{ such that } v \leq v', v'(A) = \mathbf{t}$$

$$v^+(A) = \mathbf{f} \text{ iff for all } v' \text{ such that } v \leq v', v'(A) = \mathbf{f}$$

$$v^+(A) = \mathbf{n} \text{ otherwise}$$

In the first case, we say that A is "supertrue". Using supervaluations, an inference $\langle \Sigma, B \rangle$ is "supervalid" iff for every $\mathbf{K}_3$ interpretation v, if $v^+(A) = \mathbf{t}$ for every $A \in \Sigma$, then $v^+(B) = \mathbf{t}$ (that is, if every $A \in \Sigma$ is supertrue, then B is supertrue). It turns out that an inference $\langle \Sigma, B \rangle$ is supervalid iff B is a classical consequence of Σ. So a world w_1 governed by this supervaluational logic would validate precisely the same inferences as a world w_2 governed by classical logic. In this case, we could think of w_1 as having "supertrue" as a designated value. And so w_1 and w_2 would be possible with respect to each other, because they are governed by the same logical laws. But that can't be right, because a world governed by the supervaluational logic allows for sentences that are neither

[4]For more details see Priest (2008), ch. 7.

true nor false (because $\mathbf{K}_3$ allows for sentences that are neither true nor false), whereas a classical world does not.[5]

For this reason, it is more satisfactory to use multiple conclusion inferences to capture the laws of logic. A single conclusion inference is an ordered pair $\langle \Sigma, B \rangle$, where Σ is a set of sentences and B is a single sentence. A multiple conclusion inference $\langle \Sigma, \Delta \rangle$ replaces the single sentence B with a set Δ of sentences. Multiple conclusion inferences say that if every member of the set Σ holds, then some member of the set Δ holds. If $\Sigma = \{A_1, A_2, \ldots\}$ and $\Delta = \{B_1, B_2, \ldots\}$, then we can represent the inference $\langle \Sigma, \Delta \rangle$ as $A_1, A_2, \ldots \Rightarrow B_1, B_2, \ldots$.[6]

Multiple conclusion inferences can help with the problem described at the end of the previous section. For there are multiple conclusion inferences that are classically valid, but which are invalid according to the supervaluational interpretation of $\mathbf{K}_3$. For example, the inference $A \vee B \Rightarrow A, B$ is classically valid. If $A \vee B$ holds, then at least one of A or B holds as well. Not so in the supervaluational case. So we have a multiple conclusion inference that classical logic validates, but which this supervaluational logic does not. So any two worlds, where one is governed by the laws of **FOL** and the other is governed by the laws of this supervaluational logic would be impossible with respect to one another, because the logical laws of these worlds differ.

The switch to multiple conclusion inferences allows us to distinguish between these logics, and so between their logical laws as understood in terms of inferences that preserve designated values under uniform substitution. But there are a couple of final considerations that we should mention about this proposal for deriving the logical laws of a world from what is true at that world. First, the view of logical inference here is Tarskian, and there is a well-known objection to it. The objection is presented most clearly by John Etchemendy (1988). The objection is that this view over-generates, because it says that some sentences are logical truths when we would intuitively say they are not. Take, for example, the sentence that at least two things exist:

$$\exists x \exists y (x \neq y).$$

[5] Supervaluations are used to define validity (i.e., "supervalidity") by assigning classical truth values to sentences that take the value **n**. But in terms of the world being described, those sentences still take the value **n**.

[6] For details on multiple-conclusion inferences, see Shoesmith & Smiley (1978).

This sentence contains only logical signs. Since the sentence is true, every uniform substitution instance of it is true (because there is only one uniform substitution instance: the trivial one). And so the inference from the empty set to the formula $\exists x \exists y(x \neq y)$ preserves truth under uniform substitution. On the account presented here, it is thus a logical law. Accordingly, the formula $\exists x \exists y(x \neq y)$ is then true at every logically possible world. But it does not seem to be a law of logic that at least two things exist. Plausibly, it could have been that only one thing existed, or perhaps nothing at all. At the very least, it is plausible that there could have been fewer things than there actually are. But on the Tarskian view, for however many things there actually are, it is a logical truth that there are that many things.

This is a problem for the Tarskian view of logical consequence, and so it is a problem for the view of logical law presented here. I mention two possible responses, one by Mario Gómez-Torrente, and one by Graham Priest.

Gómez-Torrente (1996, 2009) argues that we should properly interpret Tarski as requiring that our language have a nonlogical monadic predicate symbol to name the domain of quantification of the intended interpretation, and that any alternative interpretation should use the same predicate symbol to name the alternative domain under consideration. If we take that monadic predicate symbol to be N, we can say that the formula $\exists x \exists y(x \neq y)$ is an abbreviation for the formula $\exists x \exists y[N(x) \wedge N(y) \wedge x \neq y]$. Given that the extension of the predicate N may vary, the domain of quantification may vary. It may be that under a particular interpretation, only one thing satisfies the predicate N, in which case the formula $\exists x \exists y[N(x) \wedge N(y) \wedge x \neq y]$ would not be true. There are, therefore, uniform substitution instances of this formula that are not true, and so it would not be a logical law on the account presented here.

Graham Priest (1995) proposes a different solution to the problem of over-generation. One can suppose that the quantifiers of our language don't just range over the things that exist at the actual world. Rather, they range over *everything*. This includes concrete objects and abstract objects. This includes actual objects, and possible objects, and maybe impossible objects too. Given that we are counting all objects, the question of how many objects there are is no contingent matter. The number of things does not vary from world to world. And so, if we are really counting everything, then it could not fail to be that there are at least two things.

The second consideration worth mentioning is that a consequence of the view presented here is that if a world "looks" classical, then it is classical. It is well known that certain classical valuations are compatible with, e.g., intuitionistic logic. One might think that a world could be totally classical, making every instance of $A \vee \neg A$ true, and yet be governed by the laws of intuitionistic logic. That is, though every instance of $A \vee \neg A$ is true, the inference from the empty set of formulas to the set $\{A \vee \neg A\}$ is not a logical law. In the face of this consideration, I am prepared to bite the bullet and say that if a world looks classical, then it is classical. A world that makes every instance of $A \vee \neg A$ true is a world that has excluded middle as a law of logic.

For it seems plausible to me that the actual logical laws should be determined by what is true at the actual world. And the logical laws of any world should be determined by what is true, or what takes a designated value at that world. Some worlds will match ours with respect to their logical laws. They are the logically possible worlds. And some worlds will not match ours with respect to their logical laws. They are, therefore, logically impossible.

2. Metaphysical Laws

In exploring the connection between possibility and laws we have looked at two kinds of possibility: physical possibility and logical possibility. But there is also metaphysical possibility. Metaphysical possibility, if it is not simply identified with logical possibility, is sometimes thought to lie somewhere between physical and logical possibility. To say that it is between these two is to assume that there is an ordering on kinds of possibility. We will proceed with this assumption for now.

The idea is that more things are metaphysically possible than are physically possible. It is metaphysically possible, though not physically possible, for the speed of light not to be constant. More things are logically possible than are metaphysically possible. It is logically possible, though not metaphysically possible, for something to be both red and not colored. That is, it is not a law of logic that if something is red, then it is colored. A situation in which something is both red and not colored would still obey the laws of logic.[7] Given this ordering, something logically impossible would be both metaphysically and physically impossible. Something metaphysically impossible would be physically impossible, though it need not be logically impossible. And something physically impossible need not be logically or metaphysically impossible.

If we were to think of this ordering in terms of possible worlds, we would order sets of possible worlds as follows:

Physically possible worlds $\subset$ Metaphysically possible worlds $\subset$ Logically possible worlds

One might also try to fit other modalities into this ordering. For example, something ethically necessary (required) must be physically (and so logically and metaphysically) possible. And something ethically possible (permissible) should be physically possible as well. On the other side, what is epistemically possible may transcend the laws of logic. It is sometimes claimed that we can conceive of contradictions being true. We can certainly entertain what would happen had they been true. One can even believe a contradiction to be true. Consider Graham Priest's beliefs concerning the Russell set (2006).

> I, for example, believe that the Russell set is both a member of itself and not a member of itself. I do not deny that it was difficult to convince myself of this, that is, to get myself to believe it. It seemed, after all, so unlikely. … It is difficult to come to believe something that goes against everything that you have ever been taught or accepted, in logic and philosophy as elsewhere. This is just a psychological fact about the power of received views on the human mind. … In fact, the number of philosophers who have consciously believed explicit contradictions is much larger than the contemporary teaching of philosophy would lead one to expect—there are, to name but a few, Heracleitus, Plotinus, Nicholas of Cusa, Hume, Hegel, and Engels (p. 97 – 8).

Such claims suggest that contradictions, which violate the laws of (classical) logic, are epistemically possible. Some things, then, may be epistemically possible for some people, but at the same time logically impossible. However, it's not entirely clear that this thought leads to a clear ordering that includes epistemic possibility. First, epistemic possibilities are relativized to an agent, and so will be different for different agents. Second, there are many things that are logically possible, but

[7] In the sense of logical law given in Section 1 of this chapter. That is, given a language with the predicates *Red* and *Colored*, the inference from $\forall x Red(x)$ to $\forall x Colored(x)$ does not preserve truth under uniform substitution of the non-logical symbols, assuming that *Red* and *Colored* are non-logical symbols.

which are inconsistent with what a particular agent may believe, for example that the agent has two heads.

There is then the question of what may be called mathematical possibility. The standard line is that mathematical statements are necessarily true if true, necessarily false if false. There are no mathematical contingencies. The objects of mathematics, whether they are numbers, or sets, or groups, or categories, or some combination of these and other things, exist necessarily, if they exist at all.

If I had to fit mathematical possibility into this ordering, I would take it to be a variety of metaphysical possibility. For one thing, mathematical truths are not logical necessities, following from the laws of logic.[8] Nor are they physical necessities, following from the laws of physics. Nor are they ethical or epistemological necessities. The most natural fit is to group them with the metaphysical necessities. In which case, they follow from the laws of metaphysics.

What then are the laws of metaphysics? I do not have an original proposal as to what the metaphysical laws are. Whatever they are, I think that the distinction between metaphysical law and metaphysical truth is of little use. We distinguish physical laws from physical truths, and this distinction is important. Physical truths describe the physical properties of particular objects, objects that exist contingently. Physical laws are general principles that apply to all existing physical objects. Physically possible worlds have the same physical laws. They do not necessarily have the same physical truths. On the other hand, I take it that even if there was a plausible distinction between metaphysical laws and metaphysical truths, both laws and truths would hold in every metaphysically possible world. So I will only speak in terms of metaphysical truths going forward.

I take the metaphysical truths to include analytic truths. It is a metaphysical truth that all bachelors are unmarried. I take them to include truths about grounding, if there are any. If it is true that parts ground their wholes, then that is a metaphysical truth. I take them to include truths about essence, if there are any. If it is true that x is essentially F, then that is a metaphysical truth. And I take them to include mathematical truths. It is a metaphysical truth that $2+2=4$. If the Axiom

[8] Unless you are a logicist.

of Choice is true, then it is a metaphysical truth; if the Continuum Hypothesis is true, then it is a metaphysical truth. If it is true that the empty set exists, then that is a metaphysical truth.

Continuing the pattern established by physical possibility, and applied to logical possibility, we can take the set of metaphysical truths at the actual world M_α. We then have the following principle:

$$\textit{World } w \textit{ is metaphysically possible iff } M_w = M_\alpha.$$

Given this connection between metaphysical laws and metaphysical possibility, it makes sense to count mathematical truths as metaphysical truths. It is common to take mathematical truths, like "$2+2=4$" or "the empty set exists" to be true in every metaphysically possible world.

We should note that what makes something a metaphysical truth is a separate question. It could be that the mere fact of being true in every metaphysically possible world is what makes something a law of metaphysics. On this view, it is an objective fact that some worlds are metaphysically possible. The laws of metaphysics are then the sentences that happen to be true in all of these worlds. On the other hand, something else entirely may determine the laws of metaphysics. It may even be linguistic convention. We could then use these laws to pick out the metaphysically possible worlds. The only claim that we make here is that the laws of metaphysics match up with those sentences that are true in every metaphysically possible world.

Furthermore, just as in the physical case, we need not know what the metaphysical truths are. All we have to do is assume that there are some metaphysical truths. Given that some statements are metaphysically true, we can use those truths to divide the space of worlds into two groups: the metaphysically possible worlds and the metaphysically impossible worlds. Metaphysically impossible worlds are worlds that break the laws of metaphysics.

We are interested in the metaphysically impossible worlds, as they are particularly relevant to the success of the counterfactual analysis of metaphysical dependence. Given that we are interested in metaphysical dependence between mathematical objects, like sets, we are particularly interested in worlds where certain mathematical objects, like the empty set, fail to exist. We will also consider worlds where other mathematical objects, like numbers and number structures fail to exist. I take statements like "$\emptyset$ exists", "the number 9 exists", and "the natural number structure $\mathscr{N}$ exists", to be metaphysical truths, if they are truths at all. Supposing they are true, worlds where they fail to be true are metaphysically impossible.

We have given a characterization of possibility in terms of laws. This characterization is to some degree general because it applies to many important kinds of possibility: physical, metaphysical, and logical. I would like briefly to note that under this characterization, it is not clear that these kinds of possibility are ordered in any particularly obvious way. In terms of worlds, consider the ordering we suggested earlier.

Physically possible worlds $\subset$ Metaphysically possible worlds $\subset$ Logically possible worlds

This ordering is not entailed by the connection between laws and possibility endorsed here. It's just not obvious that if something is logically impossible that it must be metaphysically impossible as well. Just because a world breaks a law of logic does not entail that it breaks a law of metaphysics as well. At the very least, we need an additional argument in favor of this ordering, or any ordering for that matter. I do not offer any argument, and I am happy to leave the various kinds of possibility unordered.

3. Worlds vs. Points

We should say more about worlds in general, both possible and impossible. There are two main ways in which worlds are used in contemporary philosophy. One has to do with metaphysics; the other has to do with logic. In metaphysics, worlds allow us to give a reductive analysis of modal notions, like possibility and necessity. Many metaphysicists are unwilling to accept facts about what is possible or what is necessary to be primitive facts about the world. They cry out for explanation. A reductive analysis of these notions seeks to provide such an explanation. The most common reductive analysis of possibility and necessity are given by the following biconditionals.

(P). *It is possible that A iff there is a possible world w such that, at w, A.*

(N). *It is necessary that A iff every possible world w is such that, at w, A.*

One may worry that the reduction fails because of the word 'possible' on the right-hand side of the biconditionals. But one need not worry. First, as noted above there are different kinds of possibility; we are most interested in physical, metaphysical, and logical possibility. Taking physical possibility as an example, we can reformulate these biconditionals as follows.

(PP). *It is physically possible that A iff there is a physically possible world w such that, at w, A.*

(PN). *It is physically necessary that A iff every physically possible world w is such that, at w, A.*

Though we have not yet eliminated the word "possible" on the right-hand side, this is not hard to do, as we have connected physically possible worlds with physical laws. We should also recall that possibility, whether physical, metaphysical, or logical, is relative to the world you start from. We formulate these reductions from the perspective of the actual world.

(PP). *It is physically possible, at the actual world, that A iff there is a world w such that* $P_w = P_\alpha$ *and at w, A.*

(PN). *It is physically necessary, at the actual world, that A iff for every world w, if* $P_w = P_\alpha$*, then at w, A.*

We have taken physical possibility as an example. These principles are easily extended to the cases of metaphysical and logical possibility.

The reductive analysis of possibility and necessity takes modal claims to be claims about what holds at certain worlds. If "*A* is physically possible" is to be true at the actual world, then there must exist some world that has the same physical laws at the actual world, such that *A* is true at that world.

We have not yet said what these worlds are. They may be sets of propositions, or states of affairs, or spatiotemporally isolated concrete worlds in a Lewisian hyperuniverse. Arguments can be presented in favor of each of these views. For us it doesn't really matter what you take worlds to be. But whatever they are, according to the reductive analysis of modality, they must exist.

You need not accept the reductive analysis of modality presented here. But it is a view that one can have, and it assumes the existence of possible worlds, whatever they may be. We show in the next section that, once you have an ontology that includes possible worlds, adding impossible worlds does no harm, and can be quite useful. Of course, if you don't accept the existence of possible worlds, then you probably won't accept the existence of impossible worlds either. We discuss that alternative in the last section of this chapter.

So much for the metaphysical uses of possible worlds. Possible worlds, or things under that name, play a role in logic as well. Specifically, worlds are used to give the semantics of many intensional notions, like possibility and necessity, and knowledge and belief. The semantics for modal logic, the kind of logic used to examine notions like possibility and necessity, makes use of Kripke frames. Taking modal propositional logic as an example, Kripke frames are ordered pairs $\langle W, R \rangle$, where W is a set of points and R is a relation, called an *accessibility* relation on W. We can extend Kripke frames to Kripke interpretations by adding a valuation function that maps point-proposition pairs to truth-values.[9]

The members of W, which I have called points, are often referred to as possible worlds. Because modal logic, at least in part, is supposed to model how possibility and necessity work, these points are supposed to represent how things are in different possible worlds. But the Kripke interpretations are just mathematical constructions, and for any Kripke interpretations, the points in W are not possible worlds in the metaphysical sense. They are not concrete worlds, or sets of propositions, or states of affairs. So I will reserve the term 'point' to refer to members of W in a Kripke frame or interpretation.[10] The term 'possible world' is used to refer to the things that we appeal to in order to give a reductive analysis of possibility and necessity.

There may be, however, a particular Kripke interpretation such that points in W can be matched up with the metaphysically possible worlds, and such that the accessibility relation respects the correct relations of relative possibility between worlds. The interpretation may also accurately match point-proposition pairs to truth values (i.e., if A is true at a metaphysically possible world, then the valuation function takes the pair that consists of the point that represents this possible world and the proposition A to the truth-value 'true'). In effect, this particular interpretation (perhaps

[9]It is more common to call these Kripke models, but as we are not currently talking of these structures as modeling any particular set of sentences, we will call them interpretations.

[10]We will come back to Kripke frames and interpretations, as well as other kinds of interpretations that use points in Chapter 5.

we could call it the "standard" interpretation) would model how possibility and necessity work according to the actual world.

So these are two of the main uses of worlds in contemporary metaphysics and logic. When we talk about worlds, we are talking metaphysics; when we talk about points, we are talking logic. The metaphysically possible worlds have particular importance for the counterfactual analysis of metaphysical dependence. So let's talk metaphysics for a bit, and return to logic in Chapter 5.

4. Realist Theories of Impossible Worlds

We have characterized metaphysically impossible worlds as worlds that break the (actual) laws of metaphysics. But we have not yet argued that there are any metaphysically impossible worlds. The question of their existence is a genuine metaphysical question, in many ways similar to the question of the existence of possible worlds. Our position is that once you admit the existence of metaphysically possible worlds, you might as well admit the existence of metaphysically impossible worlds too. And I think this is true for logically impossible worlds as well. Why? For reasons very similar to those behind arguments in favor of possible worlds. Just as there are different ways the world could be, there are different ways the world couldn't be. Just as possible worlds are useful in many different ways, impossible worlds are useful in many different ways. Adding metaphysically and logically impossible worlds not only solves some of the problems that plague theories of possible worlds, they do so at very little additional cost.

But we reserve positive arguments in favor of impossible worlds for the next chapter. In the remainder of this chapter we show that one can extend mainstream theories of possible worlds to include impossible worlds, and that this extension does not cause any serious problems. When it comes to possible worlds there are two general approaches one may taken: realist or anti-realist. Realists agree that possible worlds exist. But they disagree as to the kinds of things that possible worlds are. We show that impossible worlds can be added to many realist theories without too many problems. On the other hand, anti-realists agree that possible worlds do not exist, and if they do not exist, then they are not any kind of thing. But some anti-realists do claim to be able to make use of possible worlds. If this move is allowed, then we should also be able to use impossible worlds without admitting their existence either. We start with realist theories.

4.1. Extreme Realism. By extreme realism about possible worlds I mean the theory associated most commonly with David Lewis.[11] On Lewis's view, metaphysically possible worlds are maximal mereological sums of individuals that are spatiotemporally related.[12] The worlds themselves are each spatiotemporally isolated from each other. They are concrete entities just like the actual world. These worlds are truthmakers for modal truths, truths about what is possible and what is necessary, by way of the following biconditionals:

P_L. *It is metaphysically possible that A iff there is a world w such that, at w, A.*

N_L. *It is metaphysically necessary that A iff every world w is such that, at w, A.*

Notice that these biconditionals are different from those considered earlier in this chapter. First, they are not relativized, either to any particular world, or to any particular kind of necessity. I take it that most agree that Lewis was talking about metaphysical possibility. Unless otherwise stated, talk of possibility and impossibility should be understood as metaphysical possibility and impossibility. I also take it that Lewis was interested in what is possible and necessary from the perspective of the actual world, but that he also took each world to be possible from the perspective of every other world. And so these conditionals hold at every world. For simplicity, we will usually speak of what is possible and impossible from the perspective of the actual world.

More importantly, the word 'possible' does not appear on the right-hand side of either biconditional. Indeed, there are no modal notions on the right-hand side. These biconditionals represent fully reductive analyses of possibility and necessity to non-modal notions.

We explore the consequences of adding concrete impossible worlds to the ontology of extreme realism. The most obvious problem is that we lose the reductive power of the extreme realist's view.[13] Lewis's theory of possible worlds represents the most comprehensive attempt at a reductive

[11]Most notably Lewis (1986).

[12]They are maximal in the sense that they do not leave out any individuals that are spatiotemporally related to individuals that are a part of the sum. Any sum with parts that are spatiotemporally related to things that are not parts is not maximal, and therefore is not a world.

[13]This worry is expressed by John Divers (2002), p. 69.

theory of modality. Examining the principles P_L and N_L we notice that neither one makes use of modal language on the right-hand side. Both possibility and necessity have been reduced to facts about the existence of worlds. What is possible and what is necessary is determined solely by what exists (in the sense of existence that allows for non-actual things to exist).

Introducing worlds that are impossible is problematic for Lewis. For now more things exist than are possible. That A is impossible implies that A is true at some (impossible) world. This impossible world is a world that exists. So A is true at a world that exists. But the existence of a world where impossible A is true implies, by way of P_L, that A is possible. So, if A is impossible, then A is possible. As this cannot be right, the right-to-left half of P_L must be false (though the other half remains true).

Since we now have more worlds than are possible, to differentiate the possible from the impossible, we need some way to capture all and only the metaphysically possible worlds. We can do this by appealing to metaphysical laws. Take the set M_α of metaphysical laws, or equivalently the set of metaphysical truths, at the actual world. We then have that:

World w is metaphysically possible iff $M_w = M_\alpha$.

We can combine this understanding of a metaphysically possible world with Lewis's conditions to get:

P_L^*. *It is metaphysically possible that A iff there is a world w such that $M_w = M_\alpha$ and at w, A.*

N_L^*. *It is metaphysically necessary that A iff for every world w, if $M_w = M_\alpha$, then at w, A.*

This reformulation of Lewis's principles is just as reductive, as long as we can give an accurate characterization of the metaphysical truths without appealing to what is metaphysically possible or to metaphysically possible worlds. In addition, this formulation maintains the natural understanding of "impossible" as applied to sentences as being not possible. Following from these conditions we get:

I_L^*. *It is impossible that A iff there is no world w such that $M_w = M_\alpha$ and at w, A.*

To be impossible is to fail at every possible world. Of course, A's being impossible according to this condition does not yet guarantee that there is a world, impossible as it may be, where A does hold. However, if we are to embrace impossible worlds, I am not inclined to be shy. I suggest that for any A, there is a world where A holds. And for (almost) any distinct A and B, there is a world where A holds and B does not.[14]

Given a non-modal account of metaphysical laws, extreme realism about possible and impossible worlds rises up to the challenge of providing a reductive theory of modality. But other objections, which target the coherence of an extreme realist theory of impossible worlds, remain to be answered. We consider three such objections, the first of which comes from Lewis himself. For Lewis, the term 'at world w' acts to restrict quantifiers to the things in world w. But, as a restriction on quantifiers, the term should not affect how the logical connectives work. If the logical connectives, in particular negation, work classically, then Lewis's worry is that contradictions that hold at impossible worlds imply contradictions that are true in a more general sense. Lewis takes this to be justification for rejecting impossible worlds.

> [S]uppose travellers told of a place in this world – a marvellous mountain, far away in the bush – where contradictions are true. Allegedly we have truths of the form 'On the mountain both P and not P'. But if 'on the mountain' is a restricting modifier, which works by limiting domains of implicit and explicit quantification to a certain part of all that there is, then it has no effect on the truth-functional connectives. Then the order of modifier and connectives makes no difference. So 'On the mountain both P and Q' is equivalent to 'On the mountain P, and on the mountain Q'; likewise 'On the mountain not P' is equivalent to 'Not: on the mountain P'; putting these together, the alleged truth 'On the mountain both P and not P' is equivalent to the overt contradiction 'On the mountain P, and not: on the mountain P'. That is, there is no difference between a contradiction within the scope of the modifier and a plain contradiction that has the modifier within it. So to tell the alleged truth about the marvellously contradictory things that happen on the mountain is no different from contradiction yourself. But there is no subject matter, however marvellous, about which you can tell the truth by contradicting yourself. Therefore there is no mountain where contradictions are true (1986, p. 7, n. 3).[15]

We start by reconstructing Lewis's argument:

[14]The hedge is due to sentences like $\forall xPx$ and $\forall yPy$. These sentences are distinct, but given that P is interpreted the same way in each case, they should both hold at precisely the same worlds.
[15]See also Lewis (1983b).

(1) Suppose there is a world w such that P and $\neg P$ is true at w.

(2) From 1, P is true at w, and $\neg P$ is true at w.

(3) From 2, P is true at w.

(4) From 2, $\neg P$ is true at w.

(5) From 4, Not: P is true at w.

(6) From 3 and 5, P is true at w and not: P is true at w.

Before evaluating this argument, it is important to note that it is not an argument against the existence of all metaphysically impossible worlds. Lewis's argument is against the existence of worlds in which some contradictions are true. Worlds in which some contradictions are true violate the laws of (classical) logic. They need not violate the laws of metaphysics, given that we reject the popular ordering on kinds of possibility as we have done. I maintain, then, that Lewis (and anyone else) can think this argument sound and still admit the existence of metaphysically impossible worlds.[16] In fact, Lewis's argument is not even an argument against all logically impossible worlds. Worlds governed by the laws of intuitionistic logic are logically impossible worlds (as there are classical inferences that fail to preserve truth under uniform substitution), but they do not allow for true contradictions. Lewis's argument has no consequences for the existence of these logically impossible worlds either. However, Lewis takes his argument to be an argument against impossible worlds generally, as he makes no distinction between metaphysically and logically impossible worlds. So we examine this argument, and propose a response on behalf of those who endorse logically impossible worlds.

Lewis's argument makes use of the notion of truth at a world. But Lewis also endorses a more general notion of truth, truth *simpliciter*, which I will call *Truth*. *Truth* is not truth at any specific world; in particular *Truth* is not just truth at the actual world. Truths at the actual world are sentences with their quantifiers restricted to things in the actual world. Truths at world w are sentences with their quantifiers restricted to things in world w. *Truths* are sentences with unrestricted quantifiers. Many sentences with unrestricted quantifiers, such as 'talking donkeys exist' come out *True*, even though we would in normal circumstances judge them false. In those

[16] Even with the standard ordering on possible worlds, as described above, one could reject the existence of logically impossible worlds, and still admit the existence of metaphysically impossible worlds.

circumstances, we apply an implicit restriction on the domain of quantification to include only things that exist at the actual world. But strictly speaking the sentence is *True* because there are worlds that have talking donkeys. Of course, the actual world is not such a world. So the sentence 'At the actual world, talking donkeys exist' is false. The phrase 'At the actual world' restricts the existential quantifier to range over things in the actual world. And none of those things is a talking donkey.

What Lewis's argument tries to show is that if some contradiction is true at a world, then some (though not the same) contradiction is *True*. We can rewrite the argument by appealing to a relation $\Vdash$ between sentences and worlds. Intuitively, let $w \Vdash P$ say that P is true at world w. In Lewis's terms, $w \Vdash P$ if and only if P is *True*, as long as the quantifiers in P are restricted to range over w.

The argument reconstructed with these tools runs as follows:

(1) Suppose there is a world w such that $w \Vdash P \wedge \neg P$.

(2) From 1, $w \Vdash P$ and $w \Vdash \neg P$.

(3) From 2, $w \Vdash P$.

(4) From 2, $w \Vdash \neg P$.

(5) From 4, $w \nVdash P$.

(6) From 3 and 5, $w \Vdash P$ and $w \nVdash P$.

The argument is supposed to be given in the language of *Truth*. The formula in line (6) is of the form $\varphi \wedge \neg\varphi$, representing a contradiction that is *True*. Note that this argument is not an argument to the conclusion that some contradiction is true at the actual world. But Lewis thinks it absurd enough that some contradictions are *True*. It would be easy to understand Lewis's objections to contradictions being true at the actual world. Granting that truth at the actual world is governed by the laws of classical logic, contradictions at the actual world would explode, implying that every sentence is actually true. And we wouldn't want that. It is not clear, however, that the logic of w must be governed by the laws of classical logic, especially if w is an impossible world. It may be that the logic of w is such that negation does not work classically. It may, for example, work according to a logic that allows P and $\neg P$ to both be true (at that world). In which case, the inference from (4) to (5) is unwarranted. That $\neg P$ is true at w does not entail that P is not true at w. For it may be that they are both true at w. Given that the logic of w may be non-classical, there is no guarantee that the other connectives behave classically either. And so, for example, the inference from (1) to (2) may be invalid as well.

Lewis rejects this line of reasoning, as he uses the modifier 'at world w, P' simply to restrict the quantifiers in P to world w, where P is expressed in the language of *Truth*. It is like the phrase 'on the mountain, P', or 'in the fridge, P'. As it only serves to restrict quantifiers to things in world w, it should not affect how the connectives work. So, we should be able to move from '$\neg P$ is true at w' to 'It is not the case that P is true at w'. But it's not clear that Lewis is correct on this point.

Consider the key passage of the Lewis quote.

> But if 'on the mountain' is a restricting modifier, which works by limiting the domains of implicit and explicit quantification to a certain part of all that there is, then it has no effect on the truth-functional connectives. Then the order of modifier and connectives makes no difference (Ibid.).

However, it can be the case that 'at world w' acts to restrict quantification, and that it has no effect on the connectives (we can even assume the connectives are truth-functional, though they may not be at some non-classical worlds), and yet the order of modifier and connectives can matter.[17] Consider a world w that works according to the three-valued logic **LP**. Suppose that for some object a in the domain of w, the formula Pa takes the third value i. As the third value i is interpreted as 'both true and false', it follows that Pa is true at w. But according to the truth-functional negation of **LP**, the truth value of $\neg Pa$ is also i. It follows that $\neg Pa$ is also true at w. But we cannot conclude from this that Pa is not true at w. For we have already seen that Pa is true at w. And so, even if 'at world w' is simply a restriction on the domain of quantification, and if such restrictions have no effect on the connectives, this does not guarantee that one can switch the order of modifier and connectives.

So the order of modifier and connectives does matter for some worlds that do not obey classical logic. To assume that it does not matter begs the question against someone who thinks these worlds exist. And since the claim that it does not matter does not follow from Lewis's assumptions, he has not given us reason to think that these impossible worlds do not exist.

[17] I take it that what Lewis means when he says 'at world w' has no effect on the connectives, he means that however the connectives work, the phrase 'at world w' does not force the connectives to work differently from that.

We turn now to another worry, developed by Takashi Yagisawa (1988), which involves the identity of properties. One thing that impossible worlds allow us to do is distinguish between properties that are necessarily coextensive. One way to distinguish between properties generally is to say that property P is not identical to property Q iff P and Q have different extensions. By the extension of a property, we mean the extension of the property across all worlds, where a property's extension across all worlds is the union of the property's extensions in each world.[18] If we have only possible worlds, then we cannot use this method to distinguish between properties that are necessarily coextensive, like *being triangular* and *being trilateral*.[19] In each possible world, everything that is triangular is trilateral, and vice versa — these properties hold of exactly the same things. These properties have the same extensions across all possible worlds. Identifying properties with their extensions forces us to say that being triangular and being trilateral are the same property. This identity is false, however, as one is the property of having exactly three angles while the other is the property of having exactly three sides. One can multiply these kinds of examples. The properties of being an even prime and being the square root of 4 have the same extensions across all possible worlds. As do the properties of being the largest prime and being a married bachelor. While these properties may be extensionally equivalent, we would not want to say they are the very same property.

One way to avoid these identities is to use impossible worlds. There is an impossible world where the extensions of these properties (restricted to that world) will not match. In such worlds there will be things, for example, with exactly three angles but without exactly three sides, weird as that may be. There are worlds where something is an even prime but is not the square root of 4. There are worlds where something is the largest prime but is not a married bachelor.

Yagisawa's worry is that allowing impossible worlds goes too far, as it implies that no property is ever identical with itself.

> For any property P and any property Q, either it is possible for P and Q not to be coextensive or it is impossible. If it is possible, there is a possible world where P

[18]Recall that on Lewis's extreme realism, worlds do not overlap. Anything that exists at one world does not exist at any other world (though it may have counterparts at other worlds).

[19]That these properties are necessarily coextensive follows from the more general mathematical necessity that every polygon has as many interior angles as it has sides.

> and Q are not coextensive [within that possible world]. If it is impossible, there is an impossible world where P and Q are not coextensive [within that impossible world]. Either way, the set of all possibilia and impossibilia having P is different from the set of all possibilia and impossibilia having Q. Therefore, according to the above proposal, P and Q are not the same property. This is true for any P and Q whatsoever, including P and P. So according to the proposal, no property is the same as any property, including itself! (1988, p. 195)

The first thing to note is that this is only a worry if you accept the existence of impossible worlds *and* the view that identifies properties with their extensions across worlds. One can endorse impossible worlds and drop this view of properties, and so avoid the worry.

But one can even avoid the worry while maintaing both of these views. Let's review the argument. It is impossible that P and P not be coextensive. So there is an impossible world where P and P are not coextensive with respect to the things that exist at that world. So there is an x that is in the extension of P but not in the extension of P. So, according to this argument, P and P are not the same property. But that does not follow.

Every property P has an extension $\delta^+(P,w)$ at world w: the things in w that have property P. Every property also has an anti-extension $\delta^-(P,w)$ at w: the things in w that do not have P.[20] In (classically) possible worlds the extension and anti-extension of a property are mutually exclusive and exhaustive — every thing x either does or doesn't have P (either $x \in \delta^+(P,w)$ or $x \in \delta^-(P,w)$) and not both. In impossible worlds, things can work a bit differently. For any property P, it may be that x isn't in P's extension or in its anti-extension. Or maybe x is in both P's extension and anti-extension. With these tools we can address Yagisawa's worry. It is impossible that P and P not be coextensive. So there is an impossible world w where P and P are not coextensive with respect to the things that exist at that world. So there is an x such that x has P and doesn't have P. So x is in the extension of P: $x \in \delta^+(P,w)$. And x is in the anti-extension of P: $x \in \delta^-(P,w)$. But the fact that x is in the anti-extension of P does not change the fact that x is in the extension of P. It's just in both P's extension and anti-extension. But P and P will have the same extension and anti-extension at all worlds, and so $P = P$.

[20]The functions δ^+ and δ^- take predicate-world pairs to sets of things that exist in that world: P's extension and anti-extension at that world, respectively. The notation is adopted from Priest (2005).

With the framework of extensions and anti-extensions, Yagisawa's worry is not so threatening. But there is one last worry for the extreme realist that we must examine. This worry claims that the extreme realist is not only committed to impossible worlds, she is also committed to things in those worlds as well: *impossibilia*. Consider for example the married bachelor, or the round square cupola on Berkeley College (Quine 1949), or Sylvan's box, which is both empty and occupied (Priest 1997b). These are metaphysically impossible objects. Any world that has any one of them as a part is a metaphysically impossible world. The extreme realist about impossible worlds will then be committed to the existence (in some impossible world) of these objects. As we will see, other realist theories of impossible worlds need not be committed to such objects. For example, those who think worlds are sets of sentences need only be committed to the existence of a set that has as a member the sentence "A married bachelor exists" or "Sylvan's box exists."

Commitment to metaphysically impossible objects, in addition to metaphysically impossible worlds, is not immediately problematic, though one may think it odd. Once one is committed to concrete impossible worlds, concrete impossible objects do not seem like too much of a stretch. What is more worrying is that some of these impossible objects may have influence on the actual world. Consider the following claim.[21]

There couldn't be something that exists in every world.

If this claim is true, then those who endorse concrete impossible worlds are committed to there being an impossible world *w* which has as a part some object that exists in every world, including the actual world.[22] But the extreme realist does not want to be forced to commit to something like that. Unless we can avoid this commitment, it would seem that the apparatus of impossible worlds, on the extreme realist interpretation, too easily commits one to the existence of arbitrary objects in the actual world. I think this would be grounds for rejecting extreme realism about impossible worlds.

[21] Daniel Nolan (1997) describes two similar cases.

[22] The Lewisian realist is really committed to an impossible world that has as a part some object that has a counterpart in every world. Of course, the extreme modal realist need not endorse counterpart theory. She could endorse modal realism with overlap. See McDaniel (2004).

I will look at two ways to approach this problem. The first is to be a full-blooded Lewisian realist about impossible worlds. By that, I mean a Lewisian who is committed to concrete impossible worlds, and who accepts all of the other parts of the Lewisian picture. In particular, the full-blooded realist accepts counterpart theory, as well as the view that truth-at-a-world is just *Truth* simpliciter with quantifiers restricted to the world in question. There is a fairly straightforward response available to the full-blooded Lewisian.

The full-blooded Lewisian can respond as follows. Suppose the problematic claim is true. By true, we mean *True*, not true at the actual world. In the language we have the following *True* claim:

(∃W). *There exists an impossible world w such that at w there exists an object o and o exists at every world.*[23]

This claim is not true at the actual world. For if we restricted quantifiers to the actual world, there wouldn't be any other worlds, and so (∃w) would be false. The statement (∃w) is an existence statement, claiming the existence of a world which satisfies certain conditions. We assume that existence claims in the language of *Truth* are made true in the normal way: by there being something that satisfies the existential sentence. For (∃w) to be *True*, there must exist a concrete impossible world *w* such that at *w* there exists an object *o* and *o* exists at every world, including the actual world. There must be an impossible world that makes the following sentence true:

(∃O). *There exists an object x and x exists at every world.*

The statement (∃o) is also an existence claim, with its quantifiers restricted to *w*. We suppose that existence claims at *w* are made true in the normal way. So there is an object, call it *o*, that exists at *w*. At *w*, it is true that *o* exists at every world. But 'every world' is a universal quantifier. As we are at world *w*, universal quantifiers are restricted to the domain of *w*. What exists at *w*, according to the Lewisian, are those objects that are mereological parts of *w*, including *w* itself. Other worlds, possible or impossible, are not mereological parts of *w*. And so the only world that exists according to *w* is *w*. It follows that *w* is the only world in the domain of the quantifier "every world". As *o*

[23]We do not intend to suggest that there are two kinds of quantification going on, one for worlds and one for objects. One can take worlds and objects that are proper parts of worlds to be objects one and all.

exists at *w*, *o* exists at "every world" according to *w*. But that does not entail that *o* exists at the actual world.

Lewis admits, however, that it is not a hard and fast rule that the modifier "at world *w*" must restrict the domain of every quantifier in a sentence.

> I do not suppose that [our restrictive modifiers] must restrict all quantifiers in their scope, without exception. 'In Australia, there is a yacht faster than any other' would mean less than it does if the modifier restricted both quantifiers rather than just the first. . . . In short, while our modifiers tend to impose restrictions on quantifiers, names, etc., a lot is left up to the pragmatic rule that what is said should be interpreted so as to be sensible. If that means adding extra tacit restrictions, or waiving some of the restrictions imposed by our modifiers, then – within limits – so be it.

In order for (∃o) to be sensible, we would expect that the domain of the universal quantifier is not restricted by the modifier "at world *w*". So we offer another response to the problem posed by impossibilia. This response can also be used by those who reject the understanding of truth-at-a-world as *Truth* with restricted quantifiers.

I still take it that there is an impossible world *w* that makes (∃o) true. And I still take it that the truth of (∃o) at *w* implies that there is some object at *w*, call it *o*. What is true, at *w*, is that *o* exists at every world. But it need not follow that *o* exists at the actual world. The statement (∃o) includes a quantification over all worlds. But statements that quantify over worlds are usually taken to quantify only over worlds that are accessible from the starting world, in this case *w*. We can simply claim that the actual world is not accessible from *w*, or from any world that is impossible from the perspective of the actual world. We can support this inaccessibility claim as follows. The worlds that are possible from the perspective of the actual world are those worlds that are accessible from the actual world. As *w* is an impossible world, it is not accessible from the actual world. Taking the accessibility relation to be symmetric, the actual world is not accessible from *w* either.

Given that *w* does not access the actual world, anything that, at *w*, is true at all worlds need not be true at the actual world. The claim (∃o) says that, at *w*, there exists an object *o* that exists at every world. But here every world means every world accessible from *w*. As the actual world is not accessible from *w*, *o* need not exist in the actual world. So (∃o) can be true, but its truth does not imply that something exists at the actual world. And so those who endorse concrete impossible worlds are not thereby forced to accept that some impossible worlds have influence on the actual world in undesirable ways.

That concludes our investigation of concrete, Lewisian impossible worlds. It appears that extreme realism can survive the addition of concrete impossible worlds. It is still able to provide a reductive theory of modality. It need not force us to accept contradictions that are *True simpliciter*, supposing that the notion of *Truth simpliciter* is a coherent notion. The realist can still give identity conditions on properties in terms of worlds. And the realist need not be committed to arbitrary objects that actually exist.

4.2. Abstractionism. An alternative realist theory of possible worlds takes worlds to be abstract objects. Call this view abstractionism.[24] Abstract worlds differ from concrete worlds in how they make modal truths true. For the extreme realist a talking donkey is possible because there is a concrete possible world that has a concrete object as a part, and that object is a donkey that talks. For the abstractionist a talking donkey is possible because some abstract object represents the sentence "A talking donkey exists" as being true. How this is done depends on the version of abstractionism being considered, of which there are several. A linguistic form of abstractionism takes merely possible worlds to be sets of sentences. Call this view *linguistic abstractionism*. According to linguistic abstractionism, a talking donkey is possible at the actual world because there is a certain kind of set (to be explained below) that has the sentence "A talking donkey exists" as a member. Alternatively, one might take possible worlds to be states of affairs. Call this view *state abstractionism*. According to state abstractionism, a talking donkey is possible at the actual world because there is a certain kind of state of affairs that "includes" the state of affairs <A talking donkey exists>. We examine the consequences of adding impossible worlds to both linguistic and state abstractionism.

[24]The view is also sometimes called ersatzism, as the worlds are ersatz, not concrete, worlds.

According to linguistic abstractionism, possible worlds are sets of sentence.[25] A talking donkey is possible at the actual world because there is a set of sentences that has as a member the sentence "There exists an x such that x is a donkey and x talks." Not every set of sentences is a possible world. To be a possible world, a set S of sentences must satisfy two conditions:

(1) S must contain either A or its negation, for every sentence A.

(2) It must be possible for the sentences in S to be true together.

Consider, for example, the set of all sentences. This set contains every sentence and its negation, so it satisfies the first condition. But most would agree that it is not possible for every sentence in this set to be true together. So it does not satisfy the second condition. And so the set is not a possible world. To rule this set out, the linguistic abstractionist must appeal to some notion of possibility. Possible worlds are sets of sentences such that it is possible for all of the sentences in the set to be true together. This appeal to possibility precludes linguistic abstractionism from offering a reductive account of modality.

Note that the possibility being appealed to is stronger than consistency. A set of sentences could be logically consistent, such that for no sentence A are both A and $\neg A$ members of the set. But possibility requires more than consistency. For example, the maximal set that does not include the sentence "$2+2=4$", but has its negation instead, is a consistent set (as long as there are no other contradictions). The set may even be closed under logical consequence. It's not obvious that the consequences of "$2+2=4$" being false include anything contradictory. However, given that it really is true that $2+2=4$, it is necessarily true. So any collection of sentences that includes its negation cannot all be true together, as the negation of "$2+2=4$" cannot possibly be true. This consistent set of sentences, therefore, is not a metaphysically possible world.

Most linguistic abstractionists would say, because this set is impossible, it is not a world at all. I prefer to say it is an impossible world. It is a way the world couldn't be, because one sentence that describes this way could not be true. There are many other ways the world couldn't be. A set that is not consistent describes one way. A set that does not contain A and does not contain $\neg A$, for some sentence A, describes another way. A set that is not consistent and fails to contain

[25] See Jeffrey (1965) and Lewis (1986), ch. 3.

either A or $\neg A$ describes yet another way. Some sets describe ways the world could be. They're possible worlds. Other sets describe ways the world could not be. Some sets will be logically impossible, such that they include both a sentence and its negation, or such that they are not closed under classical logical consequence, or such that they are not closed under any relation of logical consequence at all. Some will be metaphysically impossible, by including sentences like "Charles is a married bachelor" or "The empty set does not exist". All of these sets are impossible worlds. Taking this perspective to the extreme, one could say that any set of sentences is a world. Some are (logically or metaphysically or physically) possible; some are (logically or metaphysically or physically) impossible.

According to state abstractionism, possible worlds are states of affairs.[26] A talking donkey is possible at the actual world because there is a state of affairs that represents that there exists something that is a donkey and which is able to talk. Not every state of affairs is a possible world. To be a possible world, a state of affairs must be maximal. Following Sider (2003) we say that state S includes state S^* if it is not possible for S to obtain while S^* does not obtain. State S precludes S^* if it is not possible for both S and S^* to obtain. A state S is then maximal if it includes or precludes every other state of affairs. Possible worlds are the maximal states of affairs that are possible.

As with linguistic abstractionism, and for similar reasons, state abstractionism also appeals to some notion of possibility. Such an appeal is required to rule out metaphysically impossible worlds. For, consider any state that includes the state such that the $2+2=5$. Call such a state S. State S is impossible. So for any state S^*, it is not possible for both S and S^* to obtain, as it is not possible for S to obtain on its own. So S precludes every other state, which means it either includes or precludes every other state. So S is maximal. But S is not a possible world. So the state abstractionist must appeal to some notion of possibility to characterize possible worlds.

We supplemented linguistic abstractionism with impossible worlds by saying that any set of sentences is a world. Similarly, we supplement state abstractionism with impossible worlds by saying that any state of affairs is a world. Some states are inconsistent, because they entail some proposition and its negation.[27] They are logically impossible worlds. Some states violate the laws

[26] See Plantinga (1974, 1976).

[27] Where a state S entails a proposition P in the sense that if S obtains, then P is true.

of metaphysics, like the states that entail that Charles is a married bachelor, or that the empty set does not exist. They are metaphysically impossible worlds.

We can point to an example in the literature of a version of state abstractionism that incorporates impossible worlds. David Vander Laan (1997) endorses the view that possible and impossible worlds are states of affairs. Possible worlds are maximal states of affairs that are possible; impossible worlds are maximal states of affairs that are impossible.[28] But his definition of maximality differs from the standard definition given by Sider. For Vander Laan, a state of affairs that includes <San Diego's being warm> entails that 'San Diego is warm' — this sentence is true in that state of affairs. A state of affairs, then, is maximal iff for every sentence A, the state either entails A, or it entails $\neg A$, or it entails both.

For Vander Laan any non-empty collection of sentences is associated with a state of affairs. If a set of sentences makes either A or $\neg A$ true, for every sentence A, then the state that makes all and only those sentences true is a world. If the state is possible, then the world is possible; if the state is impossible, then the world is impossible. Note Vander Laan's theory is not a reductive theory of modality. He appeals to our intuitions that, for example, the states of affiars <Paul squared the circle> and <the number 9 is a Caesar salad> are impossible. And maximal states that entail the sentences "Paul squared the circle" or "the number 9 is a Caesar salad" are impossible worlds.

It is not obvious that any problems arise from adding impossible worlds to either versions of abstractionism. Presumably, if you believe in sets of sentences that make either A or $\neg A$ true, and which can all be true together, then you believe in other sets of sentences: sets of sentences that do not make either A or $\neg A$ true, sets of sentences that cannot all be true together. You may not want to call them possible worlds, which is fine. You may not want to call them worlds at all. But I prefer to say they are impossible worlds. Calling them impossible worlds is little more than a naming preference. Nothing about the sets has changed. Similarly, if you believe in states of affairs that are maximal, then you probably believe in states of affairs that are not maximal. Call a state of affairs that is maximal and possible a possible world. Call a state of affairs that is not both maximal and possible an impossible world. Again, this is simply a change in how we name things.

[28] This use of 'maximal' is Vander Laan's terminology, see Vander Laan (1997), p. 603.

We also point out that the problems that threaten extreme realism about impossible worlds do not pose threats to either form of abstractionism. First, abstractionists need not worry that adding impossible worlds will prevent them from offering a reductive theory of modality. Both theories are admittedly non-reductive, not claiming to offer a reductive theory of modality.

Second, true contradictions at impossible worlds need not entail contradictions that are *True simpliciter*. The modifier 'At world $w\ldots$' is not a restriction on quantifiers, but rather says something like "In set $S\ldots$" or "In state $S\ldots$". These modifier are similar to the modifier "In story $S\ldots$". It may be true that "In story S, A and not A". It does not then follow that "In story S, A" and "It is not the case that: in story S, A.

Third, I am inclined to think that Yagisawa's objection does not quite get off the ground. Properties do not literally have extensions at other worlds, because other worlds are not concrete objects, having particulars as parts. However, these non-actual worlds *represent* that particulars exist and that they instantiate properties and relations. If a world's representing that object a has property P is enough for a to be in the extension of P at that world, then Yagisawa's objection is relevant. But then it can be dealt with in exactly the same way as before.

As for the final objection, that objects in impossible worlds can imply the existence of objects in the actual world, I am again inclined to say that it doesn't even get off the ground. The force of the objection turns on the existence of an object in some non-actual impossible world. But objects do not exist in non-actual worlds; they are only represented as existing. For the linguistic abstractionist, there must be a set that has as a member the sentence "There exists an object x and x exists at every world." The fact that some sentence is a member of a set does not imply that some object exists at the actual world. For the state abstractionist, there must be a state <There exists an object x and x exists at every world.>, which is either identical with an impossible world, or which is included in an impossible world. Perhaps the existence of this state does entail that some object exists at an impossible world. But this "existing at every world" object need not exist at the actual world. For to exist at every world is to exist at every accessible world, that is, at every world possible from the world you started at. If the world is impossible from the perspective of the actual world, then the actual world is impossible from the perspective of it. And so this object need not exist at the actual world.

I would like to mention one last version of abstractionism that is relevant. It is a version of state abstractionism called combinatorialism. The actual world is made up of states of affairs, which are in turn made up of combinations of simple particulars and universals (i.e., properties and relations). To illustrate, suppose that the simples of the actual world $a, b, c, \ldots$ instantiate properties and relations $P, Q, R, \ldots$. We can say that a instantiates P but not Q, and b bears the relation R to c, and so on. There is one way that these particulars actually combine with these properties and relations. But there are several ways in which they *could* combine. According to combinatorialism, for every possible way that these particulars and properties and relations combine, there is a possible world in which they do so combine. Possible worlds, then, are maximal states of affairs that involve recombining the particulars, properties, and relations of the actual world.[29]

It seems that this view generates a proper subset of the worlds given by the more full blown theory of possible worlds as states of affairs. Combinatorialism leaves out some worlds, particularly those worlds that involve alien properties and relations — properties and relations not instantiated in the actual world. And in stating which combinations are worlds, the combinatorialist appeals to *possible* combinations. So, again we have a non-reductive theory of modality.

Interestingly, though, if we leave out the condition that combinations must be possible, we have a view that very easily allows for impossible, even contradictory worlds. We can generate an impossible world in the following way. Suppose, to illustrate, that particles like electrons and protons are simple. Some electron in the actual world has negative charge. Some proton in the actual world has positive charge. Presumably, we can recombine in such a way to get a state that has one of these particles instantiating both positive and negative charge. Any state that includes this state will be an impossible world.[30]

[29]By discussing combinatorialism in this section on realist theories of worlds, I have classified combinatorialism as a realist approach. But there are anti-realist versions of combinatorialism. Armstrong (1989, 1997) endorses the view that non-actual combinations are fictions. We discuss the anti-realist fictionalist approach to possible worlds in the next section.

[30]Clearly particles are not good candidates for the simples, as they have parts. But whatever the simples are (if there are simples), it is likely that of the properties that they can have, some pair of properties will be exclusive. We can

That recombination may lead to inconsistencies is often viewed as a defect in the theory. One can take steps to avoid these inconsistencies. For example, one may try to argue that having positive charge and having negative charge are not genuine properties. Perhaps having charge is a property, and things can have charge in one of two different ways. But this does not avoid the problem. Why can't some particular have this property in both ways? Something more must be said to answer this question, and the only plausible answer is that this couldn't happen because it is impossible. But this will not sway the proponent of impossible worlds. Of course this couldn't happen — it's impossible. So there is an impossible world where it does happen. Eventually, the combinatorialist must appeal to some notion of possibility to rule out these states.

But even with independent justification for ruling out inconsistent recombinations, we can generate metaphysically impossible worlds through recombination in another way. Pick some necessarily existing object. There seems no reason to think that we couldn't just recombine the particulars of the world in such a way that leaves this object out. The resulting recombination will be an impossible world.

Some versions of combinatorialism reject this move, claiming that any recombination must include all of the original simple particulars. The world cannot "contract." It is not entirely clear why this should be. In fact, not being able to contract seems to imply that every simple that exists exists necessarily, as no simple could fail to exist. That doesn't seem too plausible. But even if we accept the "no contraction" condition, we could still generate an impossible world through recombination. Take some simple that has a property necessarily and recombine so that it no longer has that property. A contentious example of this would be an electron that is recombined in such a way that it has positive (and not negative charge). That would be an impossible object. But you might just claim this particle is no longer an electron, but is a positron. So here's another example: recombine the number 9 with the property of being even, or being prime, or both. Any world that includes this state would be impossible.

---

then recombine in such a way that one simple has both of these exclusive properties, thus generating a contradictory state of affairs.

4.3. Hybrid Realism. So far we have taken realist theories of possible worlds, and tried to add impossible worlds that would be of the same ontological kind. To extreme realism we added concrete impossible worlds. To linguistic abstractionism we added impossible worlds in the form of sets of sentences. To state abstractionism we added impossible worlds in the form of states of affairs. Each theory says that possible worlds are of a particular ontological kind. What is sometimes called the Parity thesis says that once one has fixed the ontological kind of possible worlds, impossible worlds should be of the same ontological kind. One need not endorse the Parity thesis.

There is some precedent for accepting a hybrid theory. One may, for example, deny the thesis that all possible worlds are of the same ontological kind. Abstractionists of both stripes deny that merely possible worlds are of the same kind as the actual world (which is a possible world). For abstractionists, the actual world is concrete; the merely possible worlds are not. Similarly, one can deny that possible worlds and impossible worlds are of the same ontological kind.

There are *prima facie* good reasons for rejecting the parity thesis. As we have seen, even with impossible worlds abstractionism cannot give a reductive theory of modality. Extreme realism can offer such a reductive theory. But this benefit of extreme realism is threatened once concrete impossible worlds are added to the theory. The realist is forced to develop a non-modal theory of metaphysical laws. A nice alternative is to mix extreme realism about possible worlds with abstractionism about impossible worlds. Francesco Berto (2010) argues for this kind of hybrid realism. Berto endorses concrete realism for possible worlds, and he takes impossible worlds to be set-theoretic constructions that involve possible worlds, though in an indirect way. Impossible worlds are not simply sets of possible worlds. Impossible worlds are sets of propositions, though only atomic propositions are allowed. I take it that given an impossible world w, the atomic propositions that are in the set w are the atomic propositions that are true at w. So, any atomic proposition p such that $p \notin w$ is false at w. Propositions, on Berto's, view are sets as well — they are sets of possible worlds. Impossible worlds, as sets of atomic propositions, are then sets of sets of possible worlds. Note that this is not circular, as Berto take possible worlds to be Lewisian concrete objects.

Berto's hybrid view certainly has its advantages. Most important of these is that it allows one to maintain a reductive theory of modality, originally captured by Lewis's biconditionals:

P_L. *It is metaphysically possible that A iff there is a world w such that, at w, A.*

N_L. *It is metaphysically necessary that A iff every world w is such that, at w, A.*

All we need do is restrict the quantifiers to concrete worlds and we have a proper, non-modal reduction of possibility. This hybrid view can also give us contradictory worlds, where propositions of the form $A \wedge \neg A$ hold, as long as A is contingent. Given that A is contingent, there will be possible worlds $w_1, \ldots, w_i$ where A holds. On Berto's view the proposition $A = \{w_1, \ldots, w_i\}$. There will also be worlds $w_j, \ldots, w_n$ where $\neg A$ holds. The proposition $\neg A = \{w_j, \ldots, w_n\}$, the complement of A with respect to the domain of possible worlds. We can take the pair set of A and $\neg A$ to get the set $\{\{w_1, \ldots, w_i\}, \{w_j, \ldots, w_n\}\}$. According to Berto, this set, and any set that has these two sets as members, is an impossible world.

But it is not clear that Berto's view can generate all of the impossible situations that one may want.[31] Consider the proposition '9 is even'. It is not true at any possible world. And it appears to be atomic. So on the worlds view of propositions that Berto endorses, this proposition is identical to the empty set. But so is, for example, the proposition '9 is prime', as it also appears to be atomic and false at every possible world. Certainly these are different propositions. One says that 9 can be divided evenly by 2, while the other says that 9 can only be divided evenly by 1 and itself. Berto's view does not account for the difference. It follows that if one of Berto's impossible situations makes one atomic impossible proposition true, then it makes every atomic impossible proposition true. There are then no impossible situations where '9 is even' but it is not the case that '9 is prime'.

It is also unclear that Berto can account for some gappy worlds. An impossible world is a set of atomic propositions. An atomic proposition p is true at w iff $p \in w$. So if $p \notin w$, then w makes p false. So every impossible world makes each atomic proposition either true or false. There are then no impossible worlds that make some atomic propositions neither true nor false. But there

[31] Berto (2010, pp. 484 – 5) concedes this point.

are many ways in which a world may be impossible. And being gappy with respect to atomic propositions certainly seems like one of them.

Both of these worries apply to the similar view, endorsed by Greg Restall (1997), that takes impossible worlds to be sets of possible worlds (not sets of sets of possible worlds). Restall is silent on what possible worlds are. Surely, though, they are not themselves sets of worlds. For if they were, then impossible worlds would be sets of worlds, which would also be sets of worlds, which would also be sets of worlds, and so on. But whatever (else) one takes possible worlds to be, one can then take impossible worlds to be sets of them. So, given that impossible worlds are sets of worlds and possible worlds are not, this is a hybrid view.

Restall's idea is that we take possible worlds for granted. Each possible world is a different way that the world could be. As possible worlds are complete, for every proposition A, each possible world makes A either true or false. Take two possible worlds w, v such that w makes A true and v makes A false. One could represent an impossible world u that makes A both true and false as the set $\{w, v\}$. The impossible world u is characterized by the way w is and the way v is. It's as if one combined all of the information from worlds w and v into one world u.

As with Berto's view, Restall cannot account for impossible worlds that are made impossible in virtue of impossible atomic propositions being true at them. For example, there is no possible world where the number 9 is even. So no set of possible worlds will include any possible world that makes 9 even. So there is no impossible world such that the number 9 is even. And Restall cannot account for gappy worlds either. No possible world is gappy — given any possible world w and any proposition A, either w makes A true or w makes A false. If impossible worlds are combinations of the information given by possible worlds, every impossible world u will be such that, for any proposition A, either u makes A true or u makes A false.

To be sure, both of these hybrid views need work. But it is not obvious that these problems cannot be resolved. And the reductive power of a hybrid theory of worlds is motivation enough to continue work in this direction. One potential approach would be to take possible worlds to be concrete, as with Lewis, and then take impossible worlds to be combinatorially generated from the objects and properties that are parts of possible worlds. But that is a whole other project.

5. Anti-Realist Theories of Impossible Worlds

5.1. Fictionalism. Opposing the realist, who thinks possible worlds (whatever they are) exist, is the anti-realist. The anti-realist thinks that possible worlds do not exist. *Prima facie* this is a difficult position, because possible worlds are so useful. They are also deeply embedded into how contemporary philosophers speak and think. And they have the potential to give a great reductive analysis of modal notions like necessity and possibility. So the anti-realist owes us a story as to exactly what these non-existent entities are, and how they are so useful, given that they don't exist.

One answer says that possible worlds make up a fiction. Possible worlds are a fiction in very much the same way as literary works are fictions. We say that "Sherlock Holmes lives at 221b Baker Street." This sentence is not literally true, because Sherlock Holmes doesn't exist. What we mean is that "According to the Doyle stories, Sherlock Holmes lives at 221b Baker Street." And this sentence is literally true. Likewise, for the modal fictionalist, the sentence "There exists a possible world with a talking donkey" is not literally true. Merely possible worlds do not exist, and there are no actual talking donkeys. But, "According to the fiction of possible worlds, there exists a possible world with a talking donkey" is true.

As a theory of modality, fictionalism interprets modal sentences, sentences about what is possible and what is necessary, in terms of possible worlds — we can translate them into sentences about possible worlds. Modal sentences are true when their possible worlds interpretation is true in the fiction of possible worlds.[32] Take a true modal sentence A, where A is something like "Possibly P" or "Necessarily P". The modal sentence A has a possible worlds translation A^*. For example, the possible worlds translation of "Possibly, there exists a talking donkey" is "there is a possible world such that a talking donkey exists at that world." The fictionalist is then committed to the following equivalence:

(F). *A iff according to the fiction of possible worlds, A^*.*

One can construct fictionalist truth conditions for possibility claims and necessity claims as follows.

[32] Here we are following Rosen (1990).

(FP). *It is possible that A iff according to the fiction of possible worlds, there is a world w such that, at w, A.*

(FN). *It is necessary that A iff according to the fiction of possible worlds, every world w is such that, at w, A.*

The modal fictionalist is not a fictionalist about modality. The modal fictionalist does think that some modal sentences are literally true. Rather they are fictionalists about possible worlds in that sentences claiming the existence of possible worlds, though true in a particular fiction, are literally false.

This is where the main difference between realists like Lewis and anti-realists like the fictionalist lies. It may be that the realist and the fictionalist agree completely on the modal facts, on what is possible and what is necessary. Where they disagree is in how those claims of possibility and necessity are made true. The realist argues that what makes possibility and necessity claims true is the existence of certain things, called worlds, and what goes on at those worlds. The fictionalist argues that what makes these modal claims true depends on what goes on in a particular fiction — the fiction of possible worlds.

The version of fictionalism that I am describing does not take *each* possible world to be its own self-contained fiction. Rather, the fictionalist takes the whole network of merely possible worlds to be a part of one big fiction. According to that single fiction, there are possible worlds where donkeys talk, and there are possible worlds where pigs fly. There are no possible worlds where $2+2=5$.

Fictions are like stories. The fiction of possible worlds is like the fiction of Sherlock Holmes. What goes on in the fiction of Sherlock Holmes is given to us by Conan Doyle's stories. We might similarly ask: What goes on in the fiction of possible worlds? By employing the operator "According to the fiction of possible worlds ..." the fictionalist owes us some explanation as to what exactly this fiction is like. One option is to fall back on stories that others have already told. For example, Lewis has told a story about possible worlds: they are maximal spatiotemporally connected, spatiotemporally isolated mereological sums of individuals. Lewis thinks this story is true. Fictionalists do not. But fictionalists can treat Lewis's story (or something like it) as a fiction, and according to that fiction some things are true. These fictional truths imply those actual modal truths that they are translations of.

At a general level, it seems that being a fictionalist about impossible worlds as well should be relatively unproblematic. At the very least it should not introduce serious problems that do not already threaten fictionalism about possible worlds alone. The fictionalist claims to be able to take advantage of the uses of possible worlds without committing to them. Impossible worlds are also quite useful, as we will see in the next chapter. And so the fictionalist might as well take advantage of their benefits too. Again, no commitment should be required.

What happens, then, if we accept a fictionalist view of impossible worlds? The fictionalist currently has a story that delivers the truthmakers for modal truths, truths about what is possible and what is necessary. But according to this story, which is really Lewis's story about possible worlds, there are no impossible worlds. Given that we need some kind of story about impossible worlds, perhaps the best option is to modify Lewis's story to accommodate them. Essentially this new story has been told. Ira Kiourti (2010) admirably develops a theory of concrete impossible worlds, as an extension of Lewis's extreme realism. And as we have seen from our discussion of extreme realism, and as Kiourti discusses in much greater detail, this view can withstand the main objections and can be developed into a coherent story. We need not accept it as true. All we need do is recognize its coherence, recognize it as a story that can be told.[33]

5.2. Noneism. An alternative anti-realist strategy is to be a noneist about possible worlds. Noneism is the view that some things are non-existent objects.[34] Central to noneism is the distinction between two sets of quantifiers. There is the familiar existential quantifier, $\exists$, read as 'there exists' or 'there is'. Noneists treat this quantifier as existentially loaded. It ranges over things that exist, and nothing else. The noneist introduces another quantifier, $\mathfrak{S}$, read as 'some', which is not existentially loaded. This quantifier ranges over all objects, both existent and non-existent. The familiar universal quantifier $\forall$ is also existentially loaded, ranging only over existent objects. The

[33]And note that consistency is not required for coherence. The Doyle stories are coherent, though arguably inconsistent, as the question of Watson's injury indicates. If one needs further examples of coherent though inconsistent stories, see Priest's (1997b) story of Sylvan's box.
[34]See Routley (1980), and Priest (2005), ch. 7.3.

noneist introduces another universal quantifier, $\mathfrak{A}$, which is not existentially loaded, ranging over both existent and non-existent objects. However, if we are allowed an existence predicate, E, then the existentially loaded quantifiers can be defined in terms of the non-existentially loaded ones.

$$\exists x P(x) \stackrel{\text{def}}{=} \mathfrak{S}x[E(x) \wedge P(x)] \qquad \forall x P(x) \stackrel{\text{def}}{=} \mathfrak{A}x[E(x) \rightarrow P(x)]$$

The familiar existentially loaded quantifiers are simply the noneist's quantifiers restricted to those things that exist. And so there is really only one kind of quantifier.[35]

To be a noneist about merely possible worlds is to claim that merely possible worlds are non-existent objects. There seems no reason for the noneist to admit the non-existence of merely possible worlds, but reject the non-existence of impossible worlds. They can all be non-existent objects together. The noneist, like the fictionalist, can apparently make use of possible worlds without committing to their existence. Given the legitimacy of such a move, nothing should further prevent the noneist from taking advantage of impossible worlds without committing to their existence either.

As a matter of fact, most noneists already accept impossible worlds as non-existent objects. The force of the noneist's position is that beyond the concrete physical objects of the actual world, there are mere objects of thought. Objects of thought include fictional characters, abstract objects, merely possible and impossible objects, merely possible and impossible worlds.[36] The noneist calls mere objects of thought non-existent objects. We get these non-existent objects for free. We can think about them; we can talk about them. We can quantify over them with quantifiers that are not existentially loaded. We can describe properties that we take them to have.[37]

In conclusion, there is good reason to think that we can add impossible worlds to many of the mainstream theories of possible worlds, both realist and anti-realist. There is also little reason to think that impossible worlds will generate new problems that cannot be resolved. In fact, impossible worlds can help to solve well-established problems in contemporary metaphysics. These include, as we will see in Chapter 5, problems of differentiating distinct properties, problems of

[35]These equivalences require that the domain of existent objects be non-empty.

[36]On some versions of noneism, mere objects of thought also include past and future objects of the actual world.

[37]Which most likely differ from the properties that they *do* have. See Priest (2000).

differentiating distinct propositions, and problems of counterfactual reasoning. Given that impossible worlds are relatively unproblematic, our next task is to explore the value of embracing them.

CHAPTER 5

In Favor of Realism

We have seen that mainstream theories of possible worlds can accommodate their impossible counterparts without facing insurmountable problems. But compatibility is no argument in favor of existence. We now consider some positive arguments for the existence of impossible worlds. We then consider (and reject) a general argument against their existence.

1. The Argument For

The main argument that we consider builds a case in favor of the existence of impossible worlds in two stages. The first stage examines the role of impossible worlds in the logical analysis of three important concepts. The first is a conception of conditionality, often expressed by the English indicative conditional "if ... then". The second and third are the related concepts of knowledge and belief. The goal is to show that impossible worlds play a crucial role in the logical analysis of all three of these concepts. In effect, the logical analysis builds a model or representation of how they really work. One could say that the model captures the meaning of the concept, though that is a strong claim. At the very least, it gives truth conditions for claims that involve the concept. If this representation is to be accurate, if it is to give the correct truth conditions for the concept, then we should think that the entities posited by the model correspond to real things, things that are "out there" somewhere. Models of one kind of conditionality, and of knowledge and belief, posit entities that look a lot like impossible worlds. We should therefore think that impossible worlds exist.

One should find this argument compelling because there is some precedent for it. Consider the logical analysis of modality. The logical analysis of modality appeals to a mathematical representation: classes of Kripke models. Kripke models posit a domain of entities, which we will call points. Because some classes of Kripke models give an accurate representation of how we think modality works, we should think that the points in the model correspond to things that exist. We call those things possible worlds. There is therefore reason to think that possible worlds exist (whatever they may be).[1]

The second stage of our argument in favor of the existence of impossible worlds considers the implicit use of impossible worlds in common, everyday philosophical activity and debate. We will see that impossible worlds find their way into many areas of philosophical thought, from metaphysics to mathematics to ethics. Reasoning in these areas often takes the form of counterfactual reasoning, and so aspects of conditionality play a role here as well. In many cases, we are reasoning about situations that are impossible. And so, though many dare not admit it, as philosophers we take advantage of the conceptual resources of impossible worlds all of the time. Indeed, we often need to make claims about what is true in and what is true of these impossible worlds. If our philosophical views turn on the truth of these claims, claims that appeal to the existence of impossible worlds, there is then reason to think that impossible worlds exist. We will also see that impossible worlds play a necessary role in some theories of properties and propositions.

1.1. Logical Precedent. We begin with an examination of the logical analysis of the concept of modality and the role that possible worlds play in that analysis. A possible worlds interpretation of modality goes back at least to Carnap (1946), and arguably further, as Carnap was under the influence of Wittgenstein. Crucial to embracing this interpretation, however, was the development of the role of a binary relation on worlds, known as an *accessibility* relation. The accessibility relation is often attributed to Saul Kripke (1963, 1965), though it can be found in the work of A. N. Prior (1958). Something similar to an accessibility relation is even present in the work of Jónsson and Tarski (1951). Since then, the possible worlds interpretation of modality has become commonplace in the study of logic. For a history of the development of possible worlds semantics, see Copeland (2002).

[1]Essentially, this is one of Lewis's arguments. See Lewis (1986), Ch. 1.2.

Normal Modal Logics. The language of modal logic includes the language of classical propositional logic: propositional variables $p_0, p_1, \ldots$, the connectives $\neg$ for negation, $\wedge$ for conjunction, $\vee$ for disjunction, $\supset$ for the material conditional. The language of propositional logic is then extended with two propositional operators: $\Diamond$ for possibility and $\Box$ for necessity.[2]

Interpretations for the language of propositional modal logic are based on ordered pairs $\langle W, R\rangle$ called Kripke frames. The set W is a set of points; the relation R is an accessibility relation between points in W. When Kripke frames are used to interpret certain kinds of modality, the points are thought of as worlds, and the intended meaning of R is that if wRw', then world w' is possible from the perspective of world w. Kripke interpretations of propositional modal logic are ordered triples $\langle W, R, v\rangle$, which add a valuation function v from point-proposition pairs into the set $\{\mathbf{t}, \mathbf{f}\}$ of truth values.[3] Given that a proposition p is true at point w, then $v_w(p) = \mathbf{t}$. Otherwise, $v_w(p) = \mathbf{f}$. The truth conditions for propositional connectives are as usual.

$$\begin{array}{lll}
v_w(\neg A) = \mathbf{t} & \text{iff} & v_w(A) = \mathbf{f} \\
v_w(A \wedge B) = \mathbf{t} & \text{iff} & v_w(A) = \mathbf{t} \text{ and } v_w(B) = \mathbf{t} \\
v_w(A \vee B) = \mathbf{t} & \text{iff} & v_w(A) = \mathbf{t} \text{ or } v_w(B) = \mathbf{t} \\
v_w(A \supset B) = \mathbf{t} & \text{iff} & v_w(A) = \mathbf{f} \text{ or } v_w(B) = \mathbf{t}
\end{array}$$

The truth conditions for the modal operators are as follows:

$$\begin{array}{lll}
v_w(\Diamond A) = \mathbf{t} & \text{iff} & \text{for some } w' \text{ such that } wRw', v_{w'}(A) = \mathbf{t} \\
v_w(\Box A) = \mathbf{t} & \text{iff} & \text{for all } w' \text{ such that } wRw', v_{w'}(A) = \mathbf{t}
\end{array}$$

These interpretations provide a formal, mathematical structure that can be used to investigate philosophical concepts that involve modality. These concepts include possibility and necessity, belief and knowledge, time and tense. For now, we focus on the logical analysis of possibility and necessity.

Arguably, we can use these interpretations to give a strikingly accurate representation (a "model", though in a less technical sense of the word) of how we think various kinds of possibility and necessity work. This is done by looking at the interpretations in the following way: points are to

[2] More details on the logics examined in this chapter can be found in Priest (2008). We have here followed Kripke's original use of the term "normal".

[3] These are more often called Kripke models.

represent possible worlds, and the accessibility relation R is to represent a relation of relative possibility between possible worlds. We can then add restrictions on the accessibility relations of the models we look at. It is common to think that the interpretations we should be interested in, to study something like metaphysical possibility and necessity, correspond to the **S5** class of interpretations, where the accessibility relation R is reflexive, symmetric, and transitive. You might even think that, to model the metaphysical modalities, the accessibility relation should be universal, so that every world is possible from the perspective of every other. Such a view would imply that, from the perspective of any one world, there are no inaccessible (i.e., impossible) worlds.

These interpretations are just mathematical representations. But they often give us the right answers to philosophical questions that are genuinely about possibility and necessity. Understandably, one could ask why this happens. One answer says that it happens because the elements of the interpretation accurately represent things that are "out there" somewhere. As these interpretations make use of things that are supposed to correspond to possible worlds, this response suggests that possible worlds exist. And the interpretation is accurate because these possible worlds really do bear some relation to one another (call it "accessibility" or "relative possibility"), a relation that is reflexive, symmetric, and transitive, or perhaps even universal. Of course, the question of what possible worlds are is left open. They could be concrete worlds like ours, states of affairs, sets of sentences, etc. As we saw in the last chapter, there are many realist theories of possible worlds. The interpretations don't decide one way or another as to the ontological category of possible worlds. But if the truth conditions are to be correct, points must correspond to things that exist; call those things possible worlds. There is then reason to think that possible worlds exist.[4]

We therefore have a pair of concepts, possibility and necessity, whose logical analysis depends on the existence of possible worlds. There are other concepts, like knowledge and belief, time and tense, etc., whose logical analysis also makes use of possible worlds. Our goal is to find some concept whose logical analysis similarly depends on the existence of impossible worlds.

[4] We make no presupposition that "existence" in this sense is to mean existence within or at the actual world.

An impossible world, we may recall, is simply a world that has different laws. In the context of logical analysis, we are looking for worlds that are logically impossible, worlds that have different logical laws, worlds that break the (actual) laws of logic.

Non-Normal Modal Logics. Generally, the interpretations that people are interested in are "normal" modal interpretations. Alongside normal interpretations, Kripke also developed modal interpretations that he called non-normal. Non-normal modal frames $\langle W, N, R \rangle$ have as a member an additional set of normal points $N \subseteq W$. N may be a proper subset of W, and the complement $W - N$ is then the set of non-normal points. Non-normal modal interpretations $\langle W, N, R, v \rangle$ add a valuation function v from point-proposition pairs into the set $\{\mathbf{t}, \mathbf{f}\}$ of truth values. The truth conditions for propositional connectives are as usual. The truth conditions for $\Box$ and $\Diamond$ are as before when dealing with normal points. The truth conditions for $\Box$ and $\Diamond$ for non-normal points are as follows:

$$v_w(\Diamond A) = \mathbf{t}$$
$$v_w(\Box A) = \mathbf{f}$$

Interpreting the box and diamond as necessity and possibility, we have that at non-normal points, nothing is necessary and everything is possible. If there were reason to think that some particular non-normal interpretation accurately represented the space of possibility, matching up points with worlds, then it would be tempting to identify non-normal points with logically impossible worlds. Logical laws are necessary, if anything is. And we might also think that it is a law of logic that logical laws are necessary. For any logical law L, L is necessary, and it is a logical law that L is necessary. But at non-normal points, L is not necessary, because nothing is. So non-normal points violate at least one logical law, the law that logical laws are necessary. Analogously, logical falsehoods are necessarily false — they are impossible. And it is presumably a logical law that they are impossible. But non-normal points make impossible statements possible, thus violating the law that impossible statements are impossible. Non-normal points make the necessary contingent and the impossible possible. This makes them good candidates for impossible worlds.

But non-normal interpretations are only mathematical structures. Unless an interpretation gives a plausible representation of some feature of the world, some concept like possibility or knowledge or conditionality, there is no reason to believe that any component of the interpretation matches up with any feature of reality. It is the ability of the interpretation to successfully capture the semantics of a concept, and the truth conditions for sentences that use the concept, that yields an argument in favor of the existence of worlds.

One could argue that there are concepts whose semantics are accurately captured by non-normal interpretations. One way to articulate a concept is to lay down a set of axioms that the concept obeys. In fact, Kripke's non-normal interpretations are intended to provide semantics for the (C. I.) Lewis axiomatizations **S2** and **S3** of the strict conditional (Lewis and Langford 1931). What Lewis wanted was an axiomatization that allowed for a conditional that is stronger than the material conditional. The material conditional is fine for what it is, but it captures a relatively weak sense of conditionality. Lewis wanted to capture a more robust conditional, according to which conditional sentences "if A, then B" say B really *follows from* A. What Lewis came up with was the strict conditional: $A \prec B \stackrel{\text{def}}{=} \Box(A \supset B)$.[5] According to Lewis, one things's following from another is stronger than the sense of conditionality that the material conditional gives us. For example, it should not be that everything follows from a false proposition, or that every true proposition follows from any proposition whatsoever, or that for any two propositions A and B, either B follows from A or A follows from B.

It is this stronger conditional, which represents one thing following from another, that **S2** and **S3** (and **S1** for that matter) hope to capture. All of these axiomatizations can be given a semantics that appeals to non-normal worlds. For systems **S2** and **S3**, non-normal worlds are characterized by the modal truth conditions given above. For **S1**, the semantics still appeal to non-normal worlds, but the modal truth conditions are more complicated (Cresswell 1972, 1995). Lewis's preferred system to represent this conception of conditionality is **S3**.

So one might argue that there is a conception of conditionality that systems **S1** - **S3** try to capture, and which is arguably an important feature of the world. Furthermore, the truth conditions for this conditional require an appeal to non-normal points, and non-normal points look like they

[5]The phrase "follows from" can be used to mean many things. In the remainder of this section, we follow Lewis and treat "follows from" as a connective, like the material conditional. Lewis tried to capture the idea of one thing following from another with the strict conditional, as symbolized by the connective $\prec$. This section examines whether Lewis was successful.

should correspond to impossible worlds, as they break certain logical laws. However, one could respond that these systems might not be entirely satisfactory to capture the intended conception of *follows from*, where "B follows from A" is represented as the strict conditional $A \multimap B$. For instance, even though $A \multimap B$ isn't automatically true when A is false, it is automatically true when A is necessarily false. And even though $A \multimap B$ isn't automatically true when B is true, it is automatically true when B is necessarily true. In other words, everything follows from a necessarily false proposition, and a necessarily true proposition follows from anything. It doesn't seem, given Lewis's intended understanding of *follows from* as a conditional represented by $A \multimap B$, whether these paradoxes of the strict conditional can be justified. Furthermore, one might hope that a conditional representing the relation of one thing's following from another could ensure that the one thing is relevant to the other. But the strict conditional cannot guarantee relevance between antecedent and consequent. For example, regardless of what A and B are, the strict conditional formula $(A \wedge \neg A) \multimap B$ will always be true. And so, while the semantics, making use of non-normal points that could plausibly be matched up with impossible worlds, try to capture a conditional that represents the notion of *follows from*, it's not clear that it is successful. There is then not as much reason to think that these non-normal points can be matched up with any feature of reality. You might even think that the strict conditional as axiomatized by, e.g., **S5**, is closer to what we want from a representation of this kind of conditionality (even though the same paradoxes of the strict conditional arise here), and the semantics for **S5** do not make any appeal to non-normal points. So, even if we were to pursue this line of argument, we still may not need impossible worlds.

The Lewis **S** systems, therefore, are not satisfactory to capture the kind of conditionality that is supposed to represent *follows from*. But there are also the closely related Lemmon systems **E1** - **E3** (Lemmon 1957). The difference between the **S** systems and the **E** systems is that validity in the **S** systems is defined as truth preservation at every normal world of every interpretation, while validity in the **E** systems is defined as truth preservation at every world (normal or non-normal) of every interpretation.

In developing the **E** systems, Lemmon was more interested in necessity rather than conditionality. Lemmon intended these logics to capture a notion of scientific necessity. That is, $\Box A$ is to be understood as *A is scientifically necessary but not logically necessary*. Furthermore, the semantics for these logics require the use of non-normal points, which, if the semantics is to be correct, should correspond to impossible worlds. So we appear to have a concept that requires the existence of impossible worlds. But the problem here is that the notion of scientific necessity is not too clear. As we noted above, validity in these systems is defined as truth preservation at every world of every interpretation. But the truth conditions for $\Box A$ at non-normal worlds, for **E2** and **E3**, are the same as those for **S2** and **S3**: $v_w(\Box A) = \mathbf{f}$. So no instance of $\Box A$ is true at all worlds, and so the **E** systems do not validate any instance of $\Box A$. Unfortunately, Lemmon says very little about the notion of scientific necessity. Clearly it is not logical necessity. But what kind of necessity is it, especially if nothing is scientifically necessary? More would need to be said if we are to claim that these semantics represent some well-defined and relatively well-understood concept, like a conception of conditionality, or like knowledge. And so, not much case has been given that the non-normal points, as elements of the semantics, should match up with anything in the world.

First Degree Entailment. We introduce non-normal modal interpretations in order to explore the idea of a non-normal point, which will feature prominently in some of the logics we examine later. For now, we leave point-based semantics to explore a family of logics that are important because of how they treat the relationship between truth and falsity. Classically, truth and falsity are mutually exclusive and exhaustive — for every proposition A, A is either true or false, and not both. That is, for any classical interpretation, v is a total valuation function into the set $\{\mathbf{t}, \mathbf{f}\}$.

But truth and falsity need not be exclusive or exhaustive. Instead of a *total* valuation *function* v into $\{\mathbf{t}, \mathbf{f}\}$, we can take a *partial* valuation *relation* ρ, which may relate a proposition A to true, false, both, or neither. Note that there are still only two truth values, $\mathbf{t}$ and $\mathbf{f}$. But any proposition may relate to both $\mathbf{t}$ and $\mathbf{f}$, or it may relate to neither $\mathbf{t}$ nor $\mathbf{f}$. The logic of first degree entailment is a very basic logic that tries to capture this more relaxed perspective on the relationship between truth and falsity. The language of first degree entailment includes propositional variables $p_0, p_1, \ldots$, and the connectives $\neg$ for negation, $\wedge$ for conjunction, and $\vee$ for disjunction. The material conditional $A \supset B$ is defined as $\neg A \vee B$. The propositional variables are assigned truth values by way of the relation ρ. $(A)\rho\mathbf{t}$ means that ρ assigns the truth value $\mathbf{t}$ to the proposition A. The truth values of the other formulas are then defined recursively.

$$\begin{array}{lll}
(A \wedge B)\rho\mathbf{t} & \text{iff} & (A)\rho\mathbf{t} \text{ and } (B)\rho\mathbf{t} \\
(A \wedge B)\rho\mathbf{f} & \text{iff} & (A)\rho\mathbf{f} \text{ or } (B)\rho\mathbf{f} \\
(A \vee B)\rho\mathbf{t} & \text{iff} & (A)\rho\mathbf{t} \text{ or } (B)\rho\mathbf{t} \\
(A \vee B)\rho\mathbf{f} & \text{iff} & (A)\rho\mathbf{f} \text{ and } (B)\rho\mathbf{f} \\
(\neg A)\rho\mathbf{t} & \text{iff} & (A)\rho\mathbf{f} \\
(\neg A)\rho\mathbf{f} & \text{iff} & (A)\rho\mathbf{t}
\end{array}$$

As an alternative to these relational semantics, one could specify a many-valued semantics for first degree entailment. To do this, we revert to a total valuation function from atomic propositions into the set $\{\mathbf{t}, \mathbf{f}, \mathbf{b}, \mathbf{n}\}$, where $\mathbf{b}$ is to be interpreted as **both** true and false, and $\mathbf{n}$ is to be interpreted as **neither** true nor false.

A third alternative is to give a point-based semantics. This semantics is two-valued, but it requires an additional operator $*$, the Routley star. A Routley interpretation for first degree entailment is a triple $\langle W, *, v \rangle$. As with normal modal interpretations, the set W is a set of points, and v is a valuation function from point-proposition pairs to truth values. The Routley star is used to indicate that every point $w \in W$ has a mate $w^* \in W$, such that $w^{**} = w$. These pairs of points are used to give truth conditions for negation. The truth conditions for conjunction and disjunction, relativized to points, are classical.

$$\begin{array}{lll}
v_w(A \wedge B) = \mathbf{t} & \text{iff} & v_w(A) = \mathbf{t} \text{ and } v_w(B) = \mathbf{t} \\
v_w(A \vee B) = \mathbf{t} & \text{iff} & v_w(A) = \mathbf{t} \text{ or } v_w(B) = \mathbf{t} \\
v_w(\neg A) = \mathbf{t} & \text{iff} & v_{w^*}(A) = \mathbf{f}
\end{array}$$

In the next section, we extend the point-based semantics of first degree entailment to examine interpretations of a couple of relevant logics.

With first degree entailment, we get a first look at the family of logics that are paraconsistent. Paraconsistent logics try to capture the idea that contradications should not *explode*. That is, it shouldn't be the case that a contradiction logically implies any proposition whatsoever. Paraconsistent logics are defined as logics that do not validate the inference schema *ex falso quodlibet*:

EX FALSO QUODLIBET (EFQ). $A, \neg A \vDash B$

In addition to EFQ, there are other ways to capture the idea that contradictions explode. One might say that contradictions materially imply every proposition. This would give us a material version of EFQ.

MATERIAL EFQ. $\vDash (A \wedge \neg A) \supset B$

Or one might say that contradictions strictly imply every proposition, where strict implication is understood modally. This would give us a modal version of EFQ.

MODAL EFQ. $\vDash (A \wedge \neg A) \prec B$

While there are many ways to represent exploding contradictions, classical logic validates all of these versions of EFQ.[6] That classical logic validates EFQ is viewed by some to be a deficiency. Why should *everything* follow from a contradiction? Paraconsistent logics do not validate EFQ, and so they do not suffer this deficiency.

The failure of EFQ is especially important for the study of logically impossible worlds. One intuitive way for a world to be impossible is for it to make some contradiction true. If we are to take contradictory worlds seriously, we should not want EFQ to preserve truth at these worlds. For if EFQ is valid, contradictory worlds explode: they make every proposition true. As a consequence, there is then only one contradictory world, the exploded world.[7] And what happens in that world isn't very interesting.

There is another reason, however, that one might think EFQ is invalid. All three versions of EFQ express a relationship of conditionality or of consequence between an arbitrary contradiction and an arbitrary formula. It may be that the A and B in this relationship have nothing to do with each other. But there is a conception of the conditional according to which there should be some connection between A and B. They should have something in common. In other words, they should be relevant to one another.

[6]That is, non-modal classical logic validates the model-theoretic inference EFQ and the tautological Material EFQ. Modal classical logic validates all three.

[7]This assumes that there are no non-identical indiscernible worlds, i.e., no two different worlds that make exactly the same propositions true. This is a non-trivial metaphysical assumption, but nothing too important turns on the assumption in this context.

Relevant Logics. While first degree entailment does not validate the inference EFQ or the formula $(A \wedge \neg A) \supset B$, it will not capture the idea of relevance that one might want from a conditional, or from a conception of consequence.[8] Even worse, the material conditional $\supset$ as defined in first degree entailment is seriously deficient. Most worrisome is the fact that $\supset$ does not satisfy the inference schema *modus ponens*:

MODUS PONENS (MP). $A, A \supset B \vDash B$

A connective that does not satisfy MP can hardly be called a conditional. To do better, relevant logics build on the framework of first degree entailment by adding a new conditional $\rightarrow$, which we call a relevant conditional.

Relevant logics embrace the idea that for a conditional formula to be valid, the antecedent and consequent should "have something in common", they should be relevant to one another. Capturing this idea of relevance from a logical perspective begins with the idea of variable sharing: for a conditional $A \rightarrow B$ to be true, A and B must have a propositional variable in common.[9] It is a necessary (though not sufficient) condition on relevant logics that they comply with the idea of variable sharing

Rather than stipulating that antecedents and consequents of valid conditionals share a parameter, one approach to relevant logics tries to show that parameter sharing can fall out of something more basic. This can be done using a point-based semantics that includes non-normal points, and it can be done either with a truth relation ρ or with the Routley star $*$. We examine the logic $\mathbf{N}_*$, which extends the Routley star semantics of first degree entailment with truth conditions for the relevant conditional $\rightarrow$, and relativizes to points. Interpretations for the language of $\mathbf{N}_*$ are quadruples $\langle W, N, *, v \rangle$, where W is a set of points, $N \subseteq W$ is a set of normal points, and v is a function from point-proposition pairs into $\{\mathbf{t}, \mathbf{f}\}$. The truth conditions for the relevant conditional $\rightarrow$ are as follows:

If w is normal, then:
$v_w(A \rightarrow B) = \mathbf{t}$ iff for all $x \in W$ such that $v_x(A) = \mathbf{t}$, $v_x(B) = \mathbf{t}$

[8] The language of first degree entailment does not have the resources to express the formula $\Box[(A \wedge \neg A) \supset B]$.
[9] This applies when dealing with *propositional* relevant logics.

If w is non-normal, then the truth value of $A \rightarrow B$ is not defined recursively,
but are determined by v.

The logic $\mathbf{N}_*$ is a relatively weak logic. Stronger relevant logics build on the semantics of $\mathbf{N}_*$ by adding an accessibility relation R on points. Interpretations of these relevant logics are then five-tuples $\langle W, N, R, *, v\rangle$. Unlike the accessibility relation in modal logics, however, the R of relevant logics is a ternary relation on W. And it is used to give truth conditions, not for modal formulas involving $\Diamond$ and $\Box$, but for formulas using the relevant conditional $\rightarrow$.

The simplest of these strengthened relevant logics is **B**. The logic **B** gives the truth conditions for all of the connectives just as $\mathbf{N}_*$ does, except for the relevant conditional.

If w is normal, then:
$v_w(A \rightarrow B) = \mathbf{t}$ iff for all $x \in W$ such that $v_x(A) = \mathbf{t}$, $v_x(B) = \mathbf{t}$

If w is non-normal, then:
$v_w(A \rightarrow B) = \mathbf{t}$ iff for all $x, y \in W$ such that $Rwxy$, if $v_x(A) = \mathbf{t}$, then $v_y(B) = \mathbf{t}$

Uniform truth conditions in terms of the accessibility relation R can be given for the conditional under both normal and non-normal points. At non-normal points, the condition is as above. At normal points w, we use the condition for non-normal points and stipulate that $Rwxy$ iff $x = y$.

Logics stronger than **B** are achieved by modifying the properties of the accessibility relation. These modifications go beyond those made in the various systems of modal logic.[10] The best know of these logics is **R**. Given the point-based semantics for these relevant logics, it can be shown for **R** (and so for those systems weaker than **R** like **B**), that if A and B do not share any parameters, then the conditional $A \rightarrow B$, where $\rightarrow$ is the relevant conditional, is not valid. Therefore, if $A \rightarrow B$ is valid, then A and B are relevant to each other.[11]

We can also show that **R** does not validate EFQ or the related formula $(A \wedge \neg A) \rightarrow B$. To construct a counter-model, it is useful to use a many-valued logic. The relevant logic **R** is a sub-logic of another logic $\mathbf{RM}_3$, which is a three-valued logic. Because **R** is a sub-logic of $\mathbf{RM}_3$, nothing that is invalid in $\mathbf{RM}_3$ is valid in **R**. Showing that $(A \wedge \neg A) \rightarrow B$ and EFQ are invalid in

[10] For details, see Priest (2008), ch. 10.

[11] See Belnap, Jr. (1960) for a proof that the conditional of a stronger system **E** (which is different from the Lemmon **E** systems) is relevant. The proof also works for **R**.

$\mathbf{RM}_3$ is sufficient to show that they are invalid in **R** as well. In $\mathbf{RM}_3$ validity for formulas is defined as taking a designated value **t** or **i** in all interpretations, and validity for inferences as preserving a designated value in all interpretations.

Truth functions in $\mathbf{RM}_3$ for the connectives $\wedge, \vee$, and $\neg$ are given by the strong Kleene truth tables. The truth function for the conditional is given by the following truth table:

$f_{\rightarrow}$	**t**	**i**	**f**
t	**t**	**f**	**f**
i	**t**	**i**	**f**
f	**t**	**t**	**t**

It is easy to check that assigning the values $v(A) = \mathbf{i}$ and $v(B) = \mathbf{f}$ will give a counter-model to the formula $(A \wedge \neg A) \rightarrow B$ and to the inference EFQ. Given the existence of a counter-model, neither the formula nor the inference is valid in $\mathbf{RM}_3$, and so neither is valid in **R**.

The language of **R** cannot express the formula $\Box[(A \wedge \neg A) \rightarrow B]$, as $\Box$ is not a part of the language. But the logic $\Box\mathbf{R}$ can be seen as an extension of **R** in the sense that the conditional $\Rightarrow$ of $\Box\mathbf{R}$ is supposed to be a strict relevant conditional.[12] The connection between the relevant conditional $A \rightarrow B$ of **R** and the strict relevant conditional $A \Rightarrow B$ of $\Box\mathbf{R}$ is analogous to the connection between the classical material conditional $A \supset B$ and the classical strict conditional $A \multimap B$. Just as the formula $(A \wedge \neg A) \rightarrow B$ is invalid in **R**, the formula $(A \wedge \neg A) \Rightarrow B$ is invalid in $\Box\mathbf{R}$.

What we have achieved with the relevant logic **R** is a logical analysis of a concept. It is an analysis of a conditional, as often expressed by the English indicative "if ... then". This representation (or "model") of the English indicative conditional appears to be far more accurate than the material conditional of classical logic or the strict conditional of classical modal logic. First, it satisfies the requirement that the antecedent and the consequent of a conditional are relevant to each other, in the sense that they share a parameter. Second, it does not validate the inference EFQ or related formulas involving contradictions, like $(A \wedge \neg A) \rightarrow B$. We argue that there is an important conception of the English indicative conditional, such that any representation of this conditional should satisfy these requirements. The logic **R** does this in the most satisfactory way.

[12] Here, the conditional $\Rightarrow$ is not to be confused with the notation for logical inference as used in the last chapter.

It is interesting to consider why the inference EFQ, and the formula $(A \wedge \neg A) \rightarrow B$, are invalid in relevant logics like **R**. If we use a classical logic to evaluate formulas at points, there are no points where $A \wedge \neg A$ holds. That there are no such points is guaranteed by the classical semantics of negation: $\neg A$ is true at a point w just when A is not true at w. But the semantics for negation in **R** are different: $\neg A$ is true at a point w just when A is not true at some point w^*, where it needn't be the case that $w = w^*$. It follows that there may be points w such that both A and $\neg A$ hold at w. Note that the fact that $A \wedge \neg A$ holds at w does not automatically make w a non-normal point. So there can be normal points where $A \wedge \neg A$ holds, but B does not, invalidating both the formula $(A \wedge \neg A) \rightarrow B$ and the inference EFQ.

The relevant logic **R** makes use of points where contradictions, like $A \wedge \neg A$, hold. It also gives us a representation of an important conception of the English indicative conditional, one that is relevant and one that does not validate EFQ. Given the accuracy of this representation, there is reason to think that it is a plausible candidate for being the correct understanding of this English conditional. If it is correct, then we should think that the points of relevant logic correspond to things that are "out there" somewhere, things that exist. After all, it is these points that are doing the work, that make "if ... then" statements true. Given that some of these points are points where, e.g., $A \wedge \neg A$ holds, we have a *prima facie* argument for what we might call logically impossible worlds. Whatever these things are that are making formulas like $A \wedge \neg A$ true, call those things impossible worlds. From the classical perspective, a world where $A \wedge \neg A$ holds is logically impossible, as it is a world that breaks a law of logic. If worlds that break the laws of logic exist, then impossible worlds exist.

But the last step of that argument is too quick. If we think of the logic **R** as telling us about which inferences preserve truth, then it is not obvious that worlds where $A \wedge \neg A$ hold are worlds that break the laws of logic *as given according to* **R**. Logically impossible worlds are worlds that break the actual laws of logic, whatever the actual laws of logic happen to be. If the actual laws of logic correspond (or at least include) those of the logic **R**, then those are the laws we use to determine if a world is possible or impossible. Normal worlds, even if they make $A \wedge \neg A$ true, do not break the laws of logic according to **R**. (In fact, the formula $\neg(A \wedge \neg A)$ holds at these worlds as well, as it is a valid formula. One could argue, then, that the law of non-contradiction has not been broken.) So worlds that make $A \wedge \neg A$ true are not automatically impossible. On this view, however, worlds that obey the laws of classical logic are impossible, as they validate inferences (like EFQ) that are invalid according to **R**.

But there is another mechanism of relevant logic that does suggest some of the points in these interpretations correspond to logically impossible worlds. Recall that interpretations of relevant logic rely on a subset of points that are non-normal. Interestingly, at some non-normal points formulas that are logically true from the perspective of **R** fail to hold. In fact, one can show that for *any* formula, there is a point in an interpretation where it fails to be true. We show this in the Appendix to this chapter. Assuming that for every formula A, there is a point in an interpretation where A fails, we can show that at non-normal points formulas that are logical truths fail to hold. For if there is a point where some logical truth A is false, then as logical truths are true at every normal point (by the definition of validity for **R**), A must fail at a non-normal point. As non-normal points are points where logical truths may fail to hold, they intuitively should correspond to things that are logically impossible. These points break the laws of logic. If these points are to correspond to things in the real world, it is appropriate to call those things impossible worlds.

We therefore have a model of a concept, a conception of conditionality as is often captured by the English indicative conditional, the truth conditions for which require the existence of impossible worlds. The success of this model suggests that non-normal points correspond to things that are out there somewhere. Call those things impossible worlds. There is then reason to think that impossible worlds, whatever they are, exist.

There are two worries that we should address before we move on. The first worry concerns the relevant conditional; the second worry concerns relevant negation. One might worry about the relevant conditional because the truth conditions for the relevant conditional seem, to put it bluntly, unintelligible. The truth conditions for the relevant conditional employ a ternary relation R on worlds. But given worlds w, x, y, what does it mean to say that $Rwxy$? And does it have anything to do with conditionality?

In some sense, this is a worry about the nature of the things that are supposed to correspond to the points of relevant logic. We called these things worlds. Only by making sense of these worlds, and how they relate to each other, can we say that we have given a plausible logical analysis of the English conditional. Recall that the logical analysis of possibility and necessity makes use of worlds, and it is important to make some sense of what these worlds are, and how they relate to each other. In the last chapter we examined several competing accounts as to what these worlds could be — concrete objects, sets, states of affairs, etc. We didn't have to choose one over the others. The fact that there are plausible candidates is enough to make sense of the logical analysis.

In terms of relevant logic and the logical analysis of the English conditional, we need not say definitively what these worlds are or how they relate to each other. But it would be nice to have at least one plausible story of how this could go. One possible interpretation associates the worlds of relevant logic with information states. Given that these worlds are used primarily to give the truth conditions for the relevant conditional, the accessibility relation $Rwxy$ holds between states of information when the information in state w is applied to the information in state x to yield the information in state y. This interpretation is clearly described by Greg Restall (1999):

> $Rwxy$ holds when all of the information given by x, after applying the information in w, holds in y. You are to think of the information supported by x as 'data,' and that supported by w as information to be applied to the data. If the results are all supported by y, then $Rwxy$ holds. For example, w might be a state which includes laws of physics, and x might be an initial state of a system. Then if y includes all consequent states of the system, we would take it that $Rwxy$ holds.

Restall, echoing Lewis, says that there are many different ways the world could be. Restall also thinks there are many different ways the world couldn't be. And he thinks there are ways that parts of the world could or couldn't be. He takes these 'ways' to be characterized by states of information. Some states are complete, in that for every A, they represent either A or $\neg A$. Some states are incomplete. Some states are consistent, in that for no A do they represent both A and $\neg A$. Some states are inconsistent. The complete states are worlds. The complete and consistent states are possible worlds.

Restall's interpretation is certainly plausible. And it is relatively light in terms of ontological commitment. He has made no claim as to what states are. But whatever they are, the relationship of applying information to a set of data to yield results is quite natural and easy to comprehend. It even sounds something like that an indicative conditional is supposed to do.[13]

With the first worry about the relevant conditional answered, we turn to the second worry about relevant negation. Just as we needed a plausible interpretation of the relevant conditional, and the accessibility relation that governs it, it seems we also need a plausible interpretation of relevant negation, and the device that governs it — the Routley star. The truth conditions for relevant negation invoke pairs of points:

$$v_w(\neg A) = \mathbf{t} \quad \text{iff} \quad v_{w^*}(A) = \mathbf{f}$$

The worry is whether this is an acceptable conception of negation. If the symbol $\neg$ is supposed to represent some plausible conception of negation, then why should the fact that $\neg A$ holds at one point have anything to do with what goes on at other points?

In fact, relevant negation corresponds quite closely to how negation could intuitively work in a framework of information states that can be complete or incomplete, consistent or inconsistent. Simply because a state does not satisfy A, we cannot conclude that it satisfies $\neg A$, because the state may be incomplete with respect to A. And given that states of information can be inconsistent, we should allow that at some states both A and $\neg A$ hold. So there is no reason to think that negation in states of information should be classical.

The question, then, is how a state w and its partner w^* relate to one another. The answer is to be found in a conception of compatibility.[14] If a state w makes $\neg A$ true, then any state w' that makes A true is incompatible with w. Any state that does not make A true is compatible with w, as far as A goes. Starting with the idea of compatibility between states, one plausible analysis of negation would be the following:

$$v_w(\neg A) = \mathbf{t} \quad \text{iff} \quad \text{for every } w' \text{ such that } w \text{ is compatible with } w', v_{w'}(A) = \mathbf{f}$$

Suppose that $v_w(\neg A) = \mathbf{t}$. Then any state that is compatible with w could not make A true. So every state compatible with w must make A false. Conversely, suppose that every state compatible with w makes A false. There is then no state compatible with w that makes A true. In some sense

[13] For other interpretations of the ternary R relation, see Beall et al. (2012).
[14] See Restall (1999) and Beall & Restall (2006).

then, state w "rules A out". You could then take this as reason to suppose that $\neg A$ is true at state w, i.e., $v_w(\neg A) = \mathbf{t}$.

Given an intuitive understanding of compatibility, we might want to impose conditions on the formal relation that we are describing. For example, we might say that compatibility is symmetric — if w is compatible with w', then w' is compatible with w. Beyond its intuitive appeal, symmetry is desirable, as it follows from symmetry that for every state x, if $v_x(A) = \mathbf{t}$, then $v_x(\neg\neg A) = \mathbf{t}$.

We might also want to say that the compatibility relation is directed in the sense that, for every w there is a w' such that w is compatible with w'.[15] That is, for any state there is some state that it is compatible with. As long as we have no trivial state (or if we allow both the trivial state and the vacuous state), then this condition seems reasonable.

Lastly, we might want the relation of compatibility to be convergent — if there is some state compatible with w, then there is a maximal state that is compatible with w. That is, there is a state w' such that w' is compatible with w, and every state that is compatible with w is included in w'. State x is included in state y iff if $v_x(A) = \mathbf{t}$, then $v_y(A) = \mathbf{t}$. Convergence is appealing, as it follows from convergence that for every state x, if $v_x(\neg(A \vee B)) = \mathbf{t}$, then $v_x(\neg A \wedge \neg B) = \mathbf{t}$.

It is reasonable to think that compatibility is symmetric, directed and convergent. But it has been shown that if the relation of compatibility satisfies these conditions, then the clause

$$v_w(\neg A) = \mathbf{t} \quad \text{iff} \quad \text{for every } w' \text{ such that } w \text{ is compatible with } w', v_{w'}(A) = \mathbf{f}$$

is equivalent to the clause

$$v_w(\neg A) = \mathbf{t} \quad \text{iff} \quad v_{w^*}(A) = \mathbf{f}$$

where w^* is the maximal state that is compatible with w.[16] So while relevant negation is certainly different from classical negation, it is an understanding of negation that is plausible given a framework of information states.

Supposing that this interpretation is a suitable account of relevant negation, we now have an argument in favor of the existence of at least one kind of impossible world. The semantics of relevant logic give a plausible representation (a "model") of a conditional that is relevant and that does not validate EFQ. Arguably, there is a version of the English indicative conditional that these

[15] We have adopted the language of Restall (1999), though this is slightly different from the usual meaning of directed.
[16] See Dunn (1994) for the details.

semantics accurately capture. If the truth conditions given by relevant logic are correct, because the models of relevant logic make use of points, these points must correspond to things that exist, things that are "out there" somewhere. Call those things worlds. Some of the worlds (those that correspond to non-normal points) are logically impossible; they break the laws of logic. The case is strengthened by the fact that we can give a plausible account of the worlds as information states, and of the accessibility relation R and the Routley star. There is then reason to think that these impossible worlds exist.

Epistemic and Doxastic Logic. The argument in favor of the existence of impossible worlds has so far claimed that we should believe in impossible worlds because they are required to give a logical analysis of one kind of conditionality that is captured by the English indicative conditional. But even if you reject the above argument, it seems that there are other concepts whose logical analyses require an appeal to impossible worlds. In this section, we focus on the logic of knowledge and belief.[17]

Besides providing a model for possibility and necessity, normal modal logic has historically been used to model epistemic states such as knowledge and belief.[18] Using normal modal frames, we can introduce a modal operator B_a for belief, and a modal operator K_a for knowledge. We interpret B_aA as meaning agent a believes that A; and we interpret K_aA as meaning agent a knows that A.[19] Both operators work very much like the necessity operator $\Box$, and their truth conditions are the same.

$$
\begin{array}{lll}
v_w(B_aA) = \mathbf{t} & \text{iff} & \text{for all } w' \in W \text{ such that } wRw', v_{w'}(A) = \mathbf{t} \\
v_w(K_aA) = \mathbf{t} & \text{iff} & \text{for all } w' \in W \text{ such that } wRw', v_{w'}(A) = \mathbf{t}
\end{array}
$$

Normally, one would work with a single epistemic operator at a time, developing a logic for belief and a logic for knowledge separately. The $w \in W$ are taken to represent *epistemically* possible worlds if you are working with knowledge, *doxastically* possible worlds if you are working with

[17] To some degree, in the last section, we argued that classical logic is not the correct logic of the English indicative conditional. In this section, for simplicity, we assume that it is.
[18] See Hintikka (1962).
[19] The accessibility relation R should be parameterized with a as well.

belief. In a logic of belief where wR_aw', w' is a world that is compatible with what a believes at w. That is, everything that a believes at w is true at w'. Similarly for a logic of knowledge.

The modal logics of belief and knowledge both suffer from the problem of logical omniscience. The problem of logical omniscience is that this understanding of belief (knowledge) implies that one believes (knows) all of the logical consequences of his or her beliefs (knowledge). That the modal logics of belief and knowledge suffer from this problem is argued for as follows.

(P1) Suppose that agent a believes A.

(P2) A holds at all of a's epistemic possibilities, i.e., a's epistemically possible worlds.

(P3) Suppose that B is a logical consequence of A.

(P4) So, $A \supset B$ holds in every logically possible world.

(P5) Every epistemic possibility is a logical possibility.

(P6) So, $A \supset B$ holds in all of a's epistemic possibilities.

(P7) So, B holds in all of a's epistemic possibilities.

(C) So, a believes B.

Clearly we do not believe (know) all of the logical consequences of what we believe (know). If we did, then we would believe (know) every necessary truth, including every instance of every valid formula, every mathematical truth, every metaphysical truth. But we don't believe (know) these things.

An instance of the problem of logical omniscience is particularly troubling for the logic of belief. For it may be that one's beliefs are inconsistent.[20] For some A, a may believe both A and $\neg A$. It may even happen that a believes $A \wedge \neg A$, though it is possible for one to believe two conjuncts while failing to believe their conjunction. If a believes both A and $\neg A$, then a does not have any epistemic possibilities, given that all epistemic possibilities must be logically possible. Without any epistemic possibilities, a's beliefs explode — a believes every proposition. For any proposition, that proposition is true in all of a's epistemically possible worlds (show me one where it is false).

[20]This is not such a problem for the logic of knowledge because knowledge is usually taken to be factive. It is therefore far more controversial to claim that someone can *know* a contradiction, as that would imply that some contradiction is true.

There are not many responses one could make to this argument. For some time, this argument was seen as an argument against a worlds analysis of belief. From this perspective, premise (P2) ought to be rejected. Such a response is warranted. But if one wants to keep the worlds analysis, one must reject some other premise. Premises (P1) and (P3) are suppositions not open to question. Lines (P4), (P6), (P7), and (C) follow from previous lines. The only plausible candidate is premise (P5): every epistemic possibility is a logical possibility.

It is not clear that this should be so. We have seen that a can have contradictory beliefs A and $\neg A$. A world that is consistent with everything that a believes would be one where both A and $\neg A$ hold. Such a world is not logically possible. Or it may be that a believes A and believes B, but fails to "believe them together", i.e., fails to believe $A \wedge B$. Any world compatible with a's beliefs in these cases is not a logically possible world.

It is even easier to construct metaphysically impossible worlds based on a's beliefs. It may be that a believes that Fermat's Last Theorem is false, or that $2+2=5$, or that the empty set doesn't exist, or that Charles is a married bachelor. Any world compatible with these beliefs is metaphysically impossible. The class of interpretations that give the best representation of belief (and knowledge) will have to include interpretations with points that correspond to impossible worlds. And if that class of interpretations is to represent belief accurately, those impossible worlds must exist.

That concludes the first stage of the argument in favor of the existence of impossible worlds. There are a handful of concepts for which we can give a plausible logical analysis, a mathematical representation of how these concepts work. These representations include reference to points. If the representations are to be accurate, if they are to give the correct account of how these concepts work, these points must correspond to things that exist "out there" somewhere. Our logical analysis of possibility and necessity associates these points with possible worlds. The success of this representation suggests that we should think that possible worlds exist, whatever they may be. Our logical analyses of one version of the English indicative conditional, and of our concepts of knowledge and belief, associate some points with impossible worlds. The success of these representations suggests that we should think that impossible worlds exist, whatever they may be.

1.2. Philosophical Precedent. We have considered arguments in favor of impossible worlds that appeal to their use in the logical analysis of concepts that hook up with important features of the world. We turn now to arguments that appeal to the use of impossible worlds in the practice of philosophy.

Philosophical Debate. The most compelling philosophical reason to believe in impossible worlds is that appeal to impossible worlds is implicit in practically every area of philosophical reflection and debate. Philosophical debates in metaphysics, mathematics, even ethics, where the debates are genuinely between points of view that conflict, often require one to consider situations that are impossible.[21] Consider two metaphysicists, one who endorses a Spinozistic metaphysical picture, and one who endorses a Leibnizian metaphysical picture. The Spinozan will say that there is one and only one simple substance. It is infinite and eternal, and can be identified with God. The Leibnizian will say that there are many simple substances, the monads. They are the atoms of nature, and they come together to compose everyday objects that we are familiar with.

The Spinozan and the Leibnizian have a genuine disagreement about the basic nature of physical reality. The Spinozan holds that there is only one simple substance. The Leibnizian holds the opposing position that there are more than one, indeed there are many, simple substances. It is plausible to think that whether there is exactly one simple substance, many simple substances, or no simple substances, is something that holds of metaphysical necessity. It must be, then, that at least one of these metaphysicists is arguing from a position that is metaphysically impossible.

Consider another discussion between two philosophers interested in the foundations of mathematics. In set theory, there is a serious debate as to whether the Continuum Hypothesis (CH) is true. It is independent of the standard axioms of set theory. However, CH is not an obvious set-theoretic truth, in that it does not clearly follow from the generally accepted, iterative conception of set. If we are realists about mathematical truth, then CH is either true or false. Two mathematicians can genuinely debate over whether or not it is true. As mathematical truths are often taken to be necessarily true, it follows that one mathematician is arguing from a position that is impossible.

[21] Daniel Nolan (1997) makes this point.

Consider, finally, two ethicists: one a utilitarian, the other a Kantian. The utilitarian believes you should act in a way so as to maximize the overall good of society, where good is understood in terms of happiness or pleasure. And you should so act because the action that maximizes the overall good of society is the action that is morally right. The Kantian, on the other hand, believes that you should act in a way so that you could consistently will that everyone, when in similar situations, act in the same way as you. And you should so act because an action that you could consistently will that everyone choose when in similar situations is an action that is morally right. Both the utilitarian and the Kantian give conditions for the moral rightness of actions. It is plausible that conditions for the moral rightness of actions, if there are any, hold of metaphysical necessity. It follows, then, that at least one of these ethicists is arguing from a position that is metaphysically impossible.

These situations can be multiplied across many areas of philosophy. Philosophers are generally in the business of thinking about and debating over the status of propositions and theories which, if true, are necessarily true. If two philosophers take different sides in such a debate, then at least one of them will be arguing from an impossible position. To see why this might be problematic, let us return to the Spinozan and the Leibnizian, debating about simple substances. Let's suppose that both are wrong, that there are no simple substances, and that this holds necessarily. In this debate, both philosophers are arguing from impossible positions. But if there are no impossible worlds, then they may both appeal to claims that anything whatsoever follows from their opponent's position. For example, the Spinozan can claim that, had there been many simple substances, then it would be the case that $2+2=5$. And the Leibnizian can claim that, had there been only one simple substance, then there would have been a married bachelor. And these claims would be true![22]

[22] The situation is slightly more complicated for the mathematicians. Arguments in mathematics usually take the form of proof, which proceeds according to certain axioms and rules. Two mathematicians arguing about CH in the context of set theory will most likely proceed according to what is provable in ZFC. It probably isn't the case that, supposing CH is false, it follows from the axioms of ZFC that there exists a married bachelor. But we may be interested in what is true in these situations, not just what is provable. As proof does not exhaust truth in ZFC (by Gödel's incompleteness theorem), in discussions of truth, one may appeal to these kinds of claims about what is (trivially) true in certain situations.

What is problematic is that we are constantly reasoning about situations that are, in fact, metaphysically impossible. Without impossible worlds, reasoning under situations that are impossible ends up in triviality: anything follows from an impossible proposition. However, our reasoning about these situations is often principled, as it should be. We are not allowed to appeal to anything we like, justified by the claim that impossible situations make everything (trivially) true. How we reason about these situations must be coherent. And we must be able to say that, in these impossible situations, certain things just do not follow. Without impossible worlds, we cannot do this. Therefore, to engage in serious reasoning about counterpossible scenarios, we need impossible worlds.

Counterpossible Reasoning. The philosophical debates described above involve reasoning about situations that are impossible, and what follows from what in those situations. This is a form of counterpossible reasoning, involving counterpossible conditionals. A counterpossible conditional is a counterfactual conditional with an impossible antecedent. When evaluating counterfactuals with possible antecedents, we look at possible worlds where the antecedent is true and decide if the consequent is true at those worlds as well. If it is, then the counterfactual is true at the actual world. To evaluate a counterpossible, we should then look at worlds where the antecedent is true. But without impossible worlds there are no worlds like that. If there are no worlds where the antecedent of a counterfactual is true, then there are no worlds where the antecedent is true and the consequent is false — the counterfactual is vacuously true. Without impossible worlds, then, every counterpossible conditional is vacuously true, which is wrong.

It is wrong because there are plenty of counterpossibles that are false. Let us assume for the moment that classical logic is the correct logic, i.e., that it is the right theory of logical consequence. It is reasonable to think that, whatever logic is correct, the fact that it is correct holds of necessity. Now consider the following counterpossible conditionals.

- *Had intuitionistic logic been the correct logic, then the law of excluded middle would have failed.*
- *Had intuitionistic logic been the correct logic, then the law of non-contradiction would have failed.*
- *Had intuitionistic logic been the correct logic, then I would have given you \$1,000.*

The first counterpossible is true. In an intuitionistic setting, there are instances of the law of excluded middle that fail. And we know why it's true. The laws of intuitionistic logic simply do not guarantee that every instance of $A \vee \neg A$ holds. The counterpossible is not vacuously true; it is true because that's how intuitionistic logic works.

The second counterpossible is false. While instances of $A \vee \neg A$ fail in an intuitionistic setting, instances of $\neg(A \wedge \neg A)$ do not fail. Again, that's how intuitionistic logic works. As this counterpossible is false, it is certainly not trivially true. And the third counterpossible, I can assure you, is false.

That there are plausibly false counterpossible conditionals is a good reason to think that an analysis of counterfactuals based only on possible worlds is mistaken. The most intuitive way to account for false counterpossibles is to add impossible worlds to the possible worlds analysis of counterfactuals.

Unfortunately, not everyone agrees that some counterpossibles are false. In his *Counterfactuals* (1973, §1.6) Lewis maintains that he is comfortable with the trivial truth of all counterpossible conditionals. In fact, he has an argument to back up his claim. According to Lewis's preferred axiomatization for counterfactuals **VC**, a material conditional $A \supset C$ is equivalent to the corresponding counterfactual conditional $A \mathrel{\Box\!\!\rightarrow} C$, when A necessarily false, i.e., impossible.[23] Given that a material conditional with a (necessarily) false antecedent is true, it follows that the corresponding counterpossible is also true.

Given that any material conditional with a necessarily false antecedent is true, it follows that the corresponding counterpossible is also true. Lewis's axiomatization for counterfactuals is as follows.

[23] See Lewis (1973), p. 132 for the system **VC**.

Rules

R1. *Modus Ponens*

R2. *Deduction within Conditionals: for any $n \geq 1$,*

$$\frac{\vdash (X_1 \wedge \ldots \wedge X_n) \supset Y}{\vdash [(Z \mathrel{\Box\!\!\rightarrow} X_1) \wedge \ldots \wedge (Z \mathrel{\Box\!\!\rightarrow} X_n)] \supset (Z \mathrel{\Box\!\!\rightarrow} Y)}$$

R3. *Interchange of Logical Equivalents*

Axioms

A1. *Truth-functional tautologies*

A2. *Definitions of non-primitive operators*

A3. $X \mathrel{\Box\!\!\rightarrow} X$

A4. $(\neg X \mathrel{\Box\!\!\rightarrow} X) \supset (Y \mathrel{\Box\!\!\rightarrow} X)$

A5. $(X \mathrel{\Box\!\!\rightarrow} \neg Y) \vee (((X \wedge Y) \mathrel{\Box\!\!\rightarrow} Z) \equiv (X \mathrel{\Box\!\!\rightarrow} (Y \supset Z)))$[24]

A6. $(X \mathrel{\Box\!\!\rightarrow} Y) \supset (X \supset Y)$

A7. $(X \wedge Y) \supset (X \mathrel{\Box\!\!\rightarrow} Y)$

The equivalence between $A \supset B$ and $A \mathrel{\Box\!\!\rightarrow} B$, given that A is necessarily false can be derived in **VC** as follows.[25]

(1) Given that A is necessarily false, we have $\vdash \neg A$.

(2) From (1), we have $\vdash A \supset B$, for any B.

(3) By R2, $\vdash (Z \mathrel{\Box\!\!\rightarrow} A) \supset (Z \mathrel{\Box\!\!\rightarrow} B)$, for any Z.

(4) From (3), $\vdash (A \mathrel{\Box\!\!\rightarrow} A) \supset (A \mathrel{\Box\!\!\rightarrow} B)$.

(5) By A3, $\vdash A \mathrel{\Box\!\!\rightarrow} A$.

(6) From (4) and (5), by Modus Ponens, $\vdash A \mathrel{\Box\!\!\rightarrow} B$, for any B.

(7) From (6), $A \supset B \vdash A \mathrel{\Box\!\!\rightarrow} B$, for any B.

And so we have, given the assumption that A is necessarily false, $A \supset B$ logically implies $A \mathrel{\Box\!\!\rightarrow} B$. Conversely, we don't even need the assumption that A is necessarily false.

[24]Where $\equiv$ is the material bi-conditional.
[25]See Koslow (2011).

(1) $A \boxright B \vdash A \boxright B$.

(2) By A6, $A \boxright B \vdash (A \boxright B) \supset (A \supset B)$.

(3) From (1) and (2), by Modus Ponens, $A \boxright B \vdash A \supset B$.

And so we have that $A \boxright B$ logically implies $A \supset B$. Putting these two results together, under the assumption that A is necessarily false, $A \boxright B$ and $A \supset B$ are logically equivalent. Given this equivalence, we have an argument that all counterpossibles are (trivially) true. As any material conditional with a necessarily false antecedent is true, any counterfactual conditional with a necessarily false antecedent is true. So any counterfactual conditional with an impossible antecedent is true.

However, the fact that Lewis's **VC** makes all counterpossibles trivially true does not automatically justify the trivial truth of counterpossible conditionals. It simply formalizes the problem. It shows, in detail, why an axiomatization like **VC** is problematic. It gives us reason to reject **VC** as an adequate axiomatization of the counterfactual conditional.

If we are ultimately to accept that some counterpossibles really are non-trivially true, then it should also be that some things make them true. The most plausible truthmakers for counterpossible conditionals are impossible worlds. Just as possible worlds are the truth-makers of counterfactual conditionals with non-actual possible antecedents, impossible worlds should be the truth-makers of counterfactual conditionals with non-actual impossible antecedents.

Necessarily Co-Extensive Properties. As we briefly discussed in Chapter 3, one view of properties identifies them with their extensions across possible worlds. A property is identical to the collection of things that have that property in each possible world. The property of redness is identical to the set of things that contains all of the red things in this world and all of the red things in every other possible world. We noted that some properties hold of exactly the same things in all possible worlds, like the properties of being triangular and of being trilateral. Identifying properties with extensions forces one to say that being triangular and being trilateral are identical properties. But this is absurd. One says of a thing that it has three angles, while the other says of a thing that it has three sides.

One can avoid absurdity by appealing to impossible worlds. There are no possible worlds where something is triangular but not trilateral. But there are impossible worlds where such things exist. Taking properties to be extensions (or better, pairs of extensions and anti-extensions) across all worlds, both possible and impossible, will deliver the result that being triangular and being trilateral are not identical properties. Any view, therefore, that identifies properties with extensions should endorse impossible worlds.

You might think that appealing to impossible worlds to yield identity conditions on properties goes too far. There is, as Yagisawa noted, the worry that once impossible worlds are brought in, no property is identical to itself. But we were able to dispense with that worry. There is a related worry, however. You might think that the property "having 3 sides" is identical to the property "having 2 + 1 sides". But with impossible worlds, we could distinguish these.

It may be true that if you make a simple appeal to impossible worlds, then these properties will be distinct. However, there may be a way to suitably restrict the impossible worlds one considers, perhaps even on a case by case basis, to show that these properties are identical. The fact that some restriction on the impossible worlds one considers may be required in some cases does not refute the claim that at least some impossible worlds are needed to give a satisfactory account of properties as being identical to extensions across worlds.

Necessarily True/False Propositions. On some views, a proposition is a set, the members of which are exactly the worlds at which the proposition is true. The proposition "A talking donkey exists" is true in some worlds and false in others. Taking propositions to be sets of worlds, we can call this proposition D, where $D = \{w_1, w_2, \ldots\}$. In this example, none of these worlds is the actual world.

This view is open to the obvious criticism that it forces one to accept that all necessarily true propositions, which are true at every possible world, are identical. The proposition that expresses the law of non-contradiction is identical to the set of all possible worlds. The proposition that expresses the law of excluded middle is also identical to the set of all possible worlds. By the transitivity of identity, these two propositions are identical on this view. But that is absurd.

One who endorses this view must also accept that all necessarily false propositions, which are true at no possible world, are identical. "9 is even" is identical to "9 is prime", as they are both identical to the empty set of worlds. Again, this is absurd.

One can avoid absurdity by appealing to impossible worlds. Necessarily true propositions are identical because there are not enough possible worlds to discriminate between them. There are no possible worlds where they fail to be true. But there are plenty of impossible worlds where they fail to be true. There is an impossible world where the law of non-contradiction comes out true, but the law of excluded middle comes out false (a world governed by the laws of intuitionistic logic, for instance). Taking propositions to be sets of both possible and impossible worlds will allow us to distinguish between these two propositions.

Similarly for propositions that are necessarily false. Impossible worlds give us worlds where "9 is even" is true, but where "9 is prime" comes out false. Taking propositions to be sets of both possible and impossible worlds can distinguish between these two propositions. Taking this to the extreme, for any two distinct propositions, there is a world, possible or impossible, where one is true and the other is not. Such an extreme view is one way to ensure that distinct propositions really are distinct.

Given the similarities between a worlds account of properties and a worlds account of propositions, we should ask whether there is a Yagisawa style objection lurking here. Recall from Chapter 3 that Yagisawa's worry was that once we admit impossible worlds, no property will be identical with itself. Given impossible worlds, should we now worry that no proposition is identical with itself?

This is a difficult question because, while a property is identified with a set of things that exist in worlds (the union of its extensions in each world), a proposition is identified with a set of whole worlds (the worlds where it is true). Yagisawa's worry was that if it is impossible that properties P and Q not be co-extensive, then there exists an impossible world where P and Q are not co-extensive within that world. It is then supposed to follow that for every property P, it is not the case that $P = P$. The analogous worry is that if it is impossible that propositions A and B not be identical, then there is an impossible world where they are not identical. So the worry is that there is an impossible world w where proposition A is such that $A \neq A$. As propositions are sets of worlds, then according to w, there is a world v such that $v \in A$ and $v \notin A$.

Though similar, this objection is quite different from Yagisawa's, and so it does not seem that the same strategy easily applies. But we can resolve the worry as follows. While it may be true in some impossible world there exists a world v such that $v \in A$ and $v \notin A$, that need not imply that at the actual world, or at any possible world, there is a world v such that $v \in A$ and $v \notin A$. At the actual world it is still the case that $A = A$. And at every possible world it is still the case that $A = A$. Thus it is necessarily true that property A is identical with itself.

Before turning to arguments against the existence of impossible worlds, I want to point to one more argument in their favor. This argument is due to Yagisawa (1988); it is a variation on an argument that Lewis (1973) gives in favor of possible worlds. Lewis argues as follows.

> It is uncontroversially true that things might be otherwise than they are. I believe, and so do you, that things could have been different in countless ways. But what does this mean? Ordinary language permits the paraphrase: there are many ways things could have been besides the way they actually are. On the face of it, this sentence is an existential quantification. It says that there exist many entities of a certain description, to wit 'ways things could have been'. I believe that things could have been different in countless ways; I believe permissible paraphrases of what I believe; taking paraphrase at its face value, I therefore believe in the existence of entities that might be called 'ways things could have been'. I prefer to call them 'possible worlds' (p. 84).

Of course, Lewis's full argument is a bit more complicated, involving when a sentence may be taken at face value. A sentence may be taken at face value if one of two conditions are met:

- Taking it at face value is not known to lead to trouble, or
- Taking it some other way is known to lead to trouble (Ibid.).

If you accept Lewis's argument for possible worlds, there is a very similar argument in favor of impossible worlds that you should probably accept as well. There are countless ways things could *not* have been besides the way they actually are. So there exist 'ways things could not have been'. You may call them whatever you like. I prefer to call them 'impossible worlds'. And both this chapter and the previous chapter have provided arguments to show that impossible worlds do not lead to trouble. So claims of the existence of impossible worlds should be taken at face value.

This concludes our discussion of arguments in favor of the existence of impossible worlds. The logical arguments claim that there are several concepts, including a conception of conditionality, as often captured by the English "if … then" indicative conditional, and the concepts of knowledge and belief, whose logical analyses require the use of impossible worlds. If these analyses are to be correct, then those impossible worlds must represent something "out there" somewhere. Whatever they represent, I call those things impossible worlds.

The philosophical arguments make several claims. First, the implicit appeal to impossible worlds is a widespread phenomenon across many areas of philosophy. Many philosophical debates require the consideration of worlds that are impossible. We should therefore believe in the existence of impossible worlds. Second, the kind of philosophical debate described often involves the consideration of counterpossible conditionals. The importance of reflecting on counterpossible conditionals suggests that they are not all vacuously true. But to avoid the vacuous truth of counterpossibles, one must appeal to an analysis that invokes impossible worlds. In addition, impossible worlds save certain views of propositions and properties from absurdity. And then we have the Lewisian argument that claims to the existence of ways the world could not be should be taken at face value.

Given these considerations, there are good reasons to believe that impossible worlds exist. Yet many still balk at the idea of impossible worlds.

2. The Arguments Against

Anyone who mentions impossible worlds in philosophical conversation has most likely been met with looks of skepticism, perhaps even shock and horror. The current situation of impossible worlds is in this respect similar to the situation of possible worlds 30 years ago. The worry about impossible worlds (in some ways akin to the incredulous stare) is that some people claim they simply do not understand what impossible worlds are. This worry, in one form, is best articulated by Robert Stalnaker (1996).[26] Stalnaker worries about what exactly the distinction between possible and impossible worlds amounts to. There are just worlds. The phrase 'possible world', like the

[26] See also Kenneth Perszyk (1993).

phrase ‘existent entity’, is redundant. Something is possible if it is true of or exists in a world. What is the distinction between possible and impossible worlds supposed to be?

To discuss this question, Stalnaker’s paper takes the form of a dialogue. It is relatively clear who Stalnaker sides with, and it is this side who expresses the above worry. Stalnaker’s interlocutor responds that an impossible world is a world in which some contradiction is true. Notice that this differs slightly from how we have been thinking of impossible worlds so far. Impossible worlds are worlds that violate certain kinds of laws. Physically impossible worlds violate the actual world’s physical laws. Metaphysically impossible worlds violate the actual world’s metaphysical laws. The worlds that Stalnaker considers violate one of the actual world’s logical laws, the law of non-contradiction. On our view, however, there are impossible worlds, even logically impossible worlds, that do not violate this law.

Stalnaker’s rejoinder is that there are no worlds in which some contradiction is true. There are no worlds in which contradictions are true because their existence would imply that contradictions are true in the actual world. This is precisely Lewis’s worry about impossible worlds. In addressing this worry in Chapter 3, we argued that this objection begs the question against those who believe in logically impossible worlds as it assumes that negation works classically at every world.

One might, in the spirit of Quine, argue that one who endorses a form of “negation” that does not work classically is simply changing the subject. What we are talking about just isn’t negation. Stalnaker takes this line, claiming that he learned how negation works in his first logic class: $\neg A$ is true (at a world) iff A is false (at a world). Note that those who endorse impossible worlds with truth value gluts can accept this definition of negation by endorsing a truth-relational, as opposed to a truth-functional, semantics for negation. But classically A’s being false (at a world) implies that A is not true (at a world). What is at issue, however, is whether or not classical negation accurately models the relationship between truth and falsity at all worlds. Those who endorse impossible worlds claim that it does not, at least not at all worlds.

What I take to be Stalnaker’s main argument against impossible worlds invokes the idea that the intentional content of a proposition is simply its truth conditions. Possible worlds give us the conditions under which various propositions are true. In fact, for Stalnaker a proposition is identical to the set of worlds in which that proposition is true. But impossible propositions are not true under any conditions. That is what makes them impossible. We do not need impossible worlds to give the truth conditions for impossible propositions, because there are no conditions that make them true.

Stalnaker anticipates the objection that this view implies that impossible propositions are meaningless. He claims this is not his view. He takes the meaning of a proposition to be the "recipe" for discovering the conditions under which the proposition is true. One can know what this recipe is, and so know the meaning of the proposition, before computing the truth conditions, upon which one may discover that the proposition is true under no conditions.

However, there is a more serious objection that Stalnaker's view makes every impossible proposition identical. Take two impossible propositions P and Q. They are both impossible, so they have the same truth conditions: there are no conditions under which either one is true. That is, there are no worlds in which they are true. Identifying a proposition with its truth conditions, we get that $P = \emptyset = Q$. Every impossible proposition is identical with every other impossible proposition.

Clearly we do not want to identify all impossible propositions with each other. The impossible propositions "9 is even " and "9 is prime" say different things. They mean different things. They are different propositions, even though Stalnaker would identify them, as they are both true under no conditions, true in no possible worlds.

Stalnaker would also identify propositions that are necessarily true. The proposition that expresses the law of excluded middle and the proposition that expresses the law of non-contradiction are both necessarily true, they are each true in every possible world. They have the same truth conditions, and so they are the same proposition. But this can't be right, as they clearly say different things. One says that there are no truth-value gaps, the other that there are no truth-value gluts.

As we have seen, one of the best reasons to admit the existence of impossible worlds is to avoid the identification of all necessarily true (false) propositions with each other. Identifying propositions with sets of worlds, as Stalnaker would like to do, demands that for any proposition, impossible or not, there are some worlds where it is true and some worlds where it is false. If we would still like some propositions to be necessary, then we need impossible worlds.

As to the claim that one does not understand what impossible worlds are supposed to be, this kind of claim is hardly ever a justification for rejecting an idea. How can one reject the plausibility of something that one does not understand? Only after one understands what impossible worlds are can he or she be in a position to reject them. A lot of work here and elsewhere has been done to clarify the idea of an impossible world.[27] As seen in Chapter 4, there is no need to reject them on the grounds that they are incompatible with mainstream theories of possible worlds. And as seen in this chapter, there are several reasons to embrace them as a useful technical device and as required by the process of philosophical reflection and debate.

[27] See, in particular, the special issue on impossible worlds of the *Notre Dame Journal of Formal Logic*, Priest (1997a).

3. Appendix

In this Appendix, we show that in the relevant logic **R**, for any formula A, there exists an interpretation $\mathfrak{I}$ with a point w such that A is false at w according to $\mathfrak{I}$. In fact, we show the stronger result that there exists an interpretation $\mathfrak{I}$ with a point w such that every formula A is false at w according to $\mathfrak{I}$. To show this, we take an alternative, equivalent model theory of **R**, such that models are quintuples $\langle W, 0, R, *, v\rangle$, where W is a set of points and 0 is a designated point such that 0 makes all valid formulas true, R is a ternary accessibility relation on W, $*$ is the Routley star, and v is a valuation function. For these to be interpretations of **R**, the accessibility relation and Routley star must satisfy certain conditions. To state the conditions, we introduce three notational devices:

- $a \leq b$ iff $R0ab$
- $R^2(ab)cd$ iff $\exists x(Rabx \wedge Rxcd)$
- $R^2a(bc)d$ iff $\exists x(Raxd \wedge Rbcx)$

To be **R** interpretations, the accessibility relation and Routley star must satisfy the following seven conditions.

1. IDENTITY. *$R0aa$ for all points a.*
2. COMMUTATIVITY. *If $Rabc$, then $Rbac$.*
3. ASSOCIATIVITY. *If $R^2(ab)cd$, then $R^2a(bc)d$.*
4. IDEMPOTENCE. *$Raaa$ for all points a.*
5. MONOTONY. *If $Rabc$ and $a' \leq a$, then $Ra'bc$.*
6. PERIOD TWO. *$a^{**} = a$ for all points a.*
7. INVERSION. *If $Rabc$, then Rac^*b^*.*

The truth conditions for the connectives are:

$$
\begin{array}{lll}
v_w(A \wedge B) = \mathbf{t} & \text{iff} & v_w(A) = \mathbf{t} \text{ and } v_w(B) = \mathbf{t} \\
v_w(A \vee B) = \mathbf{t} & \text{iff} & v_w(A) = \mathbf{t} \text{ or } v_w(B) = \mathbf{t} \\
v_w(\neg A) = \mathbf{t} & \text{iff} & v_{w^*}(A) = \mathbf{f} \\
v_w(A \rightarrow B) = \mathbf{t} & \text{iff} & \text{for all } x, y \in W \text{ such that } Rwxy, \text{ if } v_x(A) = \mathbf{t}, \text{ then } v_y(B) = \mathbf{t}
\end{array}
$$

Consider an interpretation $\mathfrak{I}$ with two points $0, 1 \in W$. Suppose that point 0 makes every propositional parameter true and that point 1 makes every propositional parameter false. We show by induction on the complexity of formulas that point 0 makes every formula true and point 1 makes every formula false. It is straightforward that if A and B are true at 0, then $A \wedge B$ and $A \vee B$ are true at 0. Similarly, if A and B are false at 1, then $A \wedge B$ and $A \vee B$ are false at 1. To evaluate negated formulas, set $0^* = 1$, and $1^* = 0^{**} = 0$. If A is false at 1, then $\neg A$ is true at 0. If A is true at 0, then $\neg A$ is false at 1. To evaluate conditional formulas, set the accessibility relation as follows:

$$
\begin{array}{ll}
R000 & R100 \\
R010 & R101 \\
R011 & R110 \\
 & R111
\end{array}
$$

If A and B are true at 0, then $A \rightarrow B$ is true at 0. And if A and B are false at 1, then $A \rightarrow B$ is false at 1. It follows that point 0 makes every formula true (*a fortiori* point 0 makes every valid formula true as required), and that point 1 makes every formula false. Some diligent checking will confirm that R satisfies conditions $1-7$, thus making $\mathfrak{I}$ an interpretation of $\mathbf{R}$.

CHAPTER 6

The Counterfactual Analysis of Dependence

I have endorsed the existence of impossible worlds, whatever they may be, because the analysis of metaphysical dependence that I would like to give relies crucially on worlds where necessary truths fail to hold. The analysis of metaphysical dependence that I propose is a counterfactual analysis of dependence.

COUNTERFACTUAL ANALYSIS. *x metaphysically depends on y if, and only if, had y not existed, then x wouldn't have existed either.*

I take a worlds analysis of counterfactuals, as is standard. What is non-standard is that I extend the standard worlds analysis of counterfactuals with impossible worlds.

It is easy to see that, without impossible worlds, a standard worlds analysis of counterfactuals will give the wrong results when we apply a counterfactual analysis to objects like sets. In such a framework there are pairs of objects x, y such that it is true that had y not existed, then x wouldn't have existed either, but we do not want to say that x metaphysically depends on y. In other words, the condition, without impossible worlds, is not sufficient.

Consider, as we did in Chapter 3, Socrates and his singleton {Socrates}. We would like our analysis to confirm that {Socrates} depends on Socrates, and not vice versa. For sets should depend on their members, but on the whole members should not depend upon the sets that they are members of.[1] We must therefore evaluate the following counterfactual conditionals.

(1a) Had Socrates not existed, then {Socrates} wouldn't have existed either.

(1b) Had {Socrates} not existed, then Socrates wouldn't have existed either.

Suppose that we take a worlds analysis of counterfactuals that only makes use of possible worlds. Because Socrates and {Socrates} exist in all of the same possible worlds, both counterfactuals

[1] Although, see Section 2.6 of this chapter, on non-well-founded sets.

come out true. And so, by the counterfactual analysis of dependence, we have that {Socrates} depends on Socrates *and* Socrates depends on {Socrates}. These results conflict with our intuitions that members of sets should not depend on the sets that they are members of.

The problem is further illustrated if we consider pure sets, like $\emptyset$ and $\{\emptyset\}$. Pure sets, as they are pure mathematical objects, are said to exist necessarily, if they exist at all. Nevertheless, we would like it to come out that $\{\emptyset\}$ depends on $\emptyset$, and not vice versa. For sets, whether they are pure or impure, should depend on their members, and not the other way around. Accordingly, we must evaluate the relevant counterfactual conditionals.

(2a) Had $\emptyset$ not existed, then $\{\emptyset\}$ wouldn't have existed either.

(2b) Had $\{\emptyset\}$ not existed, then $\emptyset$ wouldn't have existed either.

Under a worlds analysis of counterfactuals, with only possible worlds, both (2a) and (2b) come out true, and trivially so. For there are no worlds where either of these sets fails to exist. So, again, we have the counterintuitive results that members depend on the sets to which they belong.

But the situation is worse than all this, for the counterfactual analysis implies dependence relations that are even more implausible. If there are no worlds where, for example, the empty set fails to exist, then *any object whatsoever* will metaphysically depend on the empty set. Picking Socrates as an arbitrary example, the following counterfactual is true.

(2c) Had $\emptyset$ not existed, then Socrates wouldn't have existed either.

This counterfactual is true, for the same trivial reason that (2a) is true: there are no possible worlds where $\emptyset$ fails to exist. The counterfactual analysis entails that Socrates metaphysically depends on the empty set. As Socrates was an arbitrary choice, every object metaphysically depends on the empty set. To some degree, the empty set was an arbitrary choice as well. We could have put any necessarily existing object in its place, and the resulting counterfactual would have been true. Thus, any object whatsoever metaphysically depends on any necessarily existing object.

To solve these problems, I would like to extend the worlds analysis of counterfactuals with impossible worlds. The project is to show that, with impossible worlds, a worlds analysis will make the counterfactuals (1a) and (2a) true, and counterfactuals (1b), (2b), and (2c) are false, as desired.

A few words, however, should be said about the notion of existence that I employing. A central question of metaphysics asks how we should understand the concept of existence. Some say that existence is determined by quantification: to be is to be the value of a variable, perhaps in our best theory of the world (Quine 1953). Others say that to exist is to be the referent of a singular term in a true sentence (Wright 1983). Perhaps to exist is just to be located in spacetime, or to be causally efficacious.

I do not endorse any particular theory of existence. I take it that we have a general understanding of existence, according to which I exist, and you exist, and Socrates exists (or did exist). I assume, for the sake of argument, that certain mathematical objects, like numbers and sets, exist in this sense as well. However, they do so at every metaphysically possible world. Unfortunately, I don't exist at every possible world, and neither do you.

However, the existence of sets is sometimes spoken of in a different way. As sets are mathematical objects, their existence can be spoken of as relative to a theory. For example, the empty set $\emptyset$ exists according to ZFC, and so does $\{\emptyset\}$. And these objects exist according to other (set) theories as well, like NBG. However, some things exist according to NBG which do not exist according to ZFC, like proper classes.

You might think that existence according to a theory is more precise than the more general notion of existence that I am appealing to. But it's not entirely clear that it is. For even if one would like a notion of existence according to a theory, one has serious choices to make. Does something exist according to a theory when it can be proven in that theory to exist? When its existence is consistent with the theory? When its existence is consistent with the theory relative to the consistency of some other theory?

I prefer instead to appeal to a general, unanalyzed conception of existence. That is not to say that I take this conception of existence to be unanalyzable. It is just that I do not offer an analysis of it. I take as assumptions that concrete objects, including Socrates, exist, and that some numbers and sets exist. Which numbers and sets? The only numbers that I assume to exist are the natural numbers — they are discussed in Section 4 of this chapter, which examines dependence claims made by the mathematical structuralist. The argument in favor of the counterfactual analysis of dependence, extended with impossible worlds, works by way of example. I show, for various sets, that according to the extended counterfactual analysis, sets metaphysically depend on their members, and not vice versa. So I assume certain sets to exist, in particular, those sets that serve as my examples. These include: the impure set {Socrates}, the pure sets $\emptyset$, $\{\emptyset\}$, $\{\{\emptyset\}\}$, the first limit ordinal ω, and the non-well-founded sets $a = \{a\}$ and $b = \{c, b\}$, and all those other sets that these sets metaphysically depend on. We do not argue for the existence of these sets. We assume they exist in order to examine the claim that they depend on their members.

There are some sets the existence of which I do not make assumptions about. In particular, I do not assume that large cardinals exist. However, given the nature of large cardinals, there seems no reason to think that the arguments presented in the following sections cannot be modified so as to apply to large cardinals. And so, if large cardinals exist, then there is hope that the counterfactual analysis can show that they depend on their members. Furthermore, I do not assume that there exists a set S such that $\aleph_0 < S < 2^{\aleph_0}$. That is, I do not assume any particular answer to the continuum hypothesis. However, with respect to large cardinals and sets like S, as they are purely mathematical objects, I take their existential status to be non-contingent. If they exist, then they exist necessarily; if not, then necessarily not.

1. Conditional Logic

The contemporary approach to counterfactuals gives their truth conditions in terms of worlds. This approach often appeals to some kind of similarity or resemblance relation between possible worlds. Consider Lewis's account.

> '*If kangaroos had no tails, they would topple over*' seems to me to mean something like this: in any possible state of affairs in which kangaroos have no tails, and which resembles our actual state of affairs as much as kangaroos having no tails permits it to, the kangaroos topple over (1973, p. 1).

To evaluate the conditional, we look to those possible worlds most similar to the actual world where the antecedent is true and see if the consequent is true in those worlds. But accounts like Lewis's struggle to come up with a plausible understanding of when one world is similar to another.

One need not, however, frame the discussion in terms of similarity. One could instead appeal to *ceteris paribus* clauses. To evaluate a counterfactual we look at worlds where, "other things being equal" the antecedent is true. To do this, we look at worlds where we hold some things fixed, and where the antecedent is true. These worlds differ only in what is required to make the antecedent true. In a sense, we are looking at worlds that change just enough for the antecedent to be true, and hold everything else fixed. The *ceteris paribus* clause encompasses those things we keep fixed.

Given a counterfactual "Had A occurred, then B would have occurred", the antecedent A gives rise to a set of things that we keep fixed. That is, for each antecedent A of a counterfactual, we have a set of propositions C_A such that it must be the case that any world we consider in the evaluation of the counterfactual makes A true and makes every member of C_A true as well. That the propositions in C_A are true at these worlds embodies the fact that the worlds "keep other things equal". They only differ in what is required to make the antecedent of the counterfactual true.

The *ceteris paribus* approach is analogous to the similarity approach. If certain things are kept fixed across worlds, through *ceteris paribus* clauses, then the worlds are similar to each other, at least in some respects. But the idea of similarity is a vague one. There are many ways in which one thing might be similar to another. If we evaluate counterfactuals in terms of *ceteris paribus* clauses, ideally we know precisely what differs and what remains fixed from world to world. The idea is that each counterfactual gives rise to a set of worlds that are *ceteris paribus* like the actual world.

We can model this treatment of counterfactuals with a conditional logic. Conditional logic is a modal logic.[2] The language of conditional logic includes the usual classical propositional connectives: $\wedge$ (and), $\vee$ (or), $\neg$ (not), $\supset$ (material conditional). We add a two place connective $\Box\!\!\rightarrow$. Interpretations of the language of conditional logic are ordered triples $\langle W, \{R_A : A \in \mathscr{F}\}, v\rangle$. W is a set of points, and v is a valuation function, as usual. The set $\{R_A : A \in \mathscr{F}\}$ is a set of accessibility relations between points, one for each formula A in the set of formulas $\mathscr{F}$ of our language. If wR_Aw', then we say that w' is "A-accessible" from w.

[2]A more detailed description of conditional logics can be found in Priest (2008), ch. 5.

The points semantics are used to give truth conditions for the conditional $\Box\!\!\rightarrow$. Truth conditions for the other connectives are classical (though relativized to points).

$v_w(A \Box\!\!\rightarrow B) = T \quad \text{iff} \quad \text{for all } w' \text{ such that } wR_Aw', v_{w'}(B) = T.$

These basic semantics yield the conditional logic **C**. We can add several conditions that may be desirable.

(1) The set of worlds w' such that wR_Aw' make A true.

(2) If A is true at w, then wR_Aw.

(3) If there is a world that makes A true, then there is a world w' such that wR_Aw'.

(4) If every A-accessible world from w makes B true, and every B-accessible world from w makes A true, then the set of A-accessible worlds from w is identical to the set of B-accessible worlds from w.

(5) If there is some A-accessible world from w that makes B true, then the set of $(A \wedge B)$-accessible worlds from w is a subset of the A-accessible worlds from w.

Adding conditions (1) – (2) yields the conditional logic $\mathbf{C}^+$. Also adding (3) – (5) gives the logic **S**. Further conditions give the familiar conditional logics of Lewis and Stalnaker.

(6) If there is any world w' such that wR_Aw', then it is the only w' such that wR_Aw'.

(7) If A is true at w, then w is the only w' such that wR_Aw'.

Adding condition (6) alone gives the conditional logic of Stalnaker $\mathbf{C}_2$. Stalnaker's $\mathbf{C}_2$ is stronger than Lewis's conditional logic $\mathbf{C}_1$, which one gets by adding condition (7) alone.[3]

One can argue about which conditions are appropriate. But all of these logics are trying to capture the idea that if wR_Aw', then w' is *ceteris paribus* like w in the context of the formula A. To evaluate conditionals $A \Box\!\!\rightarrow B$ at a world w, we look at worlds that are *ceteris paribus* like w with respect to A. That is, we look at worlds that keep as much as possible the same as w, except that they make A true. So the formula A will then give rise to a set of propositions C_A that should be kept fixed in the evaluation of a counterfactual that has A as an antecedent. Any world w' such that wR_Aw' should keep each formula in C_A true.

[3] $\mathbf{C}_1$ is another name for the Lewis system **VC** discussed in Chapter 5.

Of course, if A is not true at w, then making A true will consequently make some things at w' different from w. Making A true will have consequences. So whatever those consequences are, those are true at the worlds under consideration as well. And that is how these worlds differ from w. But everything else stays the same.

2. Set-Theoretic Dependence

We are interested in considering what would happen had certain things — sets — not existed. To see what would happen in these scenarios, we examine counterfactual conditionals. And to evaluate these counterfactual conditionals, we consider worlds that are *ceteris paribus* like the actual world. They are worlds that differ only in what is required to make the antecedents of the counterfactuals come out true. Nothing more, nothing less. We then see if the consequents are true in these worlds as well. For if they are, then the counterfactuals are true at the actual world. Given the counterfactual analysis of dependence, the truth of these counterfactuals will yield true claims of metaphysical dependence.

We begin with the case of Socrates and his singleton {Socrates}.

2.1. Socrates and {Socrates}. Consider the counterfactuals:

(1a) Had Socrates not existed, then {Socrates} wouldn't have existed either.

(1b) Had {Socrates} not existed, then Socrates wouldn't have existed either.

We would like (1a) to come out true and (1b) to come out false. To evaluate (1a) we must look at worlds that make the antecedent true. So we are looking at worlds that are *ceteris paribus* like the actual world, but which make it so that Socrates does not exist. If we were only considering metaphysically possible worlds, then every such world would make it that {Socrates} does not exist either. For Socrates and {Socrates} exist in all of the same possible worlds. But now we are allowing impossible worlds. And it may be that some impossible worlds are *ceteris paribus* like the actual world. The question is: are there any such impossible worlds where Socrates doesn't exist but {Socrates} does?

I argue that there are not. These worlds are ruled out by the conditions that govern which worlds are *ceteris paribus* like the actual world. We have to figure out what are the minimal conditions that must change in order to make it true that Socrates does not exist. This is not an easy question. I'm sure there are many things that must change. And I'm sure that one could not list all of them. However, I think we can be confident about some things that do not change.

Something that should not change are the identity conditions for sets. Identity conditions for sets say when two sets are identical to each other. In axiomatic set theory, these conditions are captured by the Axiom of Extensionality.

EXTENSIONALITY. *For all sets x and y, if it's true that for all z, $z \in x$ iff $z \in y$, then $x = y$.*

In short, two sets are identical if they have exactly the same members. As we are in a modal context, we should note that we take Extensionality to be a necessary truth — it is true in all metaphysically possible worlds.

It is clear that, in the worlds we are considering, nothing should change about the identity conditions of sets. These worlds are just like the actual world, except that Socrates does not exist. Nothing about the fact that Socrates no longer exists at these worlds requires that Extensionality fails. Conversely, the truth of Extensionality does not require the existence of Socrates.

Given that Extensionality holds, we know what the identity conditions for sets are at these worlds. We can ask whether there is anything in this world that is identical to the set {Socrates}. I argue that there isn't. In order for anything to be identical to {Socrates} it must have Socrates as a member. But Socrates does not exist at these worlds. So nothing at these worlds can have Socrates as a member. So nothing at these worlds is identical to {Socrates}. And so, {Socrates} does not exist at these worlds.

One might object along the following lines. Suppose that, in the worlds under consideration where Socrates does not exist, {Socrates}$=$ $\{x : x =$ Socrates$\}$. If Socrates does not exist, then {Socrates} $= \emptyset$ at these worlds. And even though Socrates does not exist, the empty set may still exist. However, worlds where {Socrates} $= \emptyset$ would violate the necessity of identity. To have {Socrates} $= \emptyset$ in these worlds would imply that {Socrates} $= \emptyset$ in the actual world, which is false. So these worlds should not make it that {Socrates} $= \emptyset$.[4]

[4]Thanks to Arnie Koslow and Kit Fine on this point.

This argument makes use of an implicit premise that says the relation of set-membership at a world is an existence entailing relation: if $x \in S$ at world w, then x exists at w. To see that this is true, suppose that the set-membership relation was not an existence entailing relation. More specifically, suppose there is a set S with members $s_1, s_2, s_3, \ldots$ such that S exists and all of the members of S exist, except for s_1. Then there could be a distinct set S^*, with members $s_2, s_3, \ldots$. There would then be two distinct sets with the same existent members, thus violating the axiom of Extensionality.[5] One might argue that this may be a world where Extensionality fails. But we are only considering worlds that differ only in what is required for it to be true that Socrates does not exist. Removing Socrates does not require that Extensionality fails.

It seems then that (1a) is true: Had Socrates not existed, then {Socrates} wouldn't have existed either. It follows, given the counterfactual analysis of dependence that {Socrates} metaphysically depends on Socrates — the set depends on its members. But we must also ensure that (1b) comes out false, so that Socrates does not metaphysically depend on {Socrates}.

Consider worlds where {Socrates} does not exist. What must change to make this happen? To answer this question, we have to ask why we think the singleton exists in the first place. We believe the singleton of Socrates exists because we think that some more general principle holds: whenever something exists, so does its singleton. Socrates exists. So his singleton exists as well. In ZF set theory, that this principle holds is implied by the Pairing axiom.

PAIRING. *For all sets x and y, there exists a set z such that for all u, u is a member of z if, and only if, u = x or u = y.* $\forall x \forall y \exists z \forall u (u \in z \leftrightarrow u = x \vee u = y)$

For simplicity, here and in the following sections, let us take Pairing to be the principle that implies the existence of singletons. Pairing implies the existence of the singleton of Socrates. Just consider the case where x and y are both Socrates. If we are looking at worlds where the singleton does not exist, then Pairing must fail to hold at these worlds.

Does the failure of Pairing imply that Socrates fails to exist? Clearly not. The existence of Socrates does not require that Pairing is true. The existence of Socrates has nothing to do with the

[5] Essentially the same argument is give by Fine (1981).

Pairing axiom. It seems, then, that (1b) above is false. And so Socrates does not metaphysically depend on his singleton.

One might object: Why don't we just get rid of Socrates? Perhaps the simplest way to have a world where {Socrates} doesn't exist is to remove Socrates. Given the arguments discussed concerning (1a) above, removing Socrates will ensure that {Socrates} no longer exists. And perhaps a world like this is *ceteris paribus* like the actual world.

What is required, however, is that *every* world that is *ceteris paribus* like the actual world, where {Socrates} doesn't exist, also makes it the case that Socrates does not exist either. And it's not clear that this is true. There may be some worlds (that are *ceteris paribus* like the actual world) where they both fail to exist. But we have no reason to think that all of them are like that. At the very least the burden is on my opponent to argue that there are no worlds that are *ceteris paribus* like the actual world where Socrates exists but {Socrates} doesn't. In order to evaluate (1b), we are looking at worlds that differ only in what is required to make it that {Socrates} doesn't exist. We recognize that

(a) Socrates exists.

(b) Pairing is true.

together imply that {Socrates} exists. If we are looking at worlds where {Socrates} fails to exist, then at least one of (a) or (b) must fail to be true at these worlds. But there seems to be no principled way to choose one over the other. Of the worlds that are *ceteris paribus* like the actual world, but which make it that {Socrates} doesn't exist, some may make (a) false and (b) true, some may make (b) false and (a) true, or some may make both false. As long as there is one world that makes (a) true, then the counterfactual (1b) is false. And there is no reason to think that there isn't such a world.

It is apparent that we are now dealing with impossible worlds. As sets and their members exist in all of the same possible worlds, any world that has the members without the set is a metaphysically impossible world. It breaks a law of metaphysics, namely, the Pairing axiom.

But we can show that the counterfactual analysis with impossible worlds gives the correct answers even if we do not need to appeal to them. Let us first examine the set-theoretic cases.

Take, for example, the set {Socrates, Plato} and its members Socrates and Plato. As usual, it should be that the set metaphysically depends on each philosopher, but that neither philosopher metaphysically depends on the set. Consider the relevant counterfactuals.

(1c) Had Socrates not existed, then {Socrates, Plato} wouldn't have existed either.

(1d) Had Plato not existed, then {Socrates, Plato} wouldn't have existed either.

(1e) Had {Socrates, Plato} not existed, then Socrates wouldn't have existed either.

(1f) Had {Socrates, Plato} not existed, then Plato wouldn't have existed either.

We would like (1c) and (1d) to be true, (1e) and (1f) to be false. To see that (1c) is true, consider worlds that make the antecedent true. We have no reason to think that these are impossible worlds. They are worlds that are *ceteris paribus* like the actual world, but which differ only in what is required to make it that Socrates does not exist. In particular, what need not change are the identity conditions for sets. Can anything in these worlds be the set {Socrates, Plato}? Clearly not, as the identity conditions for sets require that, for anything to be {Socrates, Plato}, it must be related to Socrates through the inverse membership relation. As Socrates does not exist at these worlds, nothing can be related to him in this way. And so nothing in these worlds is the set {Socrates, Plato}. The counterfactual (1c) is therefore true. According to the counterfactual analysis of dependence, the set {Socrates, Plato} metaphysically depends on Socrates. An analogous argument shows that the set {Socrates, Plato} metaphysically depends on Plato.

To see that (1e) is false, consider worlds that make the antecedent true. Again, there is no reason to think that these must be impossible worlds. They are worlds that are *ceteris paribus* like the actual world, but which differ only in what is required to make it that {Socrates, Plato} does not exist. As before, we take it that the existence of the set {Socrates, Plato} is entailed by the Pairing axiom, plus the facts that both Socrates and Plato exist. That gives us three ways to make it that {Socrates, Plato} does not exist: either Socrates does not exist, or Plato does not exist, or Pairing is false (or some combination of the three). Even if we had reason to rule out the third option — and it's not clear that we do — there is no reason to think that all of these worlds make it that Socrates does not exist. It is plausible that some of them are such that Socrates does not exist, and others are such that Plato does not exist. The counterfactual (1e) is therefore false.

According to the counterfactual analysis of dependence, Socrates does not metaphysically depend on {Socrates, Plato}. Analogously, Plato does not depend on {Socrates, Plato} either.

Other examples of dependence that do not require an appeal to impossible worlds are the purely concrete cases. These cases do not involve sets at all. Consider, for example, me. It is plausible to think that I metaphysically depend, at least in some sense, on my biological parents. That is, I metaphysically depend on my mother and I metaphysically depend on my father. This is often affirmed by saying that I couldn't have had different parents. It certainly seems that this holds on a modal/existential conception of dependence. Nevertheless, my parents do not metaphysically depend on me. The counterfactual analysis of dependence confirms this, though we need to add a couple details. For concrete objects, like myself and my parents, tend to come into and go out of existence at different times. So we need to relativize to times. To show that I metaphysically depend on my mother (let's call her Mom), and not vice versa, we should evaluate the following counterfactuals.

(1g) Had Mom not existed at any time, then John wouldn't have existed at any time.

(1h) Had John not existed at any time, then Mom wouldn't have existed at any time.

We would like (1g) to be true and (1h) to be false. To see that (1g) is true, we look at worlds where the antecedent is true. There is no need to think these are impossible worlds. They are worlds that are *ceteris paribus* like the actual world, but which make it the case that Mom does not exist at any time. It is not entirely clear what must make this happen. But surely what need not change are the identity conditions for persons. One plausible necessary identity condition on persons is that they have the parents that they do (Kripke 1972). In order for anything to be John, it must have had Mom as a parent. But, like the set-membership relation, the relation of one person having another as a parent is an existence entailing relation. That is, if x has y as a parent, then there must be a time such that x exists at that time and there must be a time such that y exists at that time. As there are no times, in the worlds we are considering, at which Mom exists, nothing can bear the "having as a parent" relation to Mom. And so nothing can be identical to John. So John does not exist at these worlds. As (1g) is therefore true, the counterfactual analysis of dependence entails that I metaphysically depend on my mother. In a similar way, the counterfactual analysis entails that I metaphysically depend on my father as well.

To see that (1g) is false, we look at worlds where the antecedent is true. Still no need for impossible worlds. These are worlds that are *ceteris paribus* like the actual world, but which make it the case that I do not exist. To see how this happens, we may consider how I came to exist in the first place. This event was the result of many facts that involve my parents, only one of them being that Mom existed at a time. Worlds where my father doesn't exist, or where my parents did not meet, or where they did not get married, are worlds where I don't exist. It's not clear that the worlds we are considering, the worlds where I failed to exist, must be worlds where Mom doesn't exist either, as one of these other situations may have been brought about to prevent my existence.[6]

So the counterfactual analysis, extended with impossible worlds, can accurately capture dependence relations even in cases where impossible worlds are not required. Though these examples do not require an appeal to impossible worlds, many of the set-theoretic cases of dependence do. This becomes more apparent when we consider cases of metaphysical dependence that involve purely mathematical sets, like $\emptyset$ and $\{\emptyset\}$.

2.2. $\emptyset$ and $\{\emptyset\}$. Essentially the same arguments given in Section 2.1 apply here as well. The relevant counterfactuals that we must consider are:

(2a) Had $\emptyset$ not existed, then $\{\emptyset\}$ wouldn't have existed either.

(2b) Had $\{\emptyset\}$ not existed, then $\emptyset$ wouldn't have existed either.

To evaluate (2a), we must look at worlds that make the antecedent true. So we are looking at worlds that are *ceteris paribus* like the actual world, but which make it so that $\emptyset$ does not exist. Notice that, as opposed to the Socrates case, when we consider these worlds, we are immediately dealing with impossible worlds. For $\emptyset$ exists necessarily, in every possible world. Any world where $\emptyset$ does not exist must therefore be impossible. Without impossible worlds, both (2a) and (2b) come

[6]When extended to concrete objects, it may be that the counterfactual analysis requires further conditions. For example, consider a painting that has been inspired by the Holocaust. For simplicity, call it "Painting". Arguably, the counterfactual *Had Hitler not existed, then Painting wouldn't have existed either* is true. On the minimal counterfactual account of metaphysical dependence, it follows that Painting metaphysically depends on Hitler. And so it may be that in order to extend the minimal counterfactual analysis beyond the set-theoretic cases, more conditions are required. However, as an account of set-theoretic dependence, the minimal counterfactual analysis is sufficient. Thanks to Kit Fine for helpful discussion on this point.

out vacuously true. So we need impossible worlds to give serious consideration to both of these counterfactuals.

The relevant impossible worlds are those that are *ceteris paribus* like the actual world, but without the empty set. We have to figure out what must be different in order to make it true that $\emptyset$ does not exist. Again, I think we can be confident about some things that are not different. Specifically, the identity conditions for sets, embodied by the Axiom of Extensionality, remain the same.

Given that Extensionality holds, we can see that nothing at any of the worlds we are considering can be the singleton of the empty set. For an object to be the singleton of the empty set, it must have $\emptyset$ as a member. But $\emptyset$ does not exist at these worlds. So nothing can have $\emptyset$ as a member. Therefore, nothing at these worlds can be $\{\emptyset\}$. And so $\{\emptyset\}$ does not exist.

The counterfactual (2a) is therefore true: Had $\emptyset$ not existed, then $\{\emptyset\}$ wouldn't have existed either. It follows from the counterfactual analysis of dependence that $\{\emptyset\}$ metaphysically depends on $\emptyset$ — the set depends on its members.

To ensure that (2b) comes out false, we consider worlds where $\{\emptyset\}$ does not exist. What must change to make this happen? As before, $\{\emptyset\}$ exists because it follows from more general principle about sets: whenever something exists, so does its singleton. In ZF, this general principle follows from the Pairing axiom. As before, we assume that Pairing is what gives us the existence of $\{\emptyset\}$. If we are looking at worlds where $\{\emptyset\}$ does not exist, then Pairing must fail to hold at these worlds.

Does the failure of Pairing imply that the empty set fails to exist? Not clearly so. The existence of the empty set does not require that Pairing is true. That is, the existence of the empty set does not entail that Pairing is true. Otherwise Pairing would be a *theorem* of set theory, not an *axiom*. It seems, then, that (2b) is false — the empty set does not metaphysically depend on its singleton.

We would also like to show that the extended counterfactual analysis can avoid other unwanted claims of metaphysical dependence. Recall that, without impossible worlds, the counterfactual analysis entails that any object whatsoever metaphysically depends on any necessarily existing object. This is due to the truth of counterfactuals like

(2c) Had $\emptyset$ not existed, then Socrates wouldn't have existed either.

We would like (2c) to come out false. To show that is does, we consider worlds where $\emptyset$ does not exist. What must change to make this happen? Again, it's not entirely clear what must happen to make it the case that $\emptyset$ fails to exist. Sometimes set theory is done with an empty set axiom, as in Kripke-Platek (KP) set theory.

EMPTY SET. *There exists a set x such that for all y, y is not a member of x.*

More often, the existence of such a set is derived from the existence of other sets. Given the existence of any set, and a principle of Separation, it follows that the empty set exist.

To ensure that the empty set does not exist, at least one of these things must fail. No matter which fails, however, there is little reason to think that this would imply that Socrates does not exist. None of these set-theoretic axioms or principles has anything to do with Socrates. So there is reason to think that (2c) is false.

2.3. $\{\emptyset\}$ and $\{\{\emptyset\}\}$. Returning to purely set-theoretic cases, we notice that as we move up into the hierarchy of sets, things do not work quite as smoothly. Instead of the empty set and its singleton, consider the singleton of the empty set $\{\emptyset\}$ and *its* singleton $\{\{\emptyset\}\}$. As before, we would like $\{\{\emptyset\}\}$ to depend on $\{\emptyset\}$, but not vice versa. So we must evaluate the relevant counterfactuals.

(3a) Had $\{\emptyset\}$ not existed, then $\{\{\emptyset\}\}$ wouldn't have existed either.

(3b) Had $\{\{\emptyset\}\}$ not existed, then $\{\emptyset\}$ wouldn't have existed either.

To uphold the counterfactual analysis of metaphysical dependence, it must be that (3a) is true and (3b) is false. With (3a), there is no problem, as it comes out true for reasons very similar to those that make (1a) and (2a) true. The problem is with (3b).

With (3b), let us try the same reasoning as before. We consider worlds where $\{\{\emptyset\}\}$ does not exist. What must change to make this happen? As before, $\{\{\emptyset\}\}$ exists because it follows from a more general principle about sets: whenever something exists, so does its singleton. As $\{\emptyset\}$ exists, so does $\{\{\emptyset\}\}$. This more general principle follows from the Pairing axiom. So we must be dealing with worlds in which Pairing fails.

We then ask: Does the failure of Pairing imply that $\{\emptyset\}$ fails to exist? And in this case it seems like it does. The existence of $\{\emptyset\}$ does require the Pairing axiom. As we saw in section 2.2, we use Pairing to get $\{\emptyset\}$ from $\emptyset$. So in worlds where $\{\{\emptyset\}\}$ fails to exist, it must be that, because Pairing fails, $\{\emptyset\}$ does not exist either. And so we would have that (3b) is true, making it true that $\{\emptyset\}$ depends on its singleton.

To avoid this, we must make a qualification. We are looking at worlds where $\{\{\emptyset\}\}$ fails to exist. And so Pairing must fail *in a specific instance*, namely, the instance where we get $\{\{\emptyset\}\}$ from $\{\emptyset\}$. And that's the only difference that is required. It need not be that Pairing fails in every instance. Just in one. So we can have this be a world such that $\emptyset$ and $\{\emptyset\}$ both exist, but $\{\{\emptyset\}\}$ doesn't.

With this response in hand, we can show that as we work up the hierarchy of sets, the counterfactual analysis gets the dependence relations right. Sets depend on their members. Members do not depend on the sets that they are members of. And the dependence between sets and their members is asymmetric. A natural question is: What happens if we go down the hierarchy, instead of up? As we started with the empty set and its singleton, you can't get much further down than that. The only other case to examine is the empty set itself. What does the empty set depend on?

2.4. The Empty Set. We would like it to be that every set depends on its members, including the empty set. But the empty set has no members. We can say, then, that trivially it depends on all of its members. Show me one of the empty set's members that it doesn't depend on. That is, for all x such that $x \in \emptyset$, it's true that had x not existed, then $\emptyset$ wouldn't have existed.

I suppose it could be that there are other things that $\emptyset$ depends on. If you are a structuralist about sets, then perhaps the empty set, and every set, depends on the set-theoretic structure, or the structure of the cumulative hierarchy. We address these kinds of structuralist claims in section 3 of this chapter. Other plausible candidates could be put to the test of the counterfactual analysis.

But it may be that the empty set is an independent mathematical object. It is, in some sense, fundamental. The idea that $\emptyset$ is fundamental agrees with the common picture of the set-theoretic universe as a cumulative hierarchy of sets. At the base of this hierarchy is the empty set. On this view, the empty set is the fundamental ground, serving as the foundation for the rest of the set-theoretic universe.

2.5. Infinite Sets. There is no obvious reason to think that those considerations of metaphysical dependence do not apply to infinite sets as well. Any set, whether finite or infinite in size, should depend on its members. The set of natural numbers metaphysically depends on each natural number. The set of real numbers metaphysically depends on each real number.

Consider the set of natural numbers, and call it ω. We examine two cases: one where the natural numbers are *sui generis* objects, and one where they are von Neumann ordinals. If numbers are taken to be *sui generis* objects, or at least objects that are not sets, then it seems pretty straightforward that ω metaphysically depends on each natural number.

Is this result confirmed by the counterfactual analysis? To answer that question, we must examine a pair of counterfactuals.

(4a) Had the number 17 not existed, then ω wouldn't have existed either.

(4b) Had ω not existed, then the number 17 wouldn't have existed either.

We have chosen the number 17 as an arbitrary member of ω. If (4a) and (4b) hold (fail) for the number 17, then they should hold (fail) for any natural number. As before, in order to confirm the adequacy of the counterfactual analysis, we would like (4a) to be true and (4b) to be false.

Treated as *sui generis* objects, it seems pretty straightforward that (4a) comes out true. We must consider a world where the number 17 doesn't exist. Admittedly, it is hard to imagine what a world like this would be like. In fact, some have found the question impossible to answer.

> It is doubtless true that nothing sensible can be said about how things would be different if there were no number 17; that is largely because the antecedent of this counterfactual gives us no hints as to what alternative mathematics is to be regarded as true in the counterfactual situation in question (Field 1991, p. 237).

It is doubtless true that the antecedent of (4a) gives us no idea of what alternative mathematics holds in the worlds we are considering. But we do not need to know anything about the alternative mathematics to say something sensible. We can say with confidence that the failure of a certain object to exist in these worlds does not require that the identity conditions for sets are different in any way. And as long as the actual identity conditions for sets hold, then we can run the same argument as we have run before. Given that the identity conditions for sets hold, nothing at any of the worlds we are considering can be ω. For an object to be ω, it must have the number 17 as a member. But the number 17 does not exist at these worlds. So nothing can have 17 as a member. And so nothing at these worlds can be ω. The set ω does not exist.

It seems then that (4a) is true: Had the number 17 not existed, then ω wouldn't have existed either. It follows from the counterfactual analysis of dependence that ω metaphysically depends on its members.

As in the previous examples, the dependence of ω on the number 17 should be asymmetric, and so we would like (4b) to come out false. To confirm that (4b) is false, we consider worlds where the set of natural numbers ω does not exist. What must change to make this happen? Again, the existence of ω follows from the axioms of set theory. The most straightforward way to show that ω (as the set containing all and only the *sui generis* natural numbers) exists is as follows. Assume that all of the natural numbers exist, and that they are ordered in the usual way according to the successor function. Then use the Axiom of Infinity to show that there is a set which has 0 as a member and which is closed under the successor function.

INFINITY. *There exists a set y such that $0 \in y$ and for all $x \in y$, $s(x) \in y$.*

The worlds that we are considering, then, must make Infinity false. But nothing about the failure of Infinity implies that any natural number does not exist. So there are worlds that are *ceteris paribus* like the actual world where ω doesn't exist, but the number 17 does. And so (4b) is false — the members do not depend on the set.

As noted several times, we have been treating ω as the set of natural numbers, which have in turn been treated as *sui generis* objects. They have not as yet been identified with sets. But it is standard practice to identify numbers with von Neumann ordinals. On this set-theoretic account of number, we have that:

$$\begin{aligned} 0 &= \emptyset \\ \text{the successor of } n &= n \cup \{n\} \end{aligned}$$

The first few numbers are then:

$$\begin{aligned}
0 &= \emptyset \\
1 &= \emptyset \cup \{\emptyset\} &&= \{\emptyset\} \\
2 &= \{\emptyset\} \cup \{\{\emptyset\}\} &&= \{\emptyset, \{\emptyset\}\} \\
3 &= \{\emptyset, \{\emptyset\}\} \cup \{\{\emptyset, \{\emptyset\}\}\} &&= \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}
\end{aligned}$$

It is still the case that ω is the set of natural numbers. But now each member of ω is itself a set, and each number is a member of every number that is greater than it (given the standard ordering).

$$\omega = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} \ldots\}$$

In fact, each member of ω is an ordinal. Given this understanding of ω, we can ask whether ω depends on its members.

We consider this case because there is a line of reasoning that yields the result that ω, treated as a set of ordinals, does not depend on its members. The line of reasoning runs as follows. We want it to be that ω depends on, e.g., the empty set $\emptyset$. Using the counterfactual analysis, the goal is then to show that the following counterfactual is true.

(5) Had $\emptyset$ not existed, then ω wouldn't have existed either.

To evaluate this counterfactual, we look at worlds where the empty set fails to exist. We need not worry about how this comes about. But notice what happens to the ordinals in a world where we remove the empty set.

$$\begin{array}{llllll}
0 & = & \emptyset & \text{doesn't exist} & & \\
1 & = & \{\emptyset\} & \text{becomes} & \{\} & = \emptyset \\
2 & = & \{\emptyset, \{\emptyset\}\} & \text{becomes} & \{, \{\}\} & = \{\emptyset\} \\
3 & = & \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} & \text{becomes} & \{, \{\}, \{, \{\}\}\} & = \{\emptyset, \{\emptyset\}\}
\end{array}$$

This happens all the way up the ordinals. In effect, by removing the empty set, each ordinal becomes its predecessor. 1 becomes 0, 2 becomes 1, etc. And since each natural number has a successor, after this "transformation", we are still left with all of the ordinals. And nothing we have said prevents us from taking sets of them. And so the set of natural numbers, ω, exists. Thus, the counterfactual (5) is false: It is not the case that, had $\emptyset$ not existed, then ω wouldn't have existed.

The empty set was an arbitrary choice in this example, and this kind of reasoning can be used if we try to remove any ordinal. Thus, on this line of reasoning, ω does not metaphysically depend on *any* of its members — we can remove any member, and yet ω still exists.

But this line of reasoning is incorrect. We are supposed to be thinking about worlds where the empty set (for example) fails to exist. In the above line of reasoning we tried to do this by simply removing the empty set and seeing what happens. But we failed. In the worlds we considered, the empty set still exists, because the empty set just is any set that has no members. If any set like that exists, then $\emptyset$ exists. That's why, on some axiomatic formulations of set theory, like KP, the existence of the empty set is guaranteed by an axiom that says there is a set with no members.

EMPTY SET. *There exists a set x such that for all y, y is not a member of x.*

The existence of the empty set is guaranteed in ZF set theory by the Axioms of Infinity and Separation. That is, the existence of a set with no members is a theorem in ZF set theory. It is standard to treat the empty set as the set that has no members. In the worlds we have been considering, there is a set that has no members. It's just that now we call it 1 instead of 0. But whatever we call it, it is the empty set. And so we are not looking at worlds where the antecedent of (5) is true.

Worlds where the antecedent of (5) is true are worlds where the empty set does not exist. As the empty set does not exist, and because every ordinal has the empty set as a member, no ordinals exist either. In order for an object to be an ordinal, it must have $\emptyset$ as a member. But $\emptyset$ does not exist at these worlds. So nothing can have $\emptyset$ as a member. And so nothing at these worlds can be an ordinal. These are worlds without ordinals. And if natural numbers are identical with ordinals, then these are worlds without any of the natural numbers.

We can continue up the hierarchy of sets. We can look at uncountable sets. We can look at inaccessible cardinals. And so on. It's not clear that there is some point in the hierarchy where the style of argument we have appealed to in this chapter will fail. Any set, whether it is empty, or finite, or infinite in size, must have each of its members as a member in order to be the set that it is.

As we are holding identity conditions for sets fixed, if any member of a set does not exist, then the set does not exist either. Each set metaphysically depends on each of its members for its existence.

2.6. Non-Well-Founded Sets. To this point, all of the sets described here have been well-founded sets. They do not involve any circular chains of membership, nor do they involve any infinitely descending chains of membership. But, according to some, there are sets that are non-well-founded. One reason for thinking so is that non-well-founded sets are useful for modeling many different kinds of phenomena, including self-referential propositions and common knowledge.[7] I would like to remain neutral as to whether non-well-founded sets exist. I would certainly not want the counterfactual analysis to rule them out. In other words, I would like the counterfactual analysis to work for both well-founded and non-well-founded sets.

One who admits the existence of non-well-founded sets usually appeals to a different conception of what a set is. According to this conception, a set is any collection of things that can be depicted by a graph, where a graph consists of nodes and directed edges as usual. Indeed, on this conception, any graph depicts a set, including graphs with infinitely descending chains, as well as graphs with proper cycles.

I would like non-well-founded sets, if they exist, to metaphysically depend on their members. When arguing that well-founded sets depend on their members, I appealed to identity conditions of sets, which are captured by the Axiom of Extensionality. But non-well-founded sets have different identity conditions. And so those kinds of arguments may not work in this context.

Non-well-founded sets have different identity conditions because the fact that non-well-founded sets can be members of themselves makes the standard identity conditions ineffective. Consider two sets $a = \{c, a\}$ and $b = \{c, b\}$. Does $a = b$? If we appeal to Extensionality, then $a = b$ iff a and b have the same members. But in this example, that question boils down to the question of whether $a = b$. In order to answer the question "Does $a = b$?" we must already know if $a = b$.

[7] See Barwise & Moss (1996).

The question of set identity is instead settled structurally. Two sets are identical iff they are depicted by exactly the same graph. To make this precise, we need some terminology.[8] We have the usual understanding of a *graph* as a set of nodes and directed edges (ordered pairs of nodes). If there is an edge from one node x to another node y, then y is a *child* of x. We then take a function $tag(x)$ which takes childless nodes to the empty set $\emptyset$.[9] Once a graph is "tagged", one can give a *decoration* of the graph. The decoration of a graph is a function $\mathscr{D}$ such that

$$\mathscr{D}(x) = \begin{cases} tag(x) & \text{if } x \text{ has no children} \\ \{\mathscr{D}(y) : y \text{ is a child of } x\} & \text{if } x \text{ has children} \end{cases}$$

By decorating a graph, we are saying that each node represents a set — the set of its children.

Non-well-founded set theory maintains that each tagged graph has a unique decoration. That is, each tagged graph is decorated with sets. Given that we can have, e.g., circular graphs, as in Figure 1, we need non-well-founded sets to decorate these graphs. And so, if every graph has a decoration, then there must be non-well-founded sets available to decorate some of these graphs.

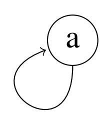

FIGURE 1. The set $a = \{a\}$

Non-well-founded set theory is characterized by the axiom that every graph has a unique decoration. That every graph has a unique decoration gives us identity conditions for all sets (not just non-well-founded ones). Two sets a, b are identical iff they are depicted by exactly the same graph. Take the example above, such that $a = \{c, a\}$ and $b = \{c, b\}$. The graphs of sets a and b are depicted in Figure 2.

[8] Here we follow Barwise & Etchemendy (1987). See also Aczel (1988). We give identity conditions for the non-well-founded set theory AFA of Aczel, as it is the non-well-founded set theory that has received the most attention. We note, however, that there are other options.

[9] If atoms are allowed, then $tag(x)$ takes childless nodes to either the empty set or to some atom.

FIGURE 2. The sets $a = \{c, a\}$ and $b = \{c, b\}$.

If we take it that c has no children, then c has no members. So $c = \emptyset$. Thus a has two members, one of which is $\emptyset$ and the other is itself. Similarly, b has two members, one of which is $\emptyset$ and the other is itself. These two sets are identical. To see this, call the leftmost node of Figure 2 x. There is a decoration $\mathscr{D}$ such that $\mathscr{D}(x) = a$. Suppose that there is another decoration $\mathscr{D}^*$ such that $\mathscr{D}^*(x) = b$. But then we have two decorations of the same graph. As every graph has a unique decoration, $\mathscr{D}(x) = \mathscr{D}^*(x)$, and so $a = b$.

Early in this chapter, we looked at several arguments to show that some sets (like $\{\emptyset\}$ and {Socrates}) metaphysically depend on their members. These arguments rely crucially on the identity conditions of sets as given by the Axiom of Extensionality. As we are now working with a different conception of set, which makes use of different identity conditions, we may need new arguments in favor of set-theoretic dependence.

In order to see exactly where the original argument breaks down (if it does), let us reconstruct it with a simple example, the empty set $\emptyset$ and its singleton $\{\emptyset\}$. The graphs of these sets are pictured in Figure 3.

FIGURE 3. The sets $\emptyset$ and $\{\emptyset\}$.

We consider the relevant counterfactual:

(2a) Had $\emptyset$ not existed, then $\{\emptyset\}$ wouldn't have existed either.

We would like (2a) to be true. (Note that our argument to show that the converse counterfactual is false did not rely on the identity conditions for sets.) Evaluating (2a), we look at worlds that differ only in what is required for it to be the case that $\emptyset$ fails to exist. We would like it to be the case that at these worlds $\{\emptyset\}$ does not exist either. In order for anything to be $\{\emptyset\}$, by Extensionality, it must have $\emptyset$ as a member. But we are no longer relying on Extensionality to give us identity conditions for sets. Now, in order for anything to be $\{\emptyset\}$, it must be depicted by the second graph in Figure 3.

Under these conditions, however, the original argument is not so far off the mark. The second graph in Figure 3 is tagged with the empty set $\emptyset$. Which means that any set that has this graph must have the empty set as a member. But the empty set does not exist. So no set can be depicted by this graph. So the singleton of the empty set does not exist. It is then true that "Had $\emptyset$ not existed, then $\{\emptyset\}$ wouldn't have existed either."

There is no reason to think that similar arguments will not carry over to any case involving well-founded sets. The question is whether they carry over to cases involving non-well-founded sets. Let us consider the simplest case, where we have a set $a = \{a\}$. Normally, to determine if this set depends on its members, we would evaluate the relevant counterfactual conditionals:

(6a) Had a not existed, then $\{a\}$ wouldn't have existed either.

(6b) Had $\{a\}$ not existed, then a wouldn't have existed either.

As $a = \{a\}$, there is good reason to think that (6a) is true iff (6b) is true.

The counterfactual "Had a not existed, then $\{a\}$ wouldn't have existed either" is true for the following reason. Consider worlds that differ only in what is required for it to be the case that a does not exist. Nothing more, nothing less. Could it be that $\{a\}$ exists in these worlds, even though a does not? It is clear that the answer is no, because $\{a\} = a$, and a does not exist. So $\{a\}$ does not exist. It follows that (6a) is true. An analogous argument shows that (6b) is true as well.

So the asymmetry that we found in the well-founded setting does not appear in the context of non-well-founded sets. But that is ok, as non-well-founded sets allow for two sets to be members of each other, and for one set to be a member of itself. Given that we want it to be the case that sets — all sets — metaphysically depend on their members, if we are to allow for the existence of non-well-founded sets, it should follow that some cases of set-theoretic dependence will be symmetric, or reflexive, or even have infinitely descending chains of dependence.

Of course, in arguing for the dependence of $\{a\}$ on a, we made use of the identity conditions for sets, which as we have noted now rely on the graphs that are used to depict sets. Consider again the graph of $a = \{a\}$.

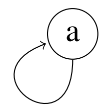

FIGURE 4. The set $a = \{a\}$

Now that we are considering worlds where a does not exist, could any other set in these worlds have the same graph? It seems not. Suppose there was some $b = \{b\}$. Then the set b would have the same graph as a.

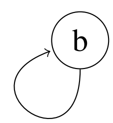

FIGURE 5. The set $b = \{b\}$

On the identity conditions for sets, we have that $a = b$. As a does not exist, b does not exist either. And so, if a does not exist, then no set with this graph exists either.

The upshot, then, is that the counterfactual "Had a not existed, then $\{a\}$ wouldn't have existed either" is true. It follows from the counterfactual analysis of dependence that $\{a\}$ metaphysically depends on a — the set depends on its members. Of course, by substitution of identicals (which we are not questioning), that means the converse counterfactual "Had $\{a\}$ not existed, then a wouldn't have existed either" is true as well. It follows from the counterfactual analysis of dependence that a metaphysically depends on $\{a\}$ — the member depends on the set. And so we must give up the asymmetry of dependence, at least for non-well-founded sets. And, because $a = \{a\}$, we must give up irreflexivity as well.

It turns out that the set $a = \{a\}$ has many different graphs, including those in Figure 6.

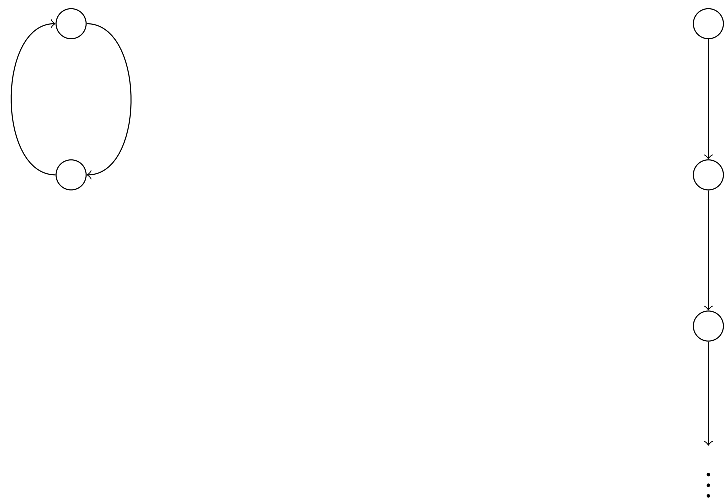

FIGURE 6. Alternative graphs of the set $a = \{a\}$.

Any set a that has a single member b, where b has a as its single member, can be shown to be identical to the set $a = \{a\}$. And any set that has an infinitely descending chain of membership can also be shown to be identical to the set $a = \{a\}$. If our arguments that the set $a = \{a\}$ metaphysically depends on its members, then the dependence relation, in the context of non-well-founded sets, need not be irreflexive, asymmetric, or well-founded.

Some have offered philosophical objections to the idea of chains of dependence that are not well-founded. Consider, for example, Ross Cameron (2008), who points out that one may think that dependent entities are only "real" because they derive their reality from independent entities.

> [T]he fundamental is what is real, with the dependent being unreal or less real. The thought is that if there were infinitely descending chains of dependence ... then nothing would be real (p. 9).

Or consider Jonathan Schaffer (2010), who argues that without an ultimate ground, nothing can have "being".

> There would be no ultimate ground. Being would be infinitely deferred, never achieved (p. 62).

According to these objections, only the ultimate ground, which is fundamental, which does not depend on anything, is real and has true being. The being, or the existence, of dependent objects is somehow derived from the being or existence of the ultimate ground. Therefore, without an ultimate ground, without objects that are fundamental, nothing at all could exist, nothing could have being. The intuitions behind these objections are appealing. But they do not constitute fully developed objections to the idea of non-well-founded dependence. It is not entirely clear why objects in an infinitely descending chain of dependence do not have being.[10] If each object gets its being from the object that comes before it in the chain, then every object gets its being from something. Beyond the vague worries that something is not quite right, there doesn't seem to be anything contradictory or incoherent about infinitely descending chains of dependence. For each thing in the chain, there is something that it depends on. For each thing in the chain there is something that makes it "real" or gives it "being". It is just that thing that comes before it in the chain.

3. An Account of Minimal Metaphysical Dependence

What has been proposed is a conception of metaphysical dependence that accurately characterizes dependence relations between sets and their members. It can plausibly capture dependence relations between other kinds of objects as well, like wholes and parts, or holes and hosts. But in some sense it is a minimal conception of dependence. It is minimalist because it does not assume that metaphysical dependence has any of the structural properties that it is normally taken to have. The dependence relation is often thought to be transitive, irreflexive, asymmetric, and well-founded. The counterfactual analysis of dependence says that:

COUNTERFACTUAL ANALYSIS. *x metaphysically depends on y if, and only if, had y not existed, then x wouldn't have existed either.*

In some of what follows, it will be more convenient to symbolize the counterfactual analysis. To do this we use an existence predicate E and the counterfactual conditional $\Box\!\!\rightarrow$. The counterfactual analysis can then be rendered as:

COUNTERFACTUAL ANALYSIS. *x metaphysically depends on y if, and only if,*

$$\neg E(y) \Box\!\!\rightarrow \neg E(x)$$

[10] For arguments against the claim that dependence must be well-founded, see Bliss (2011).

On the counterfactual analysis, dependence has none of the desired structural properties. But that is how it should be.

For some time, many have assumed that these properties must hold of the dependence relation. Any relation that fails to be transitive, irreflexive, asymmetric, or well-founded could not be the relation of metaphysical dependence. But the recent literature on dependence has started to question this position. We cannot *assume* that dependence has these properties. We must *argue* that it does. Consequently, some have thought that there is now little reason to believe that dependence is well-founded. For example, you might think that non-well-founded sets metaphysically depend on their members, just as well-founded sets do. In these cases, there would be infinitely descending chains of dependence. Aaron Cotnoir and Andrew Bacon (2012) have been looking at non-well-founded mereology, which allows for a parthood relation to be non-well-founded, symmetric, and even reflexive.[11] If you think that dependence tracks parthood, then you must also allow for a relation of dependence that has these structural properties as well. Some have even questioned the idea that dependence is transitive (Schaffer forthcoming).

What is nice about the minimal conception of metaphysical dependence, characterized by the counterfactual analysis with impossible worlds, is that with a little work, one can build in those structural properties that one believes the dependence relation should have. For example, if you have a good argument that metaphysical dependence is an irreflexive relation, then you can modify the counterfactual analysis as follows:

COUNTERFACTUAL ANALYSIS. *x metaphysically depends on y if, and only if,*

1. *had y not existed, then x wouldn't have existed either, and*
2. $x \neq y$.

This modification is not to suggest that we must think that dependence is reflexive. It is merely to show that, *given a good argument that dependence is reflexive*, we can incorporate that reflexive structure into the analysis of the dependence relation. Is there good reason to think that metaphysical dependence is irreflexive? Perhaps. The most common reasons are motivated by an association

[11] On reflexivity, also see Jenkins (2011).

of the dependence relation with other relations, such as priority, grounding, or relative fundamentality. Nothing is prior to itself, or more fundamental than itself. But it's not clear that metaphysical dependence should be associated with any of these relations, if only because it is not entirely clear as to what these relations are. In what sense is one thing prior to, or more fundamental than, another?

Recent literature on the reflexivity of dependence has focused on the notion of metaphysical explanation. The growing consensus is that we should somehow link up metaphysical dependence with metaphysical explanation. Perhaps it is a necessary condition that for x to depend on y, y must explain x (in a metaphysical sense). It may be that invoking explanation will deliver irreflexivity.[12] Things cannot be metaphysical explanations for themselves. But just like priority and fundamentality, the concept of metaphysical explanation is a concept that is far from clear. In fact, we might think that we have a better grasp of metaphysical dependence than we have of metaphysical explanation. Until work has been done to make the concept of metaphysical explanation more clear, I am skeptical that we should rely on it to solve issues involving dependence.

It may be that we sometimes want the dependence relation to be reflexive and sometimes we want it to be irreflexive. When we are looking at non-well-founded sets, the relation may allow for reflexive cases of dependence. When we are looking at sets in the cumulative hierarchy, dependence is irreflexive. Different domains may call for different properties of the dependence relation.

While it is up for debate as to whether or not the dependence relation is reflexive, it is relatively undisputed that the relation is transitive. As counterfactuals are not automatically transitive, some work must be done to deliver a transitive relation of metaphysical dependence on the counterfactual analysis. It may be that "Had A been the case, then B would have been the case", and "Had B been the case, then C would have been the case", and yet it is not true that "Had A been the case, then C would have been the case". We have the familiar counterexample (from Lewis 1973, p. 33):

(a) Had J Edgar Hoover been born a Russian, then he would have been a communist.

(b) Had J Edgar Hoover been a communist, then he would have been a traitor.

[12] As well as well-foundedness.

(c) Had J Edgar Hoover been born a Russian, then he would have been a traitor.

The inference from (a) and (b) to (c) is a bad inference. For it is plausible that (a) and (b) are true, while (c) is not clearly true at all. Had Hoover been born a Russian, he may well have been a patriotic communist.[13]

The counterfactual analysis of dependence explains metaphysical dependence solely in terms of counterfactuals. One cannot therefore piggyback on the transitivity of counterfactuals to show that dependence is transitive, because counterfactuals are not transitive. However, if one has a good independent argument that the dependence relation should be transitive, then we can modify the counterfactual analysis to make dependence a transitive relation. This modification is accomplished by introducing the idea of a chain of dependence. First take the non-transitive concept of dependence that is captured by the counterfactual analysis and call it dependence*.

x metaphysically depends on y if, and only if,* $\neg E(y) \mathrel{\Box\!\!\rightarrow} \neg E(x)$.

A chain of dependence orders objects using the dependence* relation.

a depends* on b, which depends* on $c, \ldots$

We can then define a transitive dependence relation using chains of dependence*.

COUNTERFACTUAL ANALYSIS. *x metaphysically depends on y if, and only if, there is a chain of dependence* from x to y.*

Given the idea of a chain of dependence, we get that, e.g., because $\{\{\emptyset\}\}$ depends* on $\{\emptyset\}$, and $\{\emptyset\}$ depends* on $\emptyset$, it follows that $\{\{\emptyset\}\}$ depends on $\emptyset$. Of course, we do not need transitivity to get the result that $\{\{\emptyset\}\}$ depends on $\emptyset$, as it also follows from an evaluation of the relevant counterfactual.

The relation of metaphysical dependence is also taken to be asymmetric, such that for all x and y, if x metaphysically depends on y, then y does not metaphysically depend on x. As with irreflexivity, you may think that some non-well-founded sets provide counterexamples to the asymmetry of dependence. On the other hand, you may not find those counterexamples persuasive, or you may have more persuasive arguments that the relation of dependence should be asymmetric. Or

[13] Brogaard and Salerno (2008) question whether this is really a counterexample to the transitivity of counterfactuals.

you may simply think that some domains call for an asymmetric dependence relation and other domains do not. Whatever your reason, it is easy to build asymmetry into the counterfactual analysis.

COUNTERFACTUAL ANALYSIS. *x metaphysically depends on y if, and only if,*

1. *had y not existed, then x wouldn't have existed either*

   $$\neg E(y) \mathrel{\Box\!\!\rightarrow} \neg E(x)$$

   and

2. *it is not the case that: had x not existed, then y wouldn't have existed either.*

   $$\neg[\neg E(x) \mathrel{\Box\!\!\rightarrow} \neg E(y)]$$

Or you may go for a slightly weaker, antisymmetric form of dependence.

COUNTERFACTUAL ANALYSIS. *x metaphysically depends on y if, and only if,*

1. *had y not existed, then x wouldn't have existed either*

   $$\neg E(y) \mathrel{\Box\!\!\rightarrow} \neg E(x)$$

   and

2. *if $\neg E(y) \mathrel{\Box\!\!\rightarrow} \neg E(x)$ and $\neg E(x) \mathrel{\Box\!\!\rightarrow} \neg E(y)$, then $x = y$.*

To wrap up, we can even provide for a dependence relation that is well-founded. At least, it will be well-founded in the sense that it rules out non-well-founded chains of dependence other than the reflexive cases. The dependence relation is well-founded in this sense if and only if every non-empty set of objects has an element x such that there is no y where $y \neq x$ and, had y not existed, then x wouldn't have existed either.

COUNTERFACTUAL ANALYSIS. *x metaphysically depends on y if, and only if,*

1. *had y not existed, then x wouldn't have existed either*

   $$\neg E(y) \mathrel{\Box\!\!\rightarrow} \neg E(x)$$

   and

2. *Every non-empty set S of objects has an element x, such that there does not exist an object $y \in S$ where $y \neq x$ and, had y not existed, then x wouldn't have existed either.*

   $$\forall S[\exists x(x \in S) \rightarrow \exists x(x \in S \wedge \neg\exists y(y \in S \wedge y \neq x \wedge \neg E(y) \mathrel{\Box\!\!\rightarrow} \neg E(x)))]$$

Characterizing the property of well-foundedness requires quantification over sets, treating sets as second-order objects. We thus need a second order-language (Shapiro 1991, ch. 5).

To sum up, we have a minimal conception of metaphysical dependence. This conception of dependence can be modified in various ways to achieve certain structural properties of the dependence relation that may be desirable. But one must provide independent argument that the dependence relation (on the domain of objects under consideration) should have these properties.

Beyond the structural properties of the dependence relation, it may be that there are other properties of dependence. It may be, for instance, that the dependence relation is a necessary relation. If x metaphysically depends on y, then necessarily, if x and y exist, then x metaphysically depends on y.

Does the counterfactual analysis support the claim that metaphysical dependence holds necessarily? There is reason to think that it does. We have an argument, in each case, from the perspective of the actual world. The set {Socrates} actually depends on Socrates. But the arguments we have presented rely almost exclusively on premises that are necessarily true. Most of the objects that we are interested in are purely mathematical objects. And so they exist necessarily, if they exist at all. And it is plausible to think that identity conditions generally, whether for sets or numbers or structures, are necessarily true. Furthermore, the Axioms of Extensionality, and Pairing, and Infinity, all hold with metaphysical necessity, as they are mathematical truths. So the arguments we have presented arguably go through no matter what possible world you start from. We can be confident, then, that dependence claims supported by the counterfactual analysis are necessarily true.

4. Mathematical Structuralism and Metaphysical Dependence

Having concluded my discussion of the relation of metaphysical dependence within the context of set theory, I would like to briefly discuss another area in the philosophy of mathematics where metaphysical dependence plays a role: mathematical structuralism. Mathematical structuralism says that mathematics is the study of structure, where a structure is the abstract form of a concrete system of objects that bear certain relations to one another. For example, the natural number structure is the abstract form of any system of infinitely many objects that has a unique object and a successor relation.

There are many versions of mathematical structuralism, but they can be divided roughly into two categories: eliminative and non-eliminative. Non-eliminative structuralists endorse the existence of abstract mathematical structures. According to this view, mathematical objects like numbers exist, but they are nothing more than positions in mathematical structures, where mathematical structures are abstract objects. Eliminative structuralists deny that there are abstract mathematical structures. According to this view, mathematics is about concrete structures, not abstract ones.

We will focus on non-eliminative structuralism, which we will henceforth refer to simply as structuralism. This kind of view is articulated by Resnik (1981), Shapiro (1997), among others. One main feature of the structuralist view is that mathematical objects in a structure depend on other objects in that structure, and on the structure itself.

> The number 2 is no more and no less than the second position in the natural number structure; and 6 is the sixth position. Neither of them has any independence from the structure in which they are positions, and as positions in this structure, neither number is independent of the other (Shapiro 2000, p. 258).

It is this claim of dependence that does a lot of work to distinguish structuralism from Platonist views about the nature of mathematical objects. Both the Platonist and the non-eliminative structuralist are committed to abstract mathematical objects. But the Platonist sees mathematical objects as independent from one another, just as many physical objects are independent from one another. The structuralist denies this claim. Mathematical objects depend on other mathematical objects that are in the same structure, and on the structures to which they belong. Furthermore, I contend that this dependence can be captured by the modal/existential conception of metaphysical dependence. A particular mathematical object cannot exist without the other mathematical objects that are in the same structure, and none of the objects in a structure can exist without the structure itself.

I would like to examine the dependence claims of the non-eliminative structuralist. The goal is to apply the counterfactual analysis of dependence to show that it supports these dependence claims. For convenience, I will examine them within the context of the natural number structure. We thus have two dependence claims.[14]

OBJECT DEPENDENCE. *Each natural number metaphysically depends on every other natural number.*

STRUCTURE DEPENDENCE. *Each natural number metaphysically depends on the natural number structure.*

Object Dependence and Structure Dependence should be distinguished from a third dependence claim, which we might call System Dependence.

SYSTEM DEPENDENCE. *Each concrete system that instantiates the natural number structure metaphysically depends on the natural number structure.*

Structuralists will agree on Object Dependence and Structure Dependence, in that they want them both to be true. But structuralists will disagree over whether System Dependence is true. For now, we focus on Object Dependence and Structure Dependence, returning to System Dependence at the end of this chapter.

The structuralist would like both Object Dependence and Structure Dependence to be true. Perhaps the counterfactual analysis of metaphysical dependence can help to evaluate these dependence claims. Beginning with Object Dependence, we can choose two numbers arbitrarily. Let's take the numbers 2 and 6. The structuralist would like it to be the case that the following counterfactual is true.

(7a) Had the number 2 not existed, then the number 6 wouldn't have existed either.

Notice that the structuralist would also like the converse counterfactual to be true as well.

(7b) Had the number 6 not existed, then the number 2 wouldn't have existed either.

How are we to evaluate these counterfactuals? Clearly, we need impossible worlds. If numbers exist at all, they are usually taken to exist necessarily. More carefully, from a structuralist's perspective, if numbers are merely parts (though not necessarily in a mereological sense) of an abstract

[14]See Linnebo (2008).

structure, then they too are abstract. Abstract objects are usually taken to exist necessarily. So any world where the abstract number 2 doesn't exist is an impossible world. With respect to the first counterfactual, we must consider impossible worlds that differ only in what is required to make it the case that the number 2 does not exist. What must change to make this happen? It is hard to say. But presumably what should not change are the identity conditions for natural numbers, or for mathematical objects generally.

The structuralist has a view on the identity conditions for mathematical objects. Consider Resnik (1981):

> In mathematics, I claim, we do not have objects with an 'internal' composition arranged in structures, we have only structures. The objects of mathematics, that is, the entities which our mathematical constants and quantifiers denote, are structureless points or positions in structures. As positions in structures, they have no identity or features outside a structure (p. 530).

Or Shapiro (1997):

> The essence of a natural number is its *relations* to other natural numbers. . . . The essence of 2 is to be the successor of the successor of 0, the predecessor of 3, the first prime, and so on (p. 72).

Shapiro has backed off of the claim that numbers *only* have structural properties.[15] But, presumably it is still a necessary condition of identity that, e.g., 2 be the successor of the successor of 0, the predecessor of 3, and so on.

Given these necessary identity conditions on natural numbers, we can return to evaluating our counterfactual conditionals. We are considering worlds where the number 2 does not exist, and wondering whether the number 6 could exist in these worlds. In order for anything to be the number 6, it must be the successor of the successor of the successor of the successor of the number 2. But the number 2 does not exist at these worlds. So nothing can be its successor. And so nothing can be that thing's successor. And so on. The number 6, therefore, does not exist. An analogous argument can be given, without loss of generality, concerning the converse counterfactual by appealing to the fact that 2 is the predecessor . . . of the number 6.

[15] See Shapiro (2006).

Given these arguments, the structuralist can use the counterfactual analysis of metaphysical dependence to claim that each natural number depends on every other natural number. What remains to be seen is whether the counterfactual analysis will deliver the structuralist's claim that each natural number depends on the natural number structure. I believe that it can, and I argue so below, but with a caveat. The account I give implies that the dependence between a natural number and the natural number structure is symmetric. But I think the non-eliminative structuralist can live with that.

Call the natural number structure $\mathscr{N}$. The natural number structure $\mathscr{N}$ keeps track of the fact that this structure consists of certain objects, the natural numbers $\mathbb{N}$, with a certain designated number, 0, which is the first number in the series defined by the injective successor function s. Choosing an arbitrary natural number we have the relevant counterfactuals.

(8a) Had $\mathscr{N}$ not existed, then the number 2 wouldn't have existed either.

(8b) Had the number 2 not existed, then $\mathscr{N}$ wouldn't have existed either.

Clearly, the structuralist wants (8a) to be true. I will argue that (8b) is true as well.

To evaluate both of these counterfactuals, we must look more closely at the relationship between numbers and structures. The natural number structure $\mathscr{N}$ is an abstract object. It consists of points, or positions, or places. The number 2, the number 6, and any natural number, are places in this structure. We can refer to them as objects, but they are not identical to any particular objects belonging to concrete systems that exhibit the natural number structure. They are "offices" as opposed to "officeholders". Just as there are many people who have held the office of Speaker of the House, there are many things that can hold the office of the number 2. Any object that holds the second (or third, depending on where you start) position in a concrete ω sequence is, in some sense, holding the office of the number 2. But the number 2 is not identical to any concrete object that holds this office. Any ω sequence is a "system" that exemplifies or instantiates $\mathscr{N}$, where $\mathscr{N}$ is the abstract form of this system. The number 2 is the second (or third) place in this abstract form, as given by the order created by successor function s.

The relationship between the number 2 (or any number) and the structure $\mathscr{N}$ is a bit sketchy. As we are following the version of structuralism as given by Shapiro, we can look to his attempts to explicate this relationship. Shapiro (2008) wants to treat structures as, or at least like, universals. Each structure is a one-over-many.

> Places are *components* of universals. Each *ante rem* structure consists of some places and some relations. ... A structure is constituted by its offices and the relations between them. The constitution is not that of mereology. It is not the case that a structure is just the sum of its places, since in general the places have to be related to each other via the relations of the structure. I think of an *ante rem* structure as a whole consisting of, or constituted by, its places and its relations (p. 302 – 3).

To evaluate (8a), we must consider worlds that differ only in what is required for it to be the case that the natural number structure $\mathscr{N}$ does not exist. It is as if we took the actual world, but got rid of the natural number structure. In these worlds, the abstract structure $\mathscr{N}$ does not exist. Because it is an abstract object, and abstract objects exist necessarily, we must be dealing with an impossible world. The question is whether or not, in these worlds, the number 2 exists. It seems plausible that had the natural number structure not existed, there wouldn't be any points, or positions, or places in this structure. For anything to be the number 2, it must be the second position in $\mathscr{N}$. But there is no $\mathscr{N}$. There are thus no positions in $\mathscr{N}$. Nothing, therefore, can be the second position in $\mathscr{N}$. And so the number 2 does not exist.

The counterfactual "Had $\mathscr{N}$ not existed, then the number 2 wouldn't have existed either" is true. And so the number metaphysically depends on the structure. But does the structure depend on the number? To evaluate (8b), we must consider worlds where the number 2, i.e., the second place in $\mathscr{N}$ doesn't exist. We will argue that in these worlds, $\mathscr{N}$ does not exist either, because nothing in these worlds could be identical to $\mathscr{N}$.

To show this, we appeal to Shapiro's (1997) identity conditions for structures.

> We take identity among structures to be primitive, and isomorphism is a congruence among structures. That is, we stipulate that two structures are identical if they are isomorphic. There is little need to keep multiple isomorphic copies of the same structure in our structure ontology, even if we have lots of systems that exemplify each one (p. 93).

Shapiro has argued for a sufficient condition on the identity of structures: if two structures $\mathscr{N}$ and $\mathscr{M}$ are isomorphic, then $\mathscr{N} = \mathscr{M}$. But surely the converse is true as well: if $\mathscr{N} = \mathscr{M}$, then $\mathscr{N}$ and $\mathscr{M}$ are isomorphic, for they are the same structure. So we have

$\mathscr{N} = \mathscr{M}$ *iff* $\mathscr{N}$ *and* $\mathscr{M}$ *are isomorphic.*

Suppose then that we are in a world where we remove the number 2 from the natural number structure $\mathscr{N}$. Note that, if we do this, nothing can step in to fill the place of the number 2. It is not as if the number 3 shifts to take the place of the number 2, and then the number 4 takes the place of 3, and so on. For then $\mathscr{N}$ would have a second position, and so the number 2 would still exist. We would have failed to consider a world where 2 does not exist. In fact, if we remove the number 2, we are left with a structure that is no longer $\mathscr{N}$, nor is it isomorphic to $\mathscr{N}$. The structure we are left with is not isomorphic to $\mathscr{N}$, because it has no place for the second position of an $\mathscr{N}$-like structure to be mapped to. And really no structure can be isomorphic to $\mathscr{N}$. Because if any structure was, then it would have a second position, in which case the number 2 would exist. Indeed, if any structure was isomorphic to $\mathscr{N}$, then it would be identical to $\mathscr{N}$. As $\mathscr{N}$ has a second position, the number 2 would exist. As we are, by hypothesis, considering worlds without the number 2, it follows that no structure is isomorphic to $\mathscr{N}$. And so no structure is identical to $\mathscr{N}$. The structure $\mathscr{N}$, therefore, does not exist in these worlds.

Given this argument, the counterfactual “Had the number 2 not existed, then $\mathscr{N}$ wouldn’t have existed either” is true. The natural number structure $\mathscr{N}$ metaphysically depends on the number 2. As 2 was chosen arbitrarily, $\mathscr{N}$ metaphysically depends on each natural number. Given this symmetric dependence of number on structure, we have an easier route to the claim that each number metaphysically depends on every other number. Arbitrarily choosing the numbers 2 and 6, we have

> The number 2 metaphysically depends on the structure $\mathscr{N}$.
>
> The structure $\mathscr{N}$ metaphysically depends on the number 6.

As long as metaphysical dependence is transitive, we have

> The number 2 metaphysically depends on the number 6.

It may seem that the metaphysical dependence of structure on number is incompatible with non-eliminative structuralism. But it is not obvious that this dependence causes any problems. The dependence of structure on number does not conflict with other dependence claims that the structuralist would like to make, such as Object Dependence, or the claim that numbers depend on structure. The only reason to think there is a problem is if you believe metaphysical dependence is asymmetric. But the structuralist is already committed to symmetric dependence between different numbers, and so she has a prior commitment to the possibility of symmetric dependence. Furthermore, the dependence of structure on number agrees with more naive intuitions about structure and dependence, as you couldn't have a structure without the places or positions or offices (i.e., the numbers) in that structure.

We have been focusing in this section on non-eliminative structuralism. Non-eliminative structuralists agree that abstract mathematical structures exist, and that mathematical objects are positions within these abstract structures. But they can disagree about how abstract mathematical structures relate to the concrete systems that exemplify them. *Ante rem* structuralism takes structures to exist prior to, and independent of, the systems that exemplify them. The natural number structure would exist even if no system of objects exemplified it. *In re(bus)* structuralism, on the other hand, argues that the relationship of dependence runs in the opposite direction: concrete system are prior, and abstract structures depend on them. The natural number structure could not exist without a particular system that exemplifies it. The *ante rem* and *in re* structuralists therefore disagree over System Dependence.

SYSTEM DEPENDENCE. *Each concrete system that instantiates the natural number structure metaphysically depends on the natural number structure.*

The *ante rem* structuralist believes System Dependence is true, while the *in re* structuralist believes it is false. A natural question is whether the counterfactual analysis can adjudicate between these two competing versions of non-eliminative structuralism.

Again, we may take the natural number structure $\mathscr{N}$ as an example. Given a particular system n that exemplifies $\mathscr{N}$, does $\mathscr{N}$ metaphysically depend on n? The question should actually be a bit more subtle than that. For the *in re* structuralist does not require that there exist any particular system n that exemplifies $\mathscr{N}$, just that there be some such n. The question then is whether the following counterfactual comes out true.

(9) Had there not existed any n such that n exemplifies $\mathscr{N}$, then $\mathscr{N}$ would not have existed.

It is not entirely clear that (9) is a true "contrary to fact" conditional. For all we know, the world only contains finitely many concrete systems, each consisting of only finitely many points, and so there are no systems n that exemplify $\mathscr{N}$. If this were the case, and if the *in re* structuralist is right, then the structure $\mathscr{N}$ does not exist. And so (9) is true. But if the *ante rem* structuralist is right, then $\mathscr{N}$ does exist, regardless of what kind of concrete systems exist. The upshot is that the counterfactual analysis of dependence, with impossible worlds, cannot easily determine the truth-value of (9).

The problem is that the kind of argument we have used in this chapter to evaluate other counterfactuals isn't much help here. The pattern that we have followed is to appeal to identity conditions to argue for certain connections, or lack thereof, between various objects, like sets and their members. We can try to apply the same kind of argument here by appealing to identity conditions on abstract structures.

$$\mathscr{N} = \mathscr{M} \textit{ iff } \mathscr{N} \textit{ and } \mathscr{M} \textit{ are isomorphic.}$$

But when we are considering worlds that have no concrete structure n, we need to answer the question, "Does there exist anything such that it is identical to the natural number structure $\mathscr{N}$?" And nothing we have said about these worlds, or about the connections between concrete systems and abstract structures can help us answer that question. In fact, this is precisely where the *ante rem* and *in re* structuralists disagree. The *in re* structuralist thinks that, in order for anything to be the natural number structure $\mathscr{N}$, it must be related in some way to a concrete system n that comprises countably many objects related to one another in the appropriate way. The *ante rem* structuralist denies this. Without having settled further facts about when something is identical to $\mathscr{N}$, we cannot easily evaluate (9) as we have evaluated the other counterfactuals in this chapter.

There may be more sophisticated ways to settle (9). And if so, that would give us something more to work with in the debate between *ante rem* and *in re* structuralism. But that is beyond the scope of this dissertation. That the counterfactual analysis cannot easily settle the debate between these two kinds of structuralism does not detract from its appeal. The counterfactual analysis properly evaluates many counterfactuals with impossible antecedents. In doing so, it can settle many questions about dependence relations between mathematical objects.

CHAPTER 7

Objections and Replies

In this chapter we consider objections to the counterfactual analysis of dependence as presented in Chapter 6. In the first three sections of this chapter, we consider three sets of objections. The first section looks at objections to high level assumptions that the counterfactual analysis takes for granted, like the existence of non-actual worlds and the worlds analysis of counterfactuals. We do not respond in depth to these objections, essentially mentioning them to set them aside. The second set of objections are not objections to specific aspects of the counterfactual analysis. They are rather objections to the general approach we have taken to analyze metaphysical dependence in terms of counterfactuals. In effect, they ask: why not do it this way instead? The third set of objections considers specific objections to particular aspects of the counterfactual analysis. We take each set of objections in turn. We conclude by addressing a general worry about how we can say anything about what happens at impossible worlds.

1. High Level Objections

The first high level objection that we consider points out that the counterfactual analysis takes it for granted that non-actual worlds exist, and that they can be used to evaluate claims of necessity, possibility and impossibility. But one may not think there are *any* non-actual worlds. Fair enough. But debates about metaphysical dependence and grounding have no obvious reason to reject a worlds analysis of the alethic modalities. So we need not settle once and for all whether a worlds analysis is appropriate for the modalities. Instead, we take it for granted, as is common, that a worlds analysis is appropriate.

A second objection points out that what is common is the acceptance of *possible* worlds to provide an analysis of possibility and necessity. But it is very uncommon to accept the existence of *impossible* worlds. On the standard possible worlds analysis, A is impossible iff there is no possible world w such that at w, A. Impossibility requires no commitment to impossible worlds. But we have weighed up the cases for and against impossible worlds in Chapter 5. There is good reason to think that impossible worlds exist. And the case against impossible worlds is relatively weak. So we should think that impossible worlds exist.

A third objection admits that we may accept the existence of some non-actual worlds, but it is not clear that we need evaluate counterfactuals in terms of worlds. There are some analyses (Goodman 1983, Fine 2012) of counterfactuals that do not appeal to worlds. Again, fair enough. But just as debates about dependence have no reason to reject a worlds analysis of the alethic modalities, they have no obvious reason to reject a worlds analysis of counterfactuals.

Furthermore, it is not clear that these alternative analyses of counterfactuals can handle counterpossibles in an appropriate way. Fine, for example, appeals to a semantics of possible states, as opposed to possible worlds.[1] Possible states are mereological parts of possible worlds. But looking at parts of possible worlds will not enable one to non-trivially evaluate counterfactuals with impossible antecedents. If there are no possible worlds where the antecedent of a counterpossible is true, then there are no parts of possible worlds where it is true either. All counterpossibles are therefore trivially true on Fine's account.

Goodman's classic account of counterfactuals says that a counterfactual $A \mathrel{\Box\!\!\rightarrow} B$ is true iff B follows by law from A together with a suitable set S of sentences. What goes into S is hard to decide and Goodman struggles with this a great deal. Neither can Goodman decide on what it means for one statement to "follow by law" from another. But it is taken to at least include the following of one statement from another by logical consequence. So regardless of what goes into S, if A is impossible, then simply from the laws of logic B will follow, because (at least on classical logic) everything follows from an impossible antecedent. However, though Goodman never settles on a fully satisfactory account, he does require that A and S be "self-compatible". According to Goodman, this requirement will not be satisfied by any antecedent that is impossible. And so all counterpossibles, which Goodman calls counterlegals, are false on his account. While we would

[1]What follows is not explicitly in the 2012 article. In that article, the suggestion of a semantics for counterfactuals in terms of possible states is very quickly suggested, with a promissory note to fill in the details later. Some of these details are available in an unpublished paper Fine (unpublished).

like some counterpossibles to be false, we do not want them all to be false. And so Goodman's account is unsatisfactory.

A recent account of counterfactuals, due to Leitgeb (2012a, 2012b), proposes a probabilistic semantics for counterfactuals. On this account we have the following truth conditions for counterfactuals (Leitgeb 2012a, p. 27).

LEITGEB. *$A \mathrel{\Box\!\!\rightarrow} B$ is true in a world w iff the conditional chance of B given A at w is very high.*

On this account conditional chances are not subjective epistemic or agent relative chances. They are objective, non-epistemic, and world relative. For a counterfactual $A \mathrel{\Box\!\!\rightarrow} B$ to be true, this account requires that the conditional chance of B given A be close to 1. But eventually, this probabilistic account appeals to worlds, as we have that the conditional chance of B given A being close to 1 is equivalent to: in all of the closest worlds among those in which A has a chance to be true, the conditional probability of B given A is close to 1.

Furthermore, David Lewis's system **V** for conditional logic is sound and complete with respect to Leitgeb's semantics.[2] System **V** is axiomatized as follows.

Rules

R1. *Modus Ponens*

R2. *Deduction within Conditionals: for any $n \geq 1$,*

$$\frac{\vdash (X_1 \wedge \ldots \wedge X_n) \supset Y}{\vdash [(Z \mathrel{\Box\!\!\rightarrow} X_1) \wedge \ldots \wedge (Z \mathrel{\Box\!\!\rightarrow} X_n)] \supset (Z \mathrel{\Box\!\!\rightarrow} Y)}$$

R3. *Interchange of Logical Equivalents*

Axioms

A1. *Truth-functional tautologies*

A2. $X \mathrel{\Box\!\!\rightarrow} X$

A4. $(\neg X \mathrel{\Box\!\!\rightarrow} X) \supset (Y \mathrel{\Box\!\!\rightarrow} X)$

A5. $(X \mathrel{\Box\!\!\rightarrow} \neg Y) \vee (((X \wedge Y) \mathrel{\Box\!\!\rightarrow} Z) \equiv (X \mathrel{\Box\!\!\rightarrow} (Y \supset Z)))$[3]

[2]See Leitgeb (2012a), p. 54, Theorem 3.3.
[3]Where $\equiv$ is the material bi-conditional.

In Chapter 5, Section 1.2, we proved that on the Lewis system **VC**, if A is logically false (i.e., impossible), then $A \mathrel{\Box\!\!\rightarrow} B$ is true, for any B. The proof goes through in system **V** as well. And so, on Leitgeb's semantics, all counterpossibles are true. As we would like some counterpossibles to be false, Leitgeb's semantics is unsatisfactory.

One last general objection that it is worth considering points out that the truth of counterfactuals can depend on the context in which they are asserted. Consider the familiar examples.

(10a) If Caesar had been in command in the Korean War, he would have used the atom bomb.

(10b) If Caesar had been in command in the Korean War, he would have used catapults.

It may be that in one context (10a) is true and (10b) false, while in another context it may be that (10a) is false and (10b) true. The worry is that this kind of context sensitivity may affect the truth of some of the counterfactuals we have considered in Chapter 6. Perhaps the counterfactual "Had Socrates not existed, then {Socrates} wouldn't have existed" is true in some contexts and false in others.

I agree that counterfactuals are sensitive to context. To deal with context sensitivity, I claim that if the truth value of a counterfactual varies depending on context, then we should build additional clauses into the antecedent to capture the context. It may be that (10a) is true given that Caesar has access to the weapons available during the Korean war, and perhaps the intentions to use the most effective weapons available. We should really be considering the counterfactual that adds certain clauses to the antecedent in order to capture these additional facts, because what we are interested in verifying includes what would happen given that these additional facts hold.

However, this may not be entirely necessary to evaluate some of the counterfactuals from Chapter 6. Evaluating the counterfactual "Had Socrates not existed, then {Socrates} wouldn't have existed", we are looking at worlds where Socrates doesn't exist. The argument that, in these worlds, {Socrates} also does not exist, relies only on the facts that Extensionality is true and that the inverse membership relation is an existence entailing relation. Since Socrates does not exist at these worlds, nothing at these worlds can have Socrates as a member. And so nothing at these worlds can be {Socrates}. Therefore, {Socrates} does not exist. It doesn't seem like there are many contexts where that argument would fail, and thus make it plausible that the counterfactual is false. The only contexts where that could happen would be ones where we are also claiming that, e.g., Extensionality fails, or the inverse membership relation is not an existence entailing relation. But now we are changing quite a lot in the worlds we are considering. And it's not clear that the counterfactual

(1c) Had Socrates not existed and Extensionality failed, then {Socrates} wouldn't exist.

should be relevant to the dependence relation that may or may not hold between {Socrates} and Socrates. If Extensionality fails, then it's not even clear that the objects we are talking about are really sets. The counterfactuals that are relevant to dependence relations ask us to look at worlds that keep as much as they can about the actual world fixed and differ only in what is required to make it true that some particular object does not exist. We then see if there is reason to think that some other object does not exist either.

2. Why not this way?

2.1. Contingent Mathematical Objects. To ensure that the counterfactual analysis of dependence produces the correct account of dependence relations between mathematical objects, we have extended the worlds analysis of counterfactuals with impossible worlds. Impossible worlds are required as we would like to consider what would have happened had certain mathematical object not existed. But mathematical objects exist in every possible world. So the worlds we consider must be impossible.

But an alternative approach rejects impossible worlds and embraces the (uncommon) view that mathematical objects exist contingently.[4] On this view, there are possible worlds where numbers and sets do not exist. When faced with the choice of accepting impossible worlds or rejecting the necessary existence of mathematical objects, why choose the first option over the second?

I think that one could develop a counterfactual analysis of dependence that rejected the necessary existence of mathematical objects. On this view, one would not need to appeal to impossible

[4]Hartry Field (1993) argues that the existence of mathematical objects is a contingent matter. Gideon Rosen (2006) argues that there is a sense of metaphysical necessity according to which the existence of mathematical objects, and other mathematical truths, are contingent.

worlds. In fact, I think one way to do this would look strikingly similar to the counterfactual analysis defended here. Nothing in the counterfactual analysis turns too much on the fact that some of the worlds we consider are impossible. All we need are worlds where certain mathematical objects fail to exist. If these worlds are possible worlds, and mathematical objects exist contingently, then so be it.

There is, however, independent reason for accepting the existence of impossible worlds. They are acceptable additions to many mainstream views of possible worlds. And they are useful in many different areas of philosophy. They are required to give a satisfactory logic of a relevant conditional, and to give a satisfactory logic of knowledge and belief. They can be used to extend a plausible understanding of properties to account for distinct properties that are necessarily coextensive. And they can be used to extend a plausible understanding of propositions to account for propositions that are true in all of the same possible worlds.

Most importantly for our purposes, they are required for giving an account of counterfactuals with impossible antecedents. Counterpossibles having to do with the existence of mathematical objects are not the only counterpossibles that one may be interested in. By rejecting the necessary existence of mathematical objects, we have simply redrawn the line between possible and impossible. And we still might want to know what would happen had the impossible (given this redrawn line) occurred. Presumably, on this view contradictions are still impossible. But we still might want to think about what would have happened had a particular contradiction been true. We would then be thinking about an impossible world.

So redrawing the line between possible and impossible doesn't really help. And if we keep the line where it is, we can preserve the intuitions that mathematical objects, like numbers and pure sets, exist necessarily, because they exist in every possible world.

2.2. Mathematical Dependence. We are interested in the question of whether or not a set depends on its members. Ask a mathematician if a set depends on its members and she will probably look at you funnily. She either doesn't know what you mean, or she doesn't care. Because this is a metaphysical, not a mathematical, question. You get a similar response when you ask a mathematician if numbers *really* exist. She usually just isn't interested. But press her further on the question of dependence, and she will probably say something like this. Consider the empty set $\emptyset$ and its singleton $\{\emptyset\}$. The singleton of the empty set $\{\emptyset\}$ depends on the empty set $\emptyset$ because there is no interpretation $\mathfrak{I} = \langle D, v\rangle$ such that $\{\emptyset\}$ is in the domain of $\mathfrak{I}$ but $\emptyset$ isn't. This mathematical approach gives us a more general analysis of dependence.

MATHEMATICAL ANALYSIS. *x metaphysically depends on y if, and only if, there is no interpretation* $\mathfrak{I} = \langle D, v\rangle$ *such that* $x \in D$ *and* $y \notin D$.

In this section, we explore this natural mathematical alternative to the counterfactual analysis of dependence.

The mathematical analysis is in many ways like a version of the modal analysis of dependence discussed in Chapter 2.

MODAL ANALYSIS, V1. *x metaphysically depends on y if, and only if, there is no possible world, w, such that x exists in w and y does not exist in w.*

To be in the domain of an interpretation is to exist according to that interpretation. So the mathematical analysis says that $\{\emptyset\}$ depends on $\emptyset$ because $\{\emptyset\}$ cannot exist according to an interpretation without $\emptyset$ also existing according to that interpretation. In other words, it couldn't be that $\{\emptyset\}$ exists but $\emptyset$ doesn't. A possible-worlds explanation of metaphysical dependence basically replaces the word 'interpretation' with the phrase 'possible world'. The singleton $\{\emptyset\}$ depends on the empty set $\emptyset$ because there is no possible world w such that $\{\emptyset\}$ exists at w but $\emptyset$ doesn't. But recall that the modal analysis fails. Does the mathematical analysis do any better?

Before answering that question, it is important to remember that interpretations are not possible worlds. Interpretations, assuming we are talking about interpretations of first-order logic, can tell us what sentences are first-order logical truths, and what inferences are first-order logically valid: an inference $\langle \Sigma, B\rangle$ is logically valid if there is no interpretation $\mathfrak{I}$ such that $\mathfrak{I}$ makes every member of Σ true and doesn't make B true; a sentence B is a logical truth if the inference $\langle \emptyset, B\rangle$ is valid. So even if it was the case that, for some a and b, there is no interpretation $\mathfrak{I} = \langle D, v\rangle$ such that $a \in D$ and $b \notin D$, that would tell us there is some logical connection between these two terms. Furthermore, it would tell us that the formula $\exists x(x = a) \supset \exists x(x = b)$ is a logical truth. But that doesn't seem right. It may be a metaphysical truth that b must exist if a does. It should not be a truth of logic.

But I think the idea that dependence between mathematical objects can be understood in terms of interpretations, as according to the mathematical analysis, is a natural one. So I would like to explore the idea further. Let's begin by comparing the modal analysis (v1) and the mathematical analysis. Recall that the modal analysis fails because pure sets exist in every possible world. As both the empty set and its singleton exist in every possible world, this basic analysis is unable to deliver the asymmetry of the dependence between $\{\emptyset\}$ and $\emptyset$, and between singletons and their members generally. But certainly there are interpretations such that the empty set exists according to that interpretation but its singleton doesn't. Just take the interpretation $\mathfrak{I}_1$ with one thing in its domain, and say that thing is the empty set. Done. But given the ease with which we constructed this interpretation, it seems we can construct another interpretation $\mathfrak{I}_2$ with one thing in its domain, and say that thing is the singleton of the empty set. The symmetry with respect to what interpretations exist remains, which suggests that a basic analysis using interpretations will not deliver the asymmetry that we want either.

We ran into a similar problem with impossible worlds in the Revised Modal Analysis. Simply adding impossible worlds into the mix, worlds where objects that exist necessarily fail to exist, will not yield the required asymmetry. Interpretations are not worlds, either possible or impossible. But they act, or we can treat them, a lot like impossible worlds, because one can have interpretations with domains that fail to include objects that exist necessarily. But, given any two things that exist (like the empty set and its singleton) we can have an impossible world where one exists and the other doesn't. And similarly, given any two things that exist we can have an interpretation such that one exists in the domain of the interpretation and the other doesn't.

The problem with the mathematical account is that we can build interpretations however like. An interpretation is just an ordered pair $\langle D, v\rangle$ consisting of a domain of objects and a valuation function, which maps parts of our object language (taken to be the first-order language of set theory) to objects in D or sets constructed out of objects in D. Assuming that the domain of an interpretation is taken from the (real) universe of sets, there are many different interpretations one can build. One can build an interpretation with just the empty set in the domain, or an interpretation with just its singleton in the domain. These will not be very interesting interpretations, but they are interpretations. So the straightforward mathematical analysis will not yield *any* dependence relations between any two different sets.

But we may be able to generate dependence relations by adding some restrictions on the interpretations we consider. The most straightforward way to do this is to require that these interpretations be models of certain theories. As we are interested in sets, it seems appropriate to require that they be models of some set theory. But that may be asking too much. Any axiomatic version of set theory (pick your favorite: Z or ZF or ZFC or even NBG or ZU or NF) is a theory of the infinite. Because set theory is primarily a theory of the infinite. These set theories have axioms to ensure that they are theories about infinitely many things. Models of these theories make it true that an infinite number of sets exist. That is, the domains of these the interpretations that are models of these theories are infinite. But we may want dependence relations to hold even in interpretations with finite domains. So we should be looking at some interpretations with finite domains. Furthermore, if we only consider interpretations that are models of some established set theory, then every interpretation will require the existence of both $\emptyset$ and $\{\emptyset\}$. So there are no interpretation where either fails to exist. And so there will be no asymmetric dependence between them, according to this version of the mathematical analysis. That means we need to be looking at interpretations that fail to satisfy some of the axioms of set theory.

What we want to show is that by looking at interpretations we can see that the singleton of the empty set depends on the empty set (and not vice versa), according to the mathematical analysis. Looking at all interpretations will not do this, so we need to look at a restricted class of interpretations. I contend that any plausible way to do this fails to establish an asymmetric dependence relation between a set and its members. We show this in the context of our favorite example: the case of $\emptyset$ and $\{\emptyset\}$.

One way to restrict the interpretations under consideration is to ensure that we are really talking about the empty set and its singleton. So we require that our valuation function points the terms '$\emptyset$' and '$\{\emptyset\}$' to the right objects in our domain. So we need that $v(\text{'}\emptyset\text{'}) = \emptyset$ and $v(\text{'}\{\emptyset\}\text{'}) = \{\emptyset\}$. But then it must be that $\emptyset, \{\emptyset\} \in D$. And so both exist according to every interpretation we consider. There will be no interpretation, satisfying these restrictions, where one exists and the other doesn't. And so we trivially have symmetric dependence, according to the mathematical analysis.

Another plausible restriction on the interpretations we consider is that they be models of some particularly relevant formula. For example, if we are considering the dependence relation between $\emptyset$ and $\{\emptyset\}$, we might require that each interpretation $\mathfrak{I}$ is such that $\mathfrak{I} \Vdash \emptyset \in \{\emptyset\}$. But even if we enforce this requirement, we can show that the mathematical analysis of dependence does not generate the desired results.

Consider the interpretation $\mathfrak{I}_1 = \langle D_1, v_1 \rangle$ such that

$$D_1 = \{\emptyset\}$$
$$v_1(\text{‘}\{\emptyset\}\text{’}) = v_1(\text{‘}\emptyset\text{’}) = \emptyset$$
$$v_1(\text{‘} \in \text{’}) = \{\langle \emptyset, \emptyset \rangle\}.$$

This interpretation has one thing, the empty set $\emptyset$ in its domain D_1. The terms ‘$\{\emptyset\}$’ and ‘$\emptyset$’ both refer to the one and only thing in the domain: $\emptyset$. The extension of the predicate symbol ‘$\in$’ is not the standard membership relation, but more like the identity relation for objects in D_1. As long as we are allowed to substitute terms that refer to the same object, this interpretation makes the sentence ‘$\emptyset \in \{\emptyset\}$’ true, because it makes ‘$\emptyset \in \emptyset$’ true, and ‘$\emptyset$’ and ‘$\{\emptyset\}$’ refer to the same object in the domain. But according to this interpretation $\emptyset \in D_1$ but $\{\emptyset\} \notin D_1$. And so, $\emptyset$ does not depend on $\{\emptyset\}$ according to the mathematical analysis.

But we can construct another similar interpretation to show that $\{\emptyset\}$ does not depend on $\emptyset$ either. Consider the interpretation $\mathfrak{I}_2 = \langle D_2, v_2 \rangle$ such that

$$D_2 = \{\{\emptyset\}\}$$
$$v_2(\text{‘}\{\emptyset\}\text{’}) = v_2(\text{‘}\emptyset\text{’}) = \{\emptyset\}$$
$$v_2(\text{‘} \in \text{’}) = \{\langle \{\emptyset\}, \{\emptyset\} \rangle\}.$$

This interpretation has one thing, the singleton of the empty set $\{\emptyset\}$ in its domain D_2. The terms ‘$\{\emptyset\}$’ and ‘$\emptyset$’ both refer to the one and only thing in the domain: $\{\emptyset\}$. The extension of the predicate symbol ‘$\in$’ is not the standard membership relation, but more like the identity relation for objects in D_2. As long as we are allowed to substitute terms that refer to the same object, this interpretation makes the sentence '$\emptyset \in \{\emptyset\}$' true, because it makes '$\{\emptyset\} \in \{\emptyset\}$' true, and '$\emptyset$' and '$\{\emptyset\}$' refer to the same object in the domain. But according to this interpretation $\{\emptyset\} \in D_2$ and $\emptyset \notin D_2$. It follows, given the mathematical analysis, that $\{\emptyset\}$ does not depend on $\emptyset$. And so we now have no dependence relations between sets and their members.

On the counterfactual analysis, we required the impossible worlds we considered to uphold the standard identity conditions for sets, captured by Extensionality. Perhaps that will help here. To do this, we would require that the interpretations we consider satisfy Extensionality. But both interpretations $\mathfrak{I}_1$ and $\mathfrak{I}_2$ satisfy Extensionality, as they both have only one thing in their domains. So the appropriate asymmetric dependence relations will not be derived from the mathematical analysis, even if we restrict the interpretations we consider to be models of Extensionality.

It is not obvious that any restrictions can do what the proponent of the mathematical analysis wants. There are also other considerations as to why the counterfactual analysis is more appropriate. To explain these, we must first address a more direct objection to the counterfactual analysis.

3. Specific Objections

3.1. Dependence and Non-Existent Objects. Given the counterfactual analysis, and certain plausible assumptions about which worlds are *ceteris paribus* like the actual world, objects that do not actually exist automatically depend on each other. This holds for objects, like Pegasus and Cerberus, which do not actually exist, but only contingently so. And it holds for objects that do not exist in any possible world, as they certainly do not actually exist. To illustrate, we take two impossible objects. One will be the round square cupola on Berkeley College, called 'Cupola'. The other will be Sylvan's box, the impossible box, both empty and occupied, from Priest (1997b). According to the counterfactual analysis, Cupola depends on Sylvan's box iff the following counterfactual is true.

(11a) Had Sylvan's box not existed, then Cupola wouldn't have existed either.

So we look at worlds that are *ceteris paribus* like the actual world such that Sylvan's box does not exist. But Sylvan's box (presumably) does not exist at the actual world. So, in this example, it is plausible that the set of worlds that are *ceteris paribus* like the actual world consists of just the actual world. If we are to look at those worlds that differ from the actual world only in what is required to make the antecedent true, in this example there need not be anything different from the actual world. Any non-actual world will therefore change more than is required to make the antecedent true.[5] This would hold according to Lewis's $\mathbf{C}_1$ and Stalnaker's $\mathbf{C}_2$. To evaluate (11a), then, all we need do is see if Cupola exists at the actual world. And clearly it does not, as no round squares exist at the actual world. So (11a) is true, and so Cupola metaphysically depends on Sylvan's box. And similarly Sylvan's box depends on Cupola. And any non-existent object depends on any other non-existent object.

This is a problem for the counterfactual analysis. But the fix is easy. Recall that to say one object x depends on another object y is shorthand for expressing that a dependence relation holds between certain facts (where facts are states of affairs that actually obtain): the fact that $< x$ exists $>$ metaphysically depends on the fact that $< y$ exists $>$. Clearly, if the existence of one thing actually depends on the existence of another, then both of those things should exist at the actual world. So there is an additional constraint on the relation of metaphysical dependence that we can explicitly incorporate into the counterfactual analysis.

COUNTERFACTUAL ANALYSIS. *x metaphysically depends on y iff*

(1) *x exists and y exists, and*

(2) *had y not existed, then x wouldn't have existed either.*

While this fix is easy, I think more should be said. Because I would like it to be that if certain objects had existed, then they would instantiate some dependence relations. For example, one might be a nominalist about mathematical objects. This person would not think that $\emptyset$ and $\{\emptyset\}$ actually exist, or even possibly exist. But the nominalist may still think that, if $\emptyset$ and $\{\emptyset\}$ had existed, then $\{\emptyset\}$ would metaphysically depend $\emptyset$, or if any set had existed, then it would depend on its members. In other words, one might not think that any sets exist, but still think that the membership relation between a set and its members involves metaphysical dependence.

[5]It may be that there are non-identical but indiscernible worlds, and so there are non-actual worlds that are indiscernible from the actual world. Including these in the set of worlds that we consider will not change how we handle these cases.

At the very least, it seems reasonable to me that some things could participate in a relation of dependence if they had existed. So I would like some statements of the form "Had *x* and *y* existed, then *x* would have depended on *y*" to be true. And I would like some to be false. For example, I would like the following to be false:

(11b) Had Cupola and Sylvan's box existed, then Cupola would have depended on Sylvan's box.

To see that it is, we look to worlds that are *ceteris paribus* like the actual world, but where Sylvan's box and Cupola exist. These will be impossible worlds, as Cupola and Sylvan's box are impossible objects, but they will keep a lot about the actual world the same. For (11b) to be false, it must be that at one of these worlds the counterfactual "Had Sylvan's box not existed, then Cupola wouldn't have existed either" must be false. So from the perspective of one of these worlds, there must be one world that is *ceteris paribus* like it such that Sylvan's box doesn't exist, but Cupola does. Of course, once we start looking at embedded counterfactuals, we may lose many of our intuitions about these things. But given that there is no obvious connection between these objects, there doesn't seem any good reason to think that every world that is *ceteris paribus* like the actual world, where Sylvan's box and Cupola exist, is such that every world that is *ceteris paribus* like it, where Sylvan's box does not exist, is such that Cupola doesn't exist either. So we can extend the counterfactual analysis to capture dependence relations that would hold between certain objects, even impossible objects, if they had existed.

For this reason, one might argue that the mathematical analysis we examined in Section 2.2 of this chapter, which uses the existence of interpretations to determine dependence between objects, is misguided. Because even if one could find the appropriate conditions on the class of interpretations, so that the mathematical analysis generated the appropriate asymmetric dependence relations between sets are their members, the mathematical analysis is subject to a similar objection about non-existent objects. And I don't think that the proponent of mathematical dependence has such a simple response, which gives us reason to prefer the counterfactual analysis.[6]

[6] The mathematical analysis, by the way, is not 'mathematical' because it is only supposed to apply to mathematical objects. It can be applied to any objects, mathematical or otherwise. So it can be applied to objects like Sylvan's box and Cupola.

The objection runs as follows. When we build interpretations, the domains of those interpretations consist of objects that actually exist. That is, for any interpretation $\mathfrak{I} = \langle D, v\rangle, D \subseteq U$, where U is the collection of objects that exist at the actual world. U is the actual universe, so to speak. If an object x does not actually exist, then x cannot be a member of the domain of any interpretation. Pegasus, Cerberus, Cupola, or Sylvan's box. None of these things can be a member of the domain of any interpretation, because none of them actually exists.

If we pick two of these, say Cupola and Sylvan's box, then there is no interpretation $\mathfrak{I} = \langle D, v\rangle$ such that Cupola $\in D$ and Sylvan's box $\notin D$. Because there is no interpretation $\mathfrak{I} = \langle D, v\rangle$ such that Cupola $\in D$. So, according to the mathematical analysis, Cupola metaphysically depends on Sylvan's box. And similarly, Sylvan's box depends on Cupola, and any non-existent object depends on any other non-existent object.

Those who endorse the mathematical analysis can respond by saying that dependence relations only hold between objects that actually exist. And they can add conditions on the mathematical analysis accordingly. But they may still want to extend the mathematical account to capture those dependence relations that would hold between non-existent objects, had they existed. In which case they may want to say of two non-existent objects x, y that:

(12) Had x and y existed, then x would have depended on y.

But (12) is a counterfactual conditional. We normally do not analyze counterfactuals in terms of interpretations; we analyze them in terms of worlds. So we need worlds to analyze (12). And if either x or y is an impossible object, we need impossible worlds. We would then look to these worlds where x and y exist and evaluate the dependence claim in terms of the existence, at these worlds, of particular interpretations.

I do not claim that this can't be done. But if one is to choose between two accounts of dependence, one of which only appeals to worlds, and one which appeals to worlds and interpretations, where both work equally well, it seems like the former is a better choice, as it only appeals to one kind of object.

It should be noted that this section relies heavily on the Quinean assumption that something must exist to be in the domain of an interpretation (i.e., to be quantified over). It is worth considering whether this assumption can be avoided, as there are views that allow for non-existent objects, and indeed for quantification over them.

So let us suppose that our universe is the universe of all objects: abstract and concrete objects, actual and merely possible objects, and why not impossible objects too. Our quantifiers range over *everything*. So, any objects, including Pegasus, Cerberus, Cupola, and Sylvan's box can belong to the domain of an interpretation. Clearly then, there are interpretations such that any one of these objects belongs to the domain and none of the others do. Indeed, for any objects, existent or not, there are interpretations such that one exists in the domain and the other does not. So the mathematical analysis of dependence will yield no dependence relations between any objects.

3.2. Object vs. Principle. The next objection I wish to address directly targets how we handled the evaluation of counterfactuals in Chapter 6. It will suffice to look at the first example we considered: Socrates and {Socrates}.

(1a) Had Socrates not existed, then {Socrates} wouldn't have existed either.

(1b) Had {Socrates} not existed, then Socrates wouldn't have existed either.

In evaluating these counterfactuals, I said that (1a) is true and (1b) is false. I claimed that (1a) is true because in the worlds we consider, there are certain principles about sets, like Extensionality, that should not change. Nothing about the fact that Socrates fails to exist should entail that Extensionality fails. If Extensionality holds, then for anything to be {Socrates} it must be related, via the inverse membership relation, to Socrates. But Socrates does not exist in these worlds. So nothing in these worlds can be related to Socrates is this way. And so nothing in these worlds can be {Socrates}. The set {Socrates}, therefore, does not exist in these worlds.

I claimed that (1b) is false because in the worlds we consider, there are certain principles about sets, like Pairing, that may change. Given the existence of Socrates, Pairing implies the existence of {Socrates}. So in these worlds, either Socrates does not exist, or Pairing fails. It is plausible that, at least in some of these worlds, the reason {Socrates} does not exist is because Pairing fails, even though Socrates still exists. And so (1b) is false.

One might object that there is an asymmetry in how we handled these cases. In one case I upheld a principle about sets (Extensionality) and rejected the existence of an object ({Socrates}). In the other case, I rejected a principle about sets (Pairing) and upheld the existence of an object (Socrates). Is this asymmetry justified?

I think it is. It is justified for the reasons given above. Nothing about the failure of Socrates to exist requires us to reject Extensionality. In other words, Extensionality does not require the existence of Socrates; Extensionality does not require the existence of anything. With respect to Pairing, we are not requiring that Pairing fail in every world under consideration. We require only that it fail in some world that we are considering. There is certainly no guarantee that Pairing is true in all of the worlds under consideration.

But the asymmetry is justified on other grounds. We are primarily interested in a theory of dependence in the context of sets. Sets depend on their members, and not (usually) vice versa.[7] If we reject Extensionality, then it is no longer clear that the objects we are dealing with are really sets. Extensionality holds a special place among the axioms of set theory. Recall the quote, from Chapter 2, by Boolos.

> The axiom of extensionality enjoys a special epistemological status shared by none of the other axioms of ZF. Were someone to deny another of the axioms of ZF, we would be rather more inclined to suppose, on the basis of his denial alone, that he believed that axiom false than we would if he denied the axiom of extensionality. … [I]f someone were to say, 'there are distinct sets with the same members,' he would thereby justify us in thinking his usage nonstandard far more than someone who asserted the denial of some other axiom. … That the concepts of *set* and *being a member of* obey the axiom of extensionality is a far more central feature of our use of them than is the fact that they obey any other axiom. A theory that denied, or even failed to affirm, some of the other axioms of ZF might be called a set theory, albeit a deviant or fragmentary one. But a theory that did not affirm that the objects with which it dealt were identical if they had the same members would only by charity be called a theory of *sets* alone. (1971, pp. 27 – 8).

Extensionality is what gives us the identity conditions for sets. It is what tells us that the objects we are dealing with are sets, as opposed to non-extensional objects. Pairing, on the other hand, is

[7] At least not when we're dealing with well-founded sets.

not crucial to the identity conditions for sets. If Pairing were to fail, we would have fewer sets than we do in fact have. But we would still be working with *sets*.

3.3. Symmetry. It would be nice if the counterfactual analysis was able to account for more than the dependence of one object on another. Until now, we have only been looking at cases where the existence of one object depends on the existence of another object. And we have focused on cases where these objects exist necessarily. The counterfactual analysis, extended with impossible worlds, works well for this kind of dependence. But metaphysical dependence can hold between other kinds of things. For example, it is sometimes said that the truth of a proposition depends on the thing that makes it true. Consider Aristotle, as we did in Chapter 1.

> [I]f there is a man, the statement whereby we say that there is a man is true, and reciprocally — since if the statement whereby we say that there is a man is true, there is a man. And whereas the true statement is in no way the cause of the actual thing's existence, the actual thing does seem in some way the cause of the statement's being true: it is because the actual thing exists or does not that the statement is called true or false (Aristotle, *Categories*, 14b17-23).

The claim is usually put in terms of truthmakers, and I will adopt that kind of talk here. The idea is that the truth of propositions depend in some way on other things, their truthmakers.

While it would be nice if the counterfactual analysis was able to capture these cases of dependence as well, I fear it is not to be. To handle these cases, we must either extend the counterfactual analysis, or we must invoke an alternative understanding of dependence.

To see why, we can appeal to another objection that can be raised against any counterfactual analysis. This objection is sometimes referred to as the problem of symmetry. Consider Socrates the man and the proposition 'Socrates exists'. We would like the truth of the proposition to metaphysically depend on the existence of the object. And we would like the dependence to be asymmetric. It should not be the case that the existence of the object metaphysically depends on the truth of the proposition.

So we examine the relevant counterfactuals:

(13a) Had Socrates not existed, then 'Socrates exists' wouldn't have been true.

(13b) Had 'Socrates exists' not been true, then Socrates wouldn't have existed.

We would like (13a) to be true and (13b) to be false. But the problem is that they both seem true. And it is important to note that this has nothing to do with impossible worlds. The antecedents of both (13a) and (13b) are merely contingently false. And it is not clear that to evaluate these counterfactuals, we need appeal to any impossible worlds.

One might think that the issue raised by the problem of symmetry makes an implicit appeal to something like the correspondence theory of truth. According to the correspondence theory, truths are true because they correspond to some way the world is, where a 'way the world is' is a fact.

CORRESPODENCE. *The proposition A is true if, and only if, there exists a fact f such that A corresponds to f.*

According to Correspondence it is both a necessary and sufficient condition for A to be true that it correspond to some fact. If we accept Correspondence, then we may be forced to accept both (13a) and 13b.[8] So if we don't want to do that, we could reject Correspondence.

But Correspondence is not crucial to the objection raised by the problem of symmetry. In fact, all you need to generate the problem is the Tarskian T-schema.

T-SCHEMA. $Tr\langle A\rangle \leftrightarrow A$

Here take Tr to be the truth predicate, and $\langle A\rangle$ to be a coding device for sentences A. Assume that $\leftrightarrow$ is the material biconditional. We can take the following as an instance of the T-schema.

(14a) $Tr\langle\text{Socrates exists}\rangle \leftrightarrow \text{Socrates exists}$

It is reasonable to take instances of the T-schema to be necessary truths. So we have:

(14b) $\Box[Tr\langle\text{Socrates exists}\rangle \leftrightarrow \text{Socrates exists}]$

Of course (14b) is equivalent to the conjunction of:

(14c) $\Box[Tr\langle\text{Socrates exists}\rangle \rightarrow \text{Socrates exists}]$, and

(14d) $\Box[\text{Socrates exists} \rightarrow Tr\langle\text{Socrates exists}\rangle]$

By contraposition we have:

[8] Assuming, of course, that 'Socrates exists' corresponds to the fact that Socrates exists. Denying this will allow one to reject both (13a) and (13b).

(14e) $\Box$[Socrates doesn't exist $\rightarrow \neg Tr\langle$Socrates exists$\rangle$]

(14f) $\Box[\neg Tr\langle$Socrates exists$\rangle \rightarrow$ Socrates doesn't exist]

Now, it's hard to argue with any of (14a) – (14f). But this raises problems if we would like to extend the counterfactual analysis to cases of dependence between truths and truthmakers. Consider again the relevant counterfactuals, in slightly different form.

(13a) Socrates doesn't exist $\Box\!\!\rightarrow \neg Tr\langle$Socrates exists$\rangle$

(13b) $\neg Tr\langle$Socrates exists$\rangle \Box\!\!\rightarrow$ Socrates doesn't exist

There is no reason to think that evaluating these counterfactuals requires any appeal to impossible worlds. But if the truth of these counterfactuals relies only on possible worlds, then we have that (14e) implies (13a), and (14f) implies (13b). In (14e) and (14f), the necessity operator ranges over all possible worlds; in (13a) and (13b), the necessity operator ranges over a proper subset of the possible worlds. And something that is true in all possible worlds is true in every subset of possible worlds. So if we accept those instances of the necessarily true T-schema, then we must accept both of those counterfactuals. Given the counterfactual analysis of dependence, we then have symmetric dependence between truths and their truthmakers.

There are only two ways to avoid this symmetric dependence. The first is to reject the necessary truth of the T-schema. There are some situations where this option is worth considering. For example, in the context of paradoxical sentences, like the Liar paradox or Curry's paradox, it may seem like the best thing to do is to reject the T-schema. But we are only motivated to do so in these cases because the sentences are paradoxical. For non-paradoxical sentences, it is hard to argue with the T-schema. And the examples here have no whiff of paradox. So the T-schema seems pretty solid in these cases.

The only other option is to say that the counterfactual analysis, at least in its current form, does not give a complete analysis of dependence of truth on truthmaker. Though it would be nice if it did, I can accept that it does not. It is not obvious that the dependence that holds between the existence of one object and the existence of another is exactly the same kind of dependence that holds between a truth and a truthmaker. And nothing we have said in the other chapters requires that it is. It may be that the dependence that holds between a truth and its truthmaker is counterfactual in nature, but requires an extension of the counterfactual analysis presented here. Or it may be something else altogether.

3.4. The Strangeness of Impossibility. Another objection, considered by Daniel Nolan (1997), claims that every possible world is more similar to the actual world than every impossible world. The impossible is somehow "strange" — so strange that no impossible world can be more like the actual world than some possible world. In Nolan's words, "the heavens will fall before (correct) logic fails us" (p. 550).

It should be noted, however, that Nolan's discussion of impossible worlds, in the context of counterfactual conditionals, appeals to an analysis of counterfactuals in terms of spheres of similarity. According to this picture, we can think of the actual world in the middle of the space of worlds, with concentric spheres extending outward from the actual world. This is a simplification, but it will do as a nice picture. Both Lewis and Stalnaker appeal to this picture when describing their logics of counterfactuals. It is this picture that the logics $\mathbf{C}_1$ and $\mathbf{C}_2$ try, in different ways, to capture. The space of worlds, both possible and impossible is partitioned and ordered with respect to similarity to the actual world. Each partition, or 'sphere' is an equivalence class such that if two worlds are in the same sphere, then they are equally similar to the actual world. The spheres are ordered according to similarity to the actual world. If w is in a sphere that is closer to the actual world than the sphere that contains v, then w is more similar to the actual world than v.

But the account of counterfactuals presented here does not rely on worlds being partitioned and ordered according to spheres of similarity. If this objection is to threaten the counterfactual analysis presented here, it must be formulated differently. The idea that the Strangeness of Impossibility objection is trying to capture is that the only relevantly similar worlds are possible worlds. A first pass at articulating this claim (in terms of the logical machinery presented in Chapter 6) might look like this:

STRANGENESS. *For any formula A, there is no impossible world w such that* aR_Aw.

(Here a is the actual world.) But there is no reason to think that Strangeness is true. In fact, the claim that there are counterpossibles that are false or non-trivially true forces us to explore the idea that for some counterpossible $A \boxright B$, there is an impossible world w such that $aR_A w$. In fact, even Nolan admits that Strangeness is not generally true (1997, p. 551).

However, perhaps a weaker version of Strangeness is more plausible.

WEAK STRANGENESS. *For any formula A, if there is a possible world v such that $aR_A v$, then there is no impossible world w such that $aR_A w$.*

Weak Strangeness does threaten the counterfactual analysis of dependence for some set-theoretic cases. Recall that the set {Socrates} metaphysically depends on the man Socrates, but not vice versa. We argued that the counterfactual *Had Socrates not existed, then {Socrates} wouldn't have existed either* is true. And it is true for reasons independent of one's views on Weak Strangeness. Rather it follows from considerations on the identity conditions for sets. But it is consistent with Weak Strangeness.

We also argued that the converse counterfactual *Had {Socrates} not existed, then Socrates wouldn't have existed either* is false. And this is not consistent with Weak Strangeness. We recognize that in the worlds that we consider to evaluate this counterfactual at least one of the following must fail:

(a) Socrates exists.
(b) Pairing is true.

We argued that in the worlds we consider, there is no principled way to choose which one of (a) or (b) must fail. As long as any world makes (a) true, then the counterfactual is false. But perhaps we can argue that Weak Strangeness is a principle that we should accept, thus giving us reason to think that the counterfactual is true. For if we allow that one of the worlds under consideration is such that neither {Socrates} nor Socrates exists (and which is possible in all other respects), then none of the worlds under consideration will be such that Socrates exists but {Socrates} doesn't, as such a world would be impossible.

We must consider, then, whether there is reason to think that Weak Strangeness is true. It is likely that Weak Strangeness, like Strangeness, can only be motivated by appealing to an understanding of the truth conditions for counterfactuals in terms of similarity spheres. Spheres of similarity extend out from the actual world. One can impose the condition on this picture of similarity that impossible worlds inhabit the spheres on the outer edges. Indeed, no impossible world shares a sphere with any possible world. Suppose then that we look to worlds that are relevantly similar, where {Socrates} doesn't exist. If any one of them is possible, then it will be a world where Socrates does not exist either. And so, given this ordering, every such world will be possible, and so every world will be such that Socrates does not exist either. On this account, the counterfactual *Had {Socrates} not existed, the Socrates wouldn't have existed either* would be true.

But again, this is not the account of counterfactuals we endorse. This kind of similarity ordering on classes of worlds may make for a nice picture. But having a nice picture is not a good reason to think that this is the correct way to understand counterfactuals. We understand counterfactual in terms of *ceteris paribus* clauses. And it may be that, even when we consider counterfactuals with contingently false antecedents, some worlds that are *ceteris paribus* like the actual world are impossible. At the very least, there is no reason to rule impossible worlds out in all cases just because they're impossible.

4. Conclusion

I would like to conclude by returning to an objection that we briefly considered in Chapter 6. The counterfactual analysis of dependence relies crucially on the analysis of counterfactuals with impossible antecedents. The majority of the antecedents we consider describe worlds where certain necessarily existing objects fail to exist. We looked at worlds where certain sets don't exist, where certain numbers don't exist, where certain number structures don't exist. We then tried to decide what else exists or fails to exist in these worlds.

The general worry is that it is simply too difficult to be sure of what happens in these worlds. And so, we cannot really be sure that in every *A* world that is *ceteris paribus* like the actual world, *B* holds as well. These worlds are too strange. Who knows what might happen? This worry, as we have already seen, is expressed by Field (1991).

> It is doubtless true that nothing sensible can be said about how things would be different if there were no number 17; that is largely because the antecedent of this counterfactual gives us no hints as to what alternative mathematics is to be regarded as true in the counterfactual situation in question (p. 237).

The worry is that when we consider worlds where these impossible antecedents hold, where these necessary objects fail to exist, we do not have enough information about how to think about what would happen in these worlds. For example, if we are thinking about a world where the number 17 fails to exist, does that mean this world does not have seventeen objects?

I think these worries are misguided. Remember, when we evaluate these counterfactuals, we are looking at worlds that are *ceteris paribus* like the actual world, except where the antecedents are true. They are as much like the actual world as they can be, but which differ only in what is required to make the antecedent true.

Most of the counterpossibles we consider have us look at worlds where certain objects, like numbers and sets, fail to exist. I think the best way to think about what would happen at these worlds is to ban these objects from entering into certain relations, relations that entail the existence of the relata. Crucially, what does not change are the identity conditions on any objects. We then argue that the identity conditions on certain abstract objects, like sets, or perhaps numbers if you are a structuralist about numbers, require that for something to be a particular object, it must be related in specific ways to other objects. If those other objects do not exist, then nothing can be related in this way to them. And so nothing can be the original object in question.

So I think that counterfactuals that express a relationship between the existence of two objects can be perfectly understandable, and we can be in a situation to judge whether they are true or false.

And so I disagree with Field that nothing sensible can be said about worlds where certain mathematical objects fail to exist. Of course, we might not be able to determine everything that follows from supposing that these objects do not exist. I have no idea whether or not there would be 17 objects, had the number 17 failed to exist. It's not obvious that the fact that the number 17 does not exist requires there to be fewer than 17 objects. Even nominalists who reject the existence of numbers believe that many objects exist. So I don't know how to answer that particular question.

But that does not mean that we cannot conclude anything about what would happen in such a world. Given our discussion of mathematical structuralism, a structuralist would say that, had the number 17 not existed, then none of the other natural numbers would have existed either.

So there are things that we can say about worlds where certain mathematical objects fail to exist. Once we remove these objects from the world, keeping as much as we can about the actual world fixed, we see what follows. This strategy works because one thing that need not change in these examples is the logic that holds at these worlds. We can reason about objects that do not exist at the actual world according to the rules of inference that (we believe to) hold at the actual world. Considering worlds where certain necessarily existing objects fail to exist should not force a change in which principles of reasoning about what happens at these worlds are valid. For the failure of certain objects to exist (at least in the cases we are interested in) should not require a change in logic.

Of course, there are cases when we do want to know what would happen if different rules of inference held, if a different logic was the correct logic. But we can say sensible things about these worlds as well.

(15a) Had intuitionistic logic been the correct logic, then the law of excluded middle would have failed.

(15b) Had intuitionistic logic been the correct logic, then the law of non-contradiction would have failed.

It is reasonable to think that (15a) is true and (15b) is false. But these examples have nothing to do with the existence of certain objects, necessary or otherwise.

I do not mean to say that every counterpossible is perfectly clear and guaranteed to sufficiently determine the worlds we should be looking at to evaluate them. Consider the counterpossible

(15c) Had classical logic not been correct, then the law of excluded middle would fail.

I'm inclined to say that, strictly speaking, this counterpossible is false. There are many ways that the logic of a world could have been different from classical logic. In some of these ways, excluded middle holds; in others it fails. But there are contexts in which we might be inclined to say this counterfactual is true. A debate, for example, between a classical logician and an intuitionist may establish a context where the only relevant options are classical logic and intuitionistic logic. In such a context (15c) is plausibly true, because any worlds that we look at where classical logic fails are worlds where intuitionistic logic holds. Given the approach to context discussed above, I am inclined to say that (15c) is strictly false, and what is really true is a more specific counterpossible. This counterpossible will capture the context by adding conditions into the antecedent specifying that the only alternative logics are classical and intuitionistic.

So Field's worry is justified in some instances. There are cases where it is not entirely clear what occurs in certain impossible situations. But it is not justified in all impossible situations. And it is certainly not justified when the impossible situations we are thinking about have to do with the failure of certain necessarily existing objects to exist.

Given that we keep the logic fixed, we can say sensible things about what would happen had certain necessarily existing objects failed to exist. We can say that, had the empty set not existed, then its singleton wouldn't have existed either. We can say that, had the singleton of the empty set not existed, it doesn't follow that the empty set wouldn't have existed either. If we accept a counterfactual analysis of dependence, as endorsed here, then we can further say that sets metaphysically depend on their members, but members do not metaphysically depend on the sets that they belong to.

Bibliography

Aczel, P. (1988). *Non-Well-Founded Sets*. Stanford: Center for the Study of Language and Information.

Addison, J. W., Henkin, L., & Tarski, A. (1965). *The Theory of Models*. Amsterdam: North Holland.

Aquinas (1968). *On Being and Essence,* Tr. with an introd. and notes by Armand Maurer, 2d revised edition. Toronto: Pontifical Institute of Mediaeval Studies.

Aristotle (1984a). Categories. In Barnes, J., editor, *The Complete Works of Aristotle, vol. 1*, pp. 3 – 24. Princeton University Press.

Aristotle (1984b). Metaphysics. In Barnes, J., editor, *The Complete Works of Aristotle, vol. 2*, pp. 1552 – 1728. Princeton University Press.

Armstrong, D. M. (1989). *A Combinatorial Theory of Possibility*. Cambridge: Cambridge University Press.

Armstrong, D. M. (1997). *A World of States of Affairs*. Cambridge: Cambridge University Press.

Barwise, J., & Etchemendy, J. (1987). *The Liar: An Essay on Truth and Circularity*. Oxford: Oxford University Press.

Barwise, J., & Moss, L. (1996). *Vicious Circles: On the Mathematics of Non-Wellfounded Phenomena*. Stanford: Center for the Study of Language and Information.

Beall, J., Brady, R., Dunn, J. M., Hazen, A., Mares, E., Meyer, R. K., Priest, G., Restall, G., Ripley, D., Slaney, J., & Sylvan, R. (2012). On the ternary relation and conditionality. *Journal of Philosophical Logic* **41**, 595 – 612.

Beall, J., & Restall, G. (2006). *Logical Pluralism*. Oxford: Clarendon Press.

Belnap, Jr., N. D. (1960). Entailment and relevance. *The Journal of Symbolic Logic* **25**, 144 – 146.

Berto, F. (2010). Impossible worlds and propositions: Against the parity thesis. *The Philosophical Quarterly* **60**, 471 – 486.

Bliss, R. L. (2011). *Against Metaphysical Foundationalism*. Ph. D. thesis, University of Melbourne.

Boolos, G. (1971). The iterative conception of set. *The Journal of Philosophy* **68**, 215 – 232. Reprinted in Boolos (1998), pages 13 – 29.

Boolos, G. (1989). Iteration again. *Philosophical Topics* **42**, 5 – 21. Reprinted in Boolos (1998), pages 88 – 104.

Boolos, G. (1998). *Logic, Logic, and Logic*. Cambridge, MA: Harvard University Press.

Brogaard, B., & Salerno, J. (2008). Counterfactuals and context. *Analysis* **68**, 39 – 46.

Burgess, J. P. (2004). E pluribus unum. *Philosophia Mathematica* **12**, 193 – 221. Reprinted in Burgess (2008), pp. 104 – 34.

Burgess, J. P. (2008). *Mathematics, Models, and Modality: Selected Philosophical Essays*. Cambridge: Cambridge University Press.

Cameron, R. (2008). Turtles all the way down: Regress, priority and fundamentality. *The Philosophical Quarterly* **58**, 1 – 14.

Cantor, G. (1883). Grundlagen einer allgemeinen mannigfaltigkeitslehre. ein mathematisch-philosophischer versuch in der lehre des unendlichen,. In Ewald (1996), pp. 878 – 920.

Carnap, R. (1946). Modalities and quantification. *Journal of Symbolic Logic* **11**, 33 – 64.

Chisholm, R. M. (1973). Parts as essential to their wholes. *Review of Metaphysics* **26**, 581 – 603.

Chisholm, R. M. (1976). *Person and Object: A Metaphysical Study*. La Salle, IL: Open Court.

Copeland, B. J. (2002). The genesis of possible worlds semantics. *Journal of Philosophical Logic* **31**, 99 – 137.

Correia, F. (2005). *Existential Dependence and Cognate Notions*. Munich: Philosophia Verlag.

Correia, F. (2007). (finean) essence and (priorean) modality. *dialectica* **61**, 63 – 84.

Correia, F., & Schnieder, B., editors (forthcoming). *Grounding and Explanation*. Cambridge: Cambridge University Press.

Cotnoir, A. J., & Bacon, A. (2012). Non-wellfounded mereology. *The Review of Symbolic Logic* **5**, 187 – 204.

Cresswell, M. J. (1972). The completeness of s1 and some related systems. *Notre Dame Journal of Formal Logic* **13**, 485 – 496.

Cresswell, M. J. (1995). S1 is not so simple. In Sinnott-Armstrong et al. (1995), pp. 29 – 40.

Davidson, D. (1970). Mental events. In Davidson (2001), pp. 207 – 227.

Davidson, D. (2001). *Essays on Actions and Events,* second edition. Oxford: Clarendon Press.

Descartes, R. (1637/1985a). Discourse on method. In Cottingham, J., Stoothoff, R., & Murdoch, D., editors, *The Philosophical Writings of Descartes, vol. 1*, pp. 111 – 151. Cambridge: Cambridge University Press.

Descartes, R. (1641/1985b). Meditations on first philosophy. In Cottingham, J., Stoothoff, R., & Murdoch, D., editors, *The Philosophical Writings of Descartes, vol. 2*, pp. 1 – 76. Cambridge: Cambridge University Press.

Descartes, R. (2008). *Discourse on Method and the Meditations,* trans. John Veitch. New York, NY: Cosimo.

Divers, J. (2002). *Possible Worlds*. London: Routledge.

Dunn, J. M. (1994). Star and perp: Two treatments of negation. In Tomberlin (1994a).

Etchemendy, J. (1988). Tarski on truth and logical consequence. *The Journal of Symbolic Logic* **53**, 51 – 79.

Ewald, W., editor (1996). *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*. Oxford: Clarendon Press.

Field, H. (1991). *Realism, Mathematics, and Modality*. Oxford: Basil Blackwell.

Field, H. (1993). The conceptual contingency of mathematical objects. *Mind* **102**, 285 – 299.

Fine, K. (1981). First order modal theories i: Sets. *Noûs* **15**, 177 – 205.

Fine, K. (1994). Essence and modality. In Tomberlin (1994b), pp. 1 – 16.

Fine, K. (1995a). Ontological dependence. *Proceedings of the Aristotelian Society* **95**, 269 – 290.

Fine, K. (1995b). Part-whole. In Smith (1995), pp. 463 – 85.

Fine, K. (2007). Response to fabrice correia. *dialectica* **61**, 85 – 88.

Fine, K. (2010). Towards a theory of part. *The Journal of Philosophy* **107**, 559 – 589.

Fine, K. (2012). A difficulty for the possible worlds analysis of counterfactuals. *Synthese*.

Fine, K. (unpublished). Counterfactuals without possible worlds. Available at http://philosophy.fas.nyu.edu/docs/IO/1160/Counterfactuals-jp.pdf.

Gabbay, D., & Wansing, H., editors (1999). *What is Negation?* Dordrecht: Kluwer.

Garfield, J. L. (1995). *The Fundamental Wisdom of the Middle Way: Nāgārjuna's* Mūlamadhyamakakārikā. Oxford: Oxford University Press.

Gödel, K. (1947). What is cantor's continuum problem? In Gödel (1990), pp. 176 – 187.

Gödel, K. (1990). *Collected Works, Volume II: Publications 1938-1974*. Oxford: Oxford University Press.

Gómez-Torrente, M. (1996). Tarski on logical consequence. *Notre Dame Journal of Formal Logic* **37**, 125 – 151.

Gómez-Torrente, M. (2009). Rereading tarski on logical consequence. *The Review of Symbolic Logic* **2**, 249 – 297.

Goodman, N. (1983). *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Hale, R., & Hoffman, A. (2010). *Modality: Metaphysics, Logic, and Epistemology*. Oxford: Oxford University Press.

Hintikka, J. (1962). *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca, NY: Cornell University Press.

Hoelte, M., Schnider, B., & Steinberg, A. (forthcoming). *Basic Philosophical Concepts*. Munich: Philosophia Verlag.

Husserl, E. (1901/1970). *Logical Investigations, vol 2,* trans. J. N. Findlay. London: Routledge.

Incurvati, L. (forthcoming). How to be a minimalist about sets. *Philosophical Studies*.

Jech, T. (2006). *Set Theory: The Third Millenium Edition, revised and expanded*. Berlin: Springer.

Jeffrey, R. (1965). *The Logic of Decision*. New York: McGraw Hill.

Jenkins, C. (2011). Is metaphysical dependence irreflexive? *The Monist* **94**, 267 – 276.

Jónsson, B., & Tarski, A. (1951). Boolean algebras with operators. *American Journal of Mathematics* **73**, 891 – 939.

Kanamori, A. (2009). *The Higher Infinite: Large Cardinals in Set Theory from Their Beginnings*. Springer-Verlag.

Kim, J. (2000). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.

Kim, J. (2007). *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.

Kiourti, I. (2010). *Real Impossible Worlds: The Bounds of Possibility*. Ph. D. thesis, University of St. Andrews.

Koslow, A. (2011). Structuralist modals and the combination of logics. *Logic Journal of the IGPL* **19**, 584 – 597.

Kripke, S. (1963). Semantical analysis of modal logic, i: Normal propositional calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* **8**, 113 – 116.

Kripke, S. (1965). Semantical analysis of modal logic, ii: Non-normal modal propositional calculi. In Addison et al. (1965), pp. 202 – 220.

Kripke, S. (1972). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Kunen, K. (2009). *The Foundations of Mathematics*. College Publishing.

Leitgeb, H. (2012a). A probabilistic semantics for counterfactuals. *The Review of Symbolic Logic* **5**, 26 – 84.

Leitgeb, H. (2012b). A probabilistic semantics for counterfactuals, part b. *The Review of Symbolic Logic* **5**, 85 – 121.

Lemmon, E. J. (1957). New foundations for lewis modal systems. *The Journal of Symbolic Logic* **22**, 176 – 186.

Lewis, C. I., & Langford, C. H. (1931). *Symbolic Logic*. New York: Dover.

Lewis, D. (1973). *Counterfactuals*. Cambridge, MA: Harvard University Press.

Lewis, D. (1983a). *Philosophical Papers,* vol. I. Oxford: Oxford University Press.

Lewis, D. (1983b). Postscripts to "anselm and actuality". In Lewis (1983a), pp. 21 – 25.

Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Basil Blackwell.

Lewis, D. (1991). *Parts of Classes*. Oxford: Basil Blackwell.

Linnebo, O. (2008). Structuralism and the notion of dependence. *Philosophical Quarterly* **58**, 59 – 79.

Loux, M. J., & Zimmerman, D. W., editors (2003). *The Oxford Handbook of Metaphysics*. Oxford: Oxford University Press.

Lowe, E. J. (1998). *The Possibility of Metaphysics: Substance, Identity, and Time*. Oxford: Clarendon Press.

Lowe, E. J. (2005). Ontological dependence. Stanford Encyclopedia of Philosophy. http://plato.stanford.edu/entries/dependence-ontological/.

Lowe, E. J. (2008). Two notions of being: Entity and essence. *Royal Institute of Philosophy Supplements* **83**, 23 – 48.

MacBride, F., editor (2006). *Identity and Modality*. Oxford: Clarendon Press.

Maudlin, T. (2007). *The Metaphysics Within Physics*. Oxford: Oxford University Press.

Mayberry, J. P. (2000). *The Foundations of Mathematics in the Theory of Sets*. Cambridge: Cambridge University Press.

McDaniel, K. (2004). Modal realism with overlap. *Australasian Journal of Philosophy* **82**, 137 – 152.

McLaughlin, B., & Bennett, K. (2008). Supervenience. Stanford Encyclopedia of Philosophy. http://plato.stanford.edu/entries/supervenience/.

Nolan, D. (1997). Impossible worlds: A modest approach. *Notre Dame Journal of Formal Logic* **38**(4), 535 – 572.

Parsons, C. (1990). The structuralist view of mathematical objects. *Synthese* **84**, 303 – 46.

Perszyk, K. (1993). Against extended modal realism. *Journal of Philosophical Logic* **22**, 205 – 214.

Plantinga, A. (1974). *The Nature of Necessity*. Oxford: Oxford University Press.

Plantinga, A. (1976). Actualism and possible worlds. *Theoria* **42**, 139 – 60.

Potter, M. (1990). *Sets: An Introduction*. Oxford: Clarendon Press.

Potter, M. (2004). *Set Theory and its Philosophy: A Critical Introduction*. Oxford: Oxford University Press.

Priest, G. (1979). The logic of paradox. *Journal of Philosophical Logic* **8**, 219 – 241.

Priest, G. (1995). Etchemendy and logical consequence. *Canadian Journal of Philosophy* **25**, 283 – 292.

Priest, G. (1997a). Notre Dame Journal of Formal Logic: Special Issue on Impossible Worlds, vol. 38.

Priest, G. (1997b). Sylvan's box: A short story and ten morals. *Notre Dame Journal of Formal Logic* **38**(4), 573 – 582.

Priest, G. (2000). Objects of thought. *Australasian Journal of Philosophy* **78**, 494 – 502.

Priest, G. (2005). *Towards Non-Being: The Logic and Metaphysics of Intentionality*. Oxford: Oxford University Press.

Priest, G. (2006). *In Contradiction: A Study of the Transconsistent,* expanded edition. Oxford: Clarendon Press.

Priest, G. (2008). *An Introduction to Non-Classical Logic: From If to Is,* second edition. Cambridge: Cambridge University Press.

Prior, A. N. (1958). The syntax of time-distinctions. *Franciscan Studies* **18**, 105 – 120.

Quine, W. V. O. (1948/1949). On what there is. *Review of Metaphysics* **2**, 21 – 38. Reprinted in Quine (1953), pp. 1 – 19.

Quine, W. V. O. (1953). *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

Resnik, M. D. (1981). Mathematics as a science of patterns: Ontology and reference. *Noûs* **15**, 529 – 550.

Restall, G. (1997). Ways things can't be. *Notre Dame Journal of Formal Logic* **38**, 583 – 596.

Restall, G. (1999). Negation in relevant logics (how i stopped worrying and learned to love the routley star). In Gabbay & Wansing (1999), pp. 53 – 76.

Rosen, G. (1990). Modal fictionalism. *Mind* **99**, 327 – 54.

Rosen, G. (2006). The limits of contingency. In MacBride (2006), pp. 13 – 39.

Rosen, G. (2010). Metaphysical dependence: Grounding and reduction. In Hale & Hoffman (2010), pp. 109 – 136.

Routley, R. (1980). *Exploring Meinong's Jungle*. Canberra: Australian National University, Departmental Monograph No. 3.

Schaffer, J. (2010). Monism: The priority of the whole. *Philosophical Review* **119**, 31 – 76.

Schaffer, J. (forthcoming). Grounding, transitivty, and contrastivity. In Correia & Schnieder (forthcoming).

Schnieder, B. (2007). Review of Correia existential dependence and cognate notions. *dialectic* **61**, 589 – 594.

Shapiro, S. (1991). *Foundations without Foundationalism: A Case for Second-order Logic*. Oxford: Clarendon Press.

Shapiro, S. (1997). *Philosophy of Mathematics: Structure and Ontology*. Oxford: Oxford University Press.

Shapiro, S. (2000). *Thinking About Mathematics: The Philosophy of Mathematics*. Oxford: Oxford University Press.

Shapiro, S. (2006). Structure and identity. In MacBride (2006), pp. 109 – 145.

Shapiro, S. (2008). Identity, indiscernibility, and *ante rem* structuralism: The tale of i and $-i$. *Philosophia Mathematica* **16**, 285 – 309.

Shields, C. (2007). *Aristotle*. New York, NY: Routledge.

Shoesmith, D. J., & Smiley, T. J. (1978). *Multiple-conclusion Logic*. Cambridge: Cambridge University Press.

Sider, T. (2003). Reductive theories of modality. In Loux & Zimmerman (2003), pp. 180 – 208.

Simons, P. (1982). The formalisation of husserl's theory of wholes and parts. In Smith (1982), pp. 113 – 59.

Simons, P. (1987). *Parts: A Study in Ontology*. Oxford: Clarendon Press.

Sinnott-Armstrong, E., Raffman, D., & Asher, N., editors (1995). *Modality, Morality, and Belief: Essays in Honor of Ruth Barcan Marcus*. Cambridge: Cambridge University Press.

Smith, B., editor (1982). *Parts and Moments: Studies in Logic and Formal Ontology*. Munich: Philosophia Verlag.

Smith, B., editor (1995). *The Cambridge Companion to Husserl*. Cambridge: Cambridge University Press.

Stalnaker, R. (1996). Impossibilities. *Philosophical Topics* **24**, 193 – 204.

Tarski, A. (1936). Über den begriff der logischen folgerung. *Áctes du Congrès International de Philosophie Scientifique* **7**, 1 – 11. Translated into English and reprinted in Tarski (1956).

Tarski, A. (1956). *Logic, Semantics, Metamathematics*. Oxford: Clarendon Press.

Tomberlin, J., editor (1994a). *Philosophical Perspectives,* vol. 7. Atascadero, CA: Ridgeview Publishing Company.

Tomberlin, J., editor (1994b). *Philosophical Perspectives,* vol. 8. Atascadero, CA: Ridgeview Publishing Company.

Trogdon, K. (forthcoming). An introduction to grounding. In Hoelte et al. (forthcoming).

Vander Laan, D. (1997). The ontology of impossible worlds. *Notre Dame Journal of Formal Logic* **38**, 597 – 620.

Wang, H. (1974). *From Mathematics to Philosophy*. London: Routledge and Kegan Paul.

Weber, Z. (2009). *Paradox and Foundation*. Ph. D. thesis, University of Melbourne.

Weber, Z. (2010). Extensionality and restriction in naive set theory. *Studia Logica* **94**, 87 – 104.

Wright, C. (1983). *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.

Yagisawa, T. (1988). Beyond possible worlds. *Philosophical Studies* **53**, 175 – 204.

Zermelo, E. (1930). Über grenzzahlen und mengenbereiche: Neue untersuchungen über die grundlagen der mengenlehre. *Fundamenta Mathematicae* **16**(29 - 47).