2018-04-25

Counterpossibles for Modal Normativists

Theodore Locke
University of Miami, theolocke@gmail.com

UNIVERSITY OF MIAMI

COUNTERPOSSIBLES FOR MODAL NORMATIVISTS

By
Theodore D. Locke

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

May 2018

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

COUNTERPOSSIBLES FOR MODAL NORMATIVISTS

Theodore D. Locke

Approved:

_____ Otávio Bueno, Ph.D. Professor of Philosophy	_____ Berit Brogaard, Ph.D. Professor of Philosophy
_____ Elijah Chudnoff, Ph.D. Associate Professor of Philosophy	_____ Guillermo Prado, Ph.D. Dean of the Graduate School

_____
Amie L. Thomasson, Ph.D.
Professor of Philosophy
Dartmouth College

_____
Daniel Nolan, Ph.D.
Professor of Philosophy
University of Notre Dame

LOCKE, THEODORE D. (Ph. D., Philosophy)
Counterpossibles for Modal Normativists (May 2018)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Amie L. Thomasson and Otávio Bueno.
No. of pages in text. (253)

In this dissertation, I develop an account of non-vacuous counterpossibles—counterfactuals involving metaphysical impossibilities—and related notions, e.g. metaphysical similarity between impossible worlds, that does not require us to take on questionable ontological commitments and that gives us a clear epistemological story about how we know counterpossible claims. My account of counterpossibles builds on a non-descriptivist account of metaphysical modality called *modal normativism*, which has been developed by Amie Thomasson. According to modal normativism, claims of metaphysical necessity and possibility are *not* descriptive claims in need of modal truthmakers, but instead serve the normative function of enabling language users to illustrate or express constitutive rules that govern the use of ordinary non-modal vocabulary but while staying in the object language *using*, rather than mentioning, the terms. Roughly, on my account, the evaluation of metaphysical counterpossibles only requires us to *tacitly* consider how the actual rules that govern the use of our terms are changed in a deviant, yet relevantly similar, linguistic framework that accommodates the description of some hypothetical impossible scenario. After presenting my view, I respond to two general worries one might have about my account. Finally, I offer a sketch of two easy ways to think about impossible worlds for those who might feel uncomfortable adopting such talk.

Acknowledgements

I would like to thank Amie Thomasson for being a fantastic dissertation advisor and a source of philosophical inspiration over the years. Many thanks also to my other University of Miami committee members Berit Brogaard, Otávio Bueno, and Eli Chudnoff. Also, thank you to the University of Miami Department of Philosophy, Colleges of the Arts and Sciences, Offices of the Provost and the Vice Provost for Research, and the Ethics Program for supporting grants I received during the summer of 2015 and the summer of 2017 and which allowed me to complete two papers that also function as dissertation chapters. Much gratitude goes to Daniel Nolan for agreeing to be an external committee member, for working with me during my visit to Notre Dame during the fall of 2016, and whose work has inspired me to think of as many six impossible things before breakfast every day. Thank you to all of my friends and colleagues for the good times, good conversations, and good philosophy. Finally, thank you to Kimberly Martin at the University of Miami counseling center for all of her support over the past six years.

CONTENTS

CHAPTER 1: INTRODUCTION

1.1 What are Counterpossibles?

We often use conditionals to express various relationships between certain claims, e.g. evidential or explanatory relationships. Sometimes we express these conditional relationships in an indicative mood, other times we use counterfactual conditionals. For example, when offering an explanation of *why* Smith lost the election in terms of where she campaigned, we might say:

(0) If Smith had spent more time campaigning in Jonestown, she would have won the election.

According to an "orthodox" philosophical account of counterfactual conditionals, to decide whether (0) is true we hypothetically consider a possible world relevantly similar to the actual world but where (unlike in the actual world) it is true that Smith spent more time campaigning in Jonestown.[1] If we have good reason to believe that in such a hypothetical world Smith also would have won the election, then (0) is true; otherwise, it is false. Moreover, we might provide (0) as evidence that a contributing factor to Smith losing the election is that she did not spend enough time campaigning in Jonestown.

[1] Roughly, a "possible world" is a complete way that the actual world *could have been*. For example, in the actual world I went to graduate school for philosophy, but intuitively there is possible world where I went to medical school instead. An "impossible world" is a way that the actual world *could not have been*. For example, in the actual world there is no sentence that can be both true and false. As we will see, however, we might think that there is an impossible world pretty much like the actual world but in which *some* sentence is both true and false. Explaining the nature of possible or impossible worlds is beyond the scope of the project I am embarking on here. However, I will have some things to say about them in chapter 3, section 2, and in the final chapter.

Even though Smith did not spend a lot of time campaigning in Jonestown, we all agree that *she could have*. So, talking about and considering non-actual possibilities can be useful when offering certain explanations. But, here are some interesting questions: first, is talking about and considering what *could not be the case*, i.e., what is *impossible*, ever useful in the way, say, that (0) is; and, second, if so, how do we decide when counterfactual conditionals involving impossibilities are true or *false*? I will eventually address the first question, but for now consider that the orthodox account of counterfactuals used to evaluate (0) cannot help us answer the second question.

More generally, on the orthodox account of counterfactual conditionals:

> A counterfactual 'If *p* were the case, then *q* would be the case' is true just in case either *p* and *q* are both true in the possible world most similar to the actual world or *there is no possible world in which p is true*.[2]

Note the italicized clause: on the orthodox account, conditionals with impossible antecedents get their truth for "free," i.e. they are *trivially true*. However, there are *prima facie* counterexamples to the orthodox view. Consider the following counterpossible:

(1) If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the same time would have cared.[3]

(1) seems to be *false*, not true. In an *impossible* world pretty much like the actual world but in which the 17th century English philosopher Thomas Hobbes *impossibly* squares a circle, it would *not* be the case that sick children in the mountains of South America

[2] For more specific accounts, see Lewis (1973), Stalnaker (1968).
[3] See Nolan (1997).

would care; thus, we have reason to think that (1) should be evaluated as *false*, not as trivially true. Hence, the orthodox account of counterfactual conditionals is wrong.

In this dissertation I develop a novel account of *counterpossible conditionals* and a number of related notions such as *metaphysical similarity* between worlds. I will underwrite my account with a non-descriptivist theory of metaphysical necessity called *modal normativism*, developed by Amie Thomasson.[4] Modal normativism holds that claims about what is metaphysically necessary, e.g. 'necessarily, all bachelors are male', are not descriptive claims in need of modal truthmakers, but instead are object language expressions of constitutive rules, or their consequences, governing nonmodal terms. Why modal normativism? Because, as I will explain in 3.1, modal normativism provides a plausible account of metaphysical necessity that comes with epistemological advantages, which my account of counterpossibles inherits.

This project is relevant because, as I will discuss in the next section, non-trivial counterpossibles play an important role in articulating certain philosophical views and explanations.[5] Additionally, there is good reason to think that counterpossibles have a lot of work to do in the recently discussed project of conceptual engineering and conceptual ethics.[6] So, this project will extend modal normativism in a crucial way by providing an account of counterpossibles and related phenomena that is consistent with the deflationist modal project. Conversely, approaching counterpossibles by starting with a normativist approach to metaphysical necessity gives us a novel way of thinking about many ideas

[4] See Thomasson (2007a, 2007b, 2009, 2013, and forthcoming).
[5] See Fine (1994), Nolan (1997), Kment (2006, 2014), Brogaard and Salerno (2013), and Wilson (2016), among others.
[6] For examples and discussion of conceptual engineering and conceptual ethics see: Burgess and Plunkett (2013), Haslanger (2012), Plunkett (2015), Tanswell (2017), and Thomasson (2017).

related to counterfactuals and counterpossibles such as a distinctively metaphysical notion of similarity between worlds and so-called metaphysical laws. Importantly, my account removes some of the mystery behind metaphysical impossibilities, gives a clear epistemological story of how we understand them, and shows why we might (and should) care about metaphysical impossibilities at all.

1.2 Outline

My central aim: *I will give a novel theory of counterpossibles that can account for semantic intuitions, that can do interesting philosophical work, and that avoids the worries that come with certain epistemic readings and substantive metaphysical accounts*.[7]

I think that a non-descriptivist approach to modality (modal normativism) and a deflationary approach to semantics (use-theories of meaning) provide the best resources for achieving my aim.[8] A picturesque way of summarizing my account is as follows. Considering both possible and impossible worlds, if we are going to *hypothetically* accept that the world is some way other than it actually is, then counterfactuals and counterpossibles convey a requirement about what kinds of worlds we ought accept: if you're going to hypothetically accept a *p* world, it *ought* to also be a *q* world. I grant that another way of making the same point is to say that *p q* worlds *are more relevantly similar* to our world than every *p* non-*q* world, but on my view claims about worlds and similarity are not primarily descriptive but, instead, are expressive claims that illustrate

[7] By a "substantive metaphysical account of X" I mean an account of X that posits the existence of worldly properties and relations that cannot be discovered through conceptual analysis or straightforward empirical methods.
[8] My focus will be on countermetaphysicals, and I will not consider counternomologicals.

the actual constitutive rules that govern the use of '*p*' and '*q*' given additional relevant conceptual or empirical information being held fixed by the context. So, as we will see, my account respects other extended Lewis-Stalnaker accounts of counterfactuals and counterpossibles by keeping talk of similarity and worlds *simpliciter*, which allows for talk of both possible and impossible worlds.

In less picturesque terms, in the end I develop a theory of counterpossibles with the following two features:

1. In a very qualified sense, I develop a *metalinguistic theory of counterpossibles*, where the main qualification is that my account is a *prescriptive* account of counterpossibles and *not* a descriptive.
2. In the end, I develop an account of counterpossibles whereby competently evaluating counterpossibles comes down to having:
    i. a tacit understanding of, or ability to follow, the rules and permissions that govern the terms of our "home" language, perhaps along with other relevant empirical information; and
    ii. a tacit understanding of, or ability to follow, the rules governing a language that is different yet relevantly similar to our "home" language.

The novelty of this approach can be further highlighted by noting that there are three different criterion we might appeal to when evaluating a counterpossible: an *epistemic* criterion, a *metaphysical* criterion, and a *conceptual* criterion. As far as I can tell, most of the literature has been devoted to exploring the epistemic and metaphysical criteria. So, my account breaks new ground in developing an interesting account of counterpossibles that utilizes a conceptual criterion.

The overall path I shall take to accomplish my aim is:

1. Show that non-vacuous counterpossibles are theoretically interesting and useful (Chapter 1).
2. However, there are problems with the available epistemic accounts and substantive metaphysical accounts of non-vacuous counterpossibles (Chapter 2).
3. We can overcome these problems by appealing to modal normativism and use-based accounts of meaning (Chapter 3).
4. However, there are two worries with doing this:
   (i) First, if we only appeal to deflationary approaches to modality (modal normativism) and semantics (use-theories of meaning), we won't be able to account for the theoretically interesting and useful things we can do with counterpossibles I discuss in Step 1, e.g. explaining why Socrates's existence doesn't *metaphysically depend* on {Socrates} (Chapter 4); and
   (ii) Second, that if we don't take the work I outline for counterpossibles in Step 1 seriously, as might those who are friendly to deflationary approaches to metaphysics and semantics, then we don't need counterpossibles after all (Chapter 5).
5. I show that the first worry doesn't hold because my account can do that work (Chapter 4).
6. I then address the second worry by showing that counterpossibles are useful for those that think of many disputes in philosophy are ultimately conceptual and not descriptive (Chapter 5).

7. Thus, I give an account of counterpossibles that can do interesting philosophical work that avoids the worries that come with the existing epistemic and substantive metaphysical accounts.

I will now briefly summarize each step.

Step 1:

The main point of step 1, and why it is needed, is to motivate and provide an overall background for the project. I will sometimes use expressions like "taking talk of counterpossibles (and related issues) seriously". What I mean is both: accepting counterpossibles are sometimes *non-vacuous*, e.g. there are some false counterpossibles; and, more importantly, recognizing that thinking about impossibilities, and thinking about the way some impossibilities may be more similar to actuality than others, is an indispensable tool in getting clear on many philosophical questions. So, if semantic intuitions provide reason to think there are non-vacuous counterpossibles, and if there are reasons to think that our talk and consideration of counterpossibles can play an indispensable role in our cognitive and theoretical lives, then we need a philosophical account of non-vacuous counterpossibles. Given this, there are at least three *desiderata* for any account of counterpossibles:

1. An account of counterpossibles should be consistent with our semantic intuitions, i.e. provide for the non-vacuity of counterpossibles—in particular, allow that there are some false counterpossibles;
2. An account of counterpossibles needs to enable counterpossibles to do the theoretical work we want them to;

3. An account of counterpossibles should avoid problematic ontological commitments and provide a clear story about how we come know counterpossible claims as well as related claims.

An account that meets all of these desiderata is better than alternatives that only meet some or none of them. In the end, my account meets all three.

I bring the issue of non-vacuity into my project because that is the way the literature on counterpossibles is often introduced and motivated. The reason why is that we have plausible intuitions that some counterpossibles are false and some are true (or at least many of the people in the literature I am engaging with have those intuitions). Even assuming a deflationary view of truth, as I will end up doing, we still need to account for these intuitions. For example, even if we treat the truth predicate merely as a device of generalization, and if we accept intuitions that counterpossibles are sometimes false, then if not everything Jones said is true, we still need to know if some of the false things Jones said are counterpossibles. Or, if we think that truth somehow merely comes down to assertibility, then we still need to know why it is not appropriate to assert certain counterpossibles and appropriate to assert others.

The issue of theoretical work comes into my project because, aside from being an interesting topic in its own right, the usefulness of counterpossibles in answering certain philosophical questions is another way the literature on counterpossibles is often motivated.[9] So, later in this chapter, I will provide a summary of what that work is. However, in chapter 5, I will introduce new work for a theory of counterpossibles.

---

[9] See especially Nolan (1997), Kment (2014), Bernstein (2016), and Wilson (2016), and Nolan *forthcoming*.

The issue of non-vacuity relates to the issue of theoretical work in the following ways. First, Williamson (2016, 2017) argues that we have theoretical reasons to keep counterpossibles *vacuous*. For example, Williamson argues that we must keep counterpossibles vacuous to account for their use in *reductio* proofs. I argue that we can still respect this reason while keeping an account of counterpossibles such that *some* are false. Furthermore, I argue that we have additional theoretical reasons for thinking that we *should* treat some counterpossibles as false and not vacuously true. This is the second way the issue of non-vacuity relates to the theoretical work.

It seems plausible enough to think that if an explanation for *X* implies something false, then that explanation for *X* is weak or no good at all. So, for example, if explaining the deontic properties of actions in terms of God's will implies 'if God were to (impossibly) command people to murder, then murder would be acceptable', which seems false and not vacuously true, we have reason to think that God's will isn't the right sort of thing to explain the deontic properties of actions. So, an account that allows for some counterpossibles to be false is preferable over an account that concedes Williamson's point, and thus fails to meet one of the three *desiderata*. This is something that my account of counterpossibles can do.

Some might want to add another *desideratum* my list: An account of counterpossibles should explain what *makes* counterpossibles true. As we will see, my account does *not* meet this *desideratum*. This is because I reject truthmakers for counterpossible claims. I do this because I think that there are good reasons to think that once we start requiring truthmakers for modal claims we run the risk of having to take on problematic ontological commitments, e.g. a space of Lewisian possible worlds, and run

the risk of making it unclear how we know certain modal claims, e.g. how we have cognitive access to Lewisian worlds, which are causally disconnected from the actual world. However, I also reject truthmakers for modal claims because it would be awkward to build an account of counterpossibles using normativist and expressivist tools, which reject the claim that modal claims are descriptive claims in need of truthmakers, and at the same time think that there are some facts that make counterpossibles true. So, when I talk about "giving an account of counterpossibles" or talk about "giving an account of metaphysical similarity" that meet the above *desiderata*, I do not mean also giving an account of what makes counterpossibles true. Instead, I think it is better to say *we need to give an account of what we are doing* when we make counterpossible claims, make distinctions between impossibilities, compare impossibilities, etc.—an account which can also guide us in figuring out when we do these things properly, and preferably an account that meets the first three desiderata listed above.

Step 2:

The main point of step 2, which will be carried out in Chapter 2, is to outline worries for other approaches to non-vacuous counterpossibles that are currently on the market. The main worry with the epistemic approaches to counterpossibles I look at, which focus on how things appear or are represented to some agent in a context, is that they are too restrictive in their readings. An account of counterpossibles should not imply that 'if a steel Penrose triangle were placed in a 4000 deg. F oven, it would melt' is *about* how things impossibly appear to someone. While substantive metaphysical readings avoid this worry, it does so by appealing to substantive theoretical tools, e.g. substantive

properties of essence or substantive relations of metaphysical dependence, which come with theoretical and epistemological worries that are also best to avoid.

Establishing these worries, is important for two reasons. First of all, if our *only* available options for non-vacuous counterpossibles have unwanted or unavoidable worries, then my arguments for counterpossibles in step 1 will be potentially undermined—after all, maybe the theoretical cost of non-vacuous counterpossibles isn't worth it if they raise larger questions than questions as to why we sometimes think that counterpossibles are false or why we sometimes appeal to counterpossibles when offering certain explanations. However, if we have an account of counterpossibles that avoids these worries, then there is no problem. Secondly, outlining worries for prominent approaches to non-vacuous counterpossibles that are currently on the market further motivates why anyone should take my project seriously, and motivates why we should even bother looking at a non-descriptivist account of modality (modal normativism) and a deflationary view of semantics (use-theory of meaning) to explain counterpossibles and metaphysical similarity.

Step 3:

The main point of step 3, which will be carried out in Chapter 3, is to give my account of counterpossibles, which can allow for our semantic intuitions, can do interesting philosophical work, and avoids the worries that come with certain epistemic readings and substantive metaphysical readings. This step is needed because I establish the need for a novel approach to counterpossibles in steps 1 and 2. More importantly, this step is where I demonstrate my central achievement: *a novel account of counterpossibles that utilizes resources from a non-descriptivist account of modality (modal normativism)*

and deflationary account of semantics (use-theory of meaning). The main novelty of my account stems from my development of an account of metaphysical similarity in terms of relations of conceptual similarity between non-modal expressions of relevantly similar languages, which has not been done before.

Nolan (1997) shows that we can use a Stalnaker-Lewis framework to give an account of counterpossibles by generalizing the framework to include impossible worlds in addition to possible worlds. However, this leaves open two questions: how should we think about metaphysical similarity between possible and impossible worlds; and how should we think about impossible worlds? In my project, I focus on metaphysical similarity; why? I focus on metaphysical similarity because judgments of metaphysical similarity do more work in thinking about counterpossibles than however it is we think about what impossible worlds are. For example, most of the problems that come with the epistemic and substantive realist readings come from how they account for similarity between worlds: the epistemic readings are too restrictive, and the substantive realist account appeals to questionable resources such as essences, grounding, and structure. So, I take giving an account of metaphysical similarity to be where the primary theoretical need for an account of counterpossibles lies. That said, I will have some things to say about worlds in chapter 3, section 2, and chapter 6.

My account of counterfactuals and counterpossibles utilizes an *expressivist element*:

> A counterfactual 'If *p* were the case, then q *would* be the case' *expresses* a *requirement* to accept that *q* given *p* as a hypothetical supposition and a range of *relevant auxiliary assumptions* determined by the context. (*Expressive Element*)

and a *similarity element*:

> A counterfactual 'If *p* were the case, then q *would* be the case' is *assertible* in *w* just in case there is some *p q* world that more relevantly similar to w than every *p* non-*q* world. (*Similarity Element*).

One might think that there is a certain amount of awkwardness in giving a presentation of counterfactuals in expressivist terms *and* in terms of worlds and similarity. In chapter 3, section 2, I will show how this is not as awkward as one might think. For now, let me say that I adopt the expressivist element because it seems to avoid problematic ontological commitments and provides a clear story about what we do with counterpossibles and how we come know counterpossible claims. I adopt the similarity element, in part, for dialectical reasons since the alternative views of counterpossibles I am engaging with all work within the extended Lewis-Stalnaker framework. But, more importantly, I do this because I take counterfactual and counterpossible claims *and* claims about similarity between worlds, e.g. similarity between various impossible worlds and the actual world with respect to certain essential properties or other necessary metaphysical truths, to *all* be object language expressions of rules and permissions that govern the use of non-modal terms found in our "home" language and languages relevantly similar to our "home" language. This is a novel way of thinking about metaphysical similarity that avoids heavy ontological commitments and only relies on conceptual analysis to explain how we come to know certain counterpossibles.

Steps 4-6:

In steps 4-6 I do two things, both of which are subservient to my central aim. First, I make good on my claim that my account of counterpossibles can account for

semantic intuitions, can do interesting philosophical work, both of which are discussed in step 1, and my account does so in a way that avoids the worries that come with the readings discussed in step 2.

These steps are needed because the substantive realist justifies many of the theoretical resources in her account of counterpossibles by arguing that they are needed to make sense of the role counterpossibles play in offering metaphysical explanations or explanations of comparative modal claims. So, one might be worried that if we want to use counterpossibles to do these things, we will be required to use the substantive realist's resources. If the worry is true, then I am forced into a dilemma. Either: a.) I must retract the job description I outline for counterpossibles later in this chapter, which undermines my argument that we need an account of counterpossibles and metaphysical similarity; or b.) I must after all appeal to the substantive metaphysical properties and relations I raise worries for in chapter 2. However, in Chapter 4, I show how to use a non-descriptivist account of modality (modal normativism) and a deflationary account of semantics (use-theory of meaning) to understand non-epistemic counterpossibles and, more importantly, given an account of the role counterpossibles play in metaphysical explanations, e.g., an explanation of why Socrates's existence doesn't metaphysically depend on {Socrates}.

In chapter 5, I will do two related things. First, I will look at a worry to my overall argument that comes from a theoretical perspective that is friendlier to the overall deflationary approach I am taking here. Roughly, the worry is that maybe we really shouldn't care about debates over essences, over the nature of things, over metaphysical dependence, etc. It might be that many of these debates are intractable because people are

simply playing word games when debating about the essential properties or natures of things. If that's the case, then maybe the related counterpossible claims are not that interesting and not worth all the theoretical fuss I use to motivate this project. In my response to this worry, I will look at recent literature on conceptual engineering and argue that this literature provides new work for a theory of counterpossibles. Furthermore, I will show that my account of counterpossibles can even do that work and do work for those who want to embrace certain forms of metaphysical conventionalism.

Finally, in chapter 6, I will offer a brief summary of the strengths of my account. I will also briefly say something about impossible worlds, which I freely talk about throughout this project. Giving and defending a full account of the nature and epistemology of impossible worlds is a dissertation project all on its own. However, I hope to at least sketch two ways it can be done while keeping with the overall deflationary spirt I take in developing my account of counterpossibles for modal normativists.

1.3 Orthodoxy Vs. Non-Orthodoxy

Counterpossibles are counterfactual conditionals involving some impossibility, usually in the antecedent.[10] An *impossibility* is simply a way things could not be, i.e. a circumstance that cannot be the case. There are a variety of ways things could not possibly be. In more austere contexts, scenarios in which the actual physical laws are violated will count as *nomologically impossible*. In more liberal contexts, some violations

[10] Nolan (1997, 2011) and Vander Laan (2004) give examples of potentially non-vacuous counterpossibles with possible antecedents and impossible consequents, which will be briefly discussed shortly.

of physical laws will be possible, but scenarios in which certain metaphysical truths are different will count as *metaphysically impossible*. We can plausibly include mathematical and logical truths in the set of metaphysically necessary truths. So, worlds in which certain logical or mathematical truths fail to obtain will count as metaphysically impossible. However, there are many other truths besides which can also plausibly be taken to be metaphysically necessary. For example, many facts about essential properties, the nature of free will, the nature of God (if she exists), the nature of mereological sums, etc. are neither mathematical nor logical truths but are plausibly still metaphysically necessary. I am not only concerned with mathematical or logical impossibilities, but with this broader range of impossibilities. So, I will assume that any circumstance in which these facts are violated will count as metaphysically impossible.[11]

Recently there has been a lot of debate about the metaphysics and semantic status of counterpossibles. Consider the following *orthodox account of counterfactuals* taken from Vander Laan (2004), and which is based on accounts developed in Lewis (1973) and Stalnaker (1968):

> If *A* were the case, *C* would be the case" is true iff (i) there are no possible *A* worlds; or (ii) some possible world in which both the antecedent and the consequent are true (an *A C* world) is more similar to the actual world than every possible world in which the antecedent is true and the consequent is false (an *A* $\sim C$ world). (260)

Note that, by (i), counterfactuals with impossible antecedents are *vacuously true* in the sense that, whenever *A* is impossible, the truth value of the consequent makes no

[11] Whether there is a set of truths that is distinctly metaphysical and what constitutes being included in that set are interesting questions that I cannot fully address in this project. For an interesting and detailed discussion see Nolan (2011). Again, I will not say anything about distinctively nomological impossibilities in this dissertation.

difference to the truth value of the counterfactual as a whole. However, critics of the orthodox account think that this is wrong. To start, there is *prima facie* evidence in the form of intuitions that some counterpossible conditionals are false, not true, and others that are true but non-vacuous in the sense that they might be made false by making changes to the consequent. Consider the following examples:

(1) If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the same time would have cared.
(2) If water had not been H2O, then water would not have been water.
(3) If George Eliot had been a biological parent of Sean Spicer, then Sean Spicer would have been the empty set.
(4) If there were a 36-sided steel Platonic solid, then it would have more than 12 sides.[12]

Given defensible metaphysical assumptions, all four conditionals have a necessarily false antecedent. However, it seems that (1)-(3) are false, not true. For example, under the assumption that facts about our biological ancestry are essential, it still seems reasonable to consider what would be the case if someone (impossibly) had different biological parents. In most contexts, we would not think that in such impossible circumstances the person would be the empty set, as opposed to simply still being a human being. On the other hand, (4) seems true, but non-vacuously insofar as it seems that if we change the consequent to claim that the 36-sided steel Platonic solid would have had more than 100

[12] Some of these examples (and many more like them) can be found in Nolan (1997), Jenkins and Nolan (2013), Brogaard and Salerno (2013), Vander Laan (2004), and Kment (2014). See also Williamson (2007, 2016, and 2017), though Williamson argues against the idea that there are non-vacuous counterpossibles, which will be discussed shortly.

sides, we get something false.[13] So, the argument goes, to say that all four claims are vacuously true is a mistake.

Nolan (1997) offers a way to accommodate these intuitions by extending the Lewis-Stalnaker account of counterfactuals to include impossible worlds in addition to possible worlds. Following this strategy, Vander Laan (2004) proposes the following generalized analysis of counterfactuals:

> A counterfactual, 'If *A* were the case, then *C* would be the case', is true iff some (possible or impossible) *A C* world is more similar to the actual world than every *A* non-*C* world.[14] (265)

For example, we can say that (1)-(3) are false because, given an appropriate similarity relation determined by context, at the impossible worlds where the antecedent obtains that are most similar to the actual world it doesn't look as though the consequents obtain as well.[15] In other words, impossible worlds where Sean Spicer has different biological parents but is still otherwise the same are still more similar to the actual world than impossible worlds where Sean Spicer is an abstract object or impossible worlds where everything is the case. More needs to be said about how to determine similarity; in particular, more need to be said about how to think about a distinctively *metaphysical notion of similarity*, e.g. similarity with respect to facts about essence. I will discuss this notion of similarity more in the next chapter, and I will argue in chapter 3 that the modal normativist view gives us a plausible and preferable way of thinking about metaphysical

---

[13] Platonic solids are polyhedrons with either 4, 6, 8, 12, or 20 faces. Mathematically, there are only those five Platonic solids. So, the existence of a Platonic solid with 36 faces is impossible.

[14] See also Brogaard and Salerno (2013).

[15] In addition to Nolan (1997) and Vander Laan (2004), see also Krakauer (2012), Brogaard and Salerno (2013), and Kment (2006, 2014) for similar expanded Lewis-Stalnaker accounts. Others that seem to support non-vacuous readings counterpossibles include Kim and Maslen (2006), Jago (2014), Bernstein (2016), Vetter (2016), among others.

similarity. However, not everyone agrees that intuitions regarding (1)-(4) require that we abandon orthodoxy.

1.4 Why Epistemic Two-Dimensionalism is Not Enough

It was just suggested that we can accommodate our counterpossible intuitions by appealing to an extended Lewis-Stalnaker framework of counterfactuals that includes *both* possible and impossible worlds in our modal space. However, two potential objections to this *un*orthodox approach are: one, we don't need to talk about impossible worlds to accommodate our intuitions; and, two, we would be better off without impossible worlds.

One alternative suggestion is that we can and should reject the unorthodox approach in favor of an approach that utilizes an *epistemic two-dimensional semantics* with possible worlds only. Consider (2) and (3) above, which seem false and not true. On an epistemic reading, the antecedents of (2) and (3) are meant to be epistemic possibilities that only *appear* to be metaphysically impossible. What we are considering when we evaluate the above counterfactuals are not worlds where *actual* water is not H_2O or where *actual* Sean Spicer has a different biological origin than he actually has. Instead, we are considering worlds that are qualitatively very much like the actual world, i.e. a world in which the watery stuff in our environment is not H_2O and a world where a person qualitatively indistinguishable from Sean Spicer has a different biological origin. However, these are not *impossible* worlds; they are quite possible. The contrary appearance results from the fact that antecedent sentences are associated with two different sets of worlds—one set that provides an "epistemic" meaning for the sentence

and another that provides an "external" meaning—but these worlds inhabit a single modal space of *metaphysically possible worlds only*.[16]

There are at least two suggestions for why this epistemic two-dimensionalist approach might seem appealing. One suggestion is that by taking this approach we can avoid potential worries about what exactly impossible circumstances or worlds are, as opposed to possible circumstances or worlds, thus giving us a simpler modal ontology. The second suggestion is that if we do not have to appeal to impossibilities, we might be able to keep a reliable link between conceivability and possibility, thus giving us a plausible modal epistemology. The worry is that on the extended Lewis-Stalnaker approach proposed above, we can sometimes conceive of worlds that are epistemically possible but not metaphysically possible. In which case, conceivability will not be a good guide to what is metaphysically possible. So, the epistemic two-dimensionalist wants to argue that if we want to have a simple modal ontology and reliable guide to the limits of possibility, we need to reject impossible worlds and the extended Lewis-Stalnaker approach. However, there are three reasons why this is either not a good argument or we should in fact reject the epistemic two-dimensionalist approach to counterpossibles.

First, Ripley (2012) and Yagisawa (2010) independently argue that even the epistemic two-dimensionalist must rely on impossible circumstances when evaluating

---

[16] See Chalmers (2002) and Schroeter (2017), Section5 for a general overview of rationalist two-dimensional semantics. Very roughly, on this view, expressions, e.g. 'water', are associated with two *intensions*, or meanings: an external meaning determined by external facts, e.g. the chemical properties of a substance such as water (chemical properties that are had *necessarily*), and an epistemic meaning determined by internal facts, e.g. how that substance *appears* to an agent. The rough idea is that we could (mistakenly) associate the term 'water' with two qualitatively identical substances that each fill two different drinking glasses, but the chemical structure of the substance in one glass is H_2O while XYZ in the other (and so the two substances are necessarily distinct). Brogaard and Salerno (2013) offer something like an epistemic two-dimensionalist account, but one that appeals to impossible worlds. I will critically discuss this view in the next chapter.

certain sentences. In that case, epistemic two-dimensionalism will not yield a simpler modal ontology. Suppose that the actual chemical composition of a substance such as water is essential to that substance. Consider a circumstance, *w**, where the watery stuff in *w** is not H_2O but XYZ instead. Now consider the sentence: 'water is XYZ'. According to the epistemic two-dimensionalist, if we interpret 'water' according to its epistemic intension, then this sentence is true *in w**. But, presumably, 'water is XYZ' is also true *about* the circumstance *w**.[17] But water actually just is H_2O, and for the epistemic two-dimensionalist this fact is reflected in the second, external intension of the term 'water'. Given that the actual chemical composition of a substance like water is essential to that substance, or given that water is necessarily identical to H_2O, this means that *any* circumstance about which 'water is XYZ' is true—namely, circumstance *w**—is therefore an impossible circumstance. Thus, it is simply false to say that the antecedent worlds in (2) and (3) only appear to be metaphysically impossible—they are metaphysically impossible. Thus, the epistemic two-dimensionalist is stuck with impossible circumstances.[18]

However, being stuck with impossible circumstances shouldn't be a worry since we can easily incorporate impossible circumstances into our ontology. For example, Nolan (1997) outlines a number of plausible ontological accounts of impossible worlds after arguing for their theoretical use.[19] In Chapter 6 (and more briefly in chapter 3, section 2), I will sketch two easy ways to account for impossible circumstances. One easy way is to be a fictionalist about impossible worlds, which is an attractive view for those

[17] See fn. 17 above for a rough explanation of the difference between external and epistemic intensions.
[18] See Ripley (2012), pp. 105-106 and Yagisawa (2010). pp. 221-227.
[19] Nolan (1997), pp. 541-543.

already committed to fictionalism about possible worlds. The second easy way builds on the work of Schiffer (2003), Thomasson (2007b, 2013), and Steinberg (2013). I will offer the beginnings of a novel ontological account of impossible worlds based on this work that takes them to be pleonastic entities. I will then outline reasons to think that both options can provide a reasonable metaphysics of impossible worlds and a clear epistemological story of how we know claims about impossible worlds. The point, for now, is that the simple fact we can easily incorporate impossible worlds into our ontology undermines the idea that a possible worlds only ontology is a substantial motivation to accept epistemic two-dimensionalism to begin with.

Second, the theory of modality I build my particular account of counterpossibles on, modal normativism, provides a plausible epistemology for metaphysical modality that *need not* rely on a link between conceivability and possibility, which again undermines the epistemic two-dimensionalist's motivation to get rid of impossible worlds. In general, one need not think that the function of modal expressions is to give an exhaustive list of everything that is metaphysically possible. Instead, discourse about what is necessary functions to make explicit those truths that point to certain *inviolability*.[20] At this point, disagreement will come down to what exactly is inviolable, with some proposing it to be some deep feature of reality, with others proposing it to be what is entertainable, and still others proposing it to be the rules that govern the use of our ordinary non-modal vocabulary. On this last approach, which is called *modal normativism*, understanding what is metaphysically possible or metaphysically impossible comes down to conceptual analysis, in particular, it comes down to a *tacit* understanding the rules, or their

[20] See Blackburn (1993), Thomasson (2007b), and Kment (2014).

consequences, that govern the use of our ordinary non-modal vocabulary, perhaps along with other empirical facts.

I will build on modal normativism and argue that understanding the difference between various metaphysical impossibilities also simply comes down to a tacit understanding of the rules that govern the use of our ordinary non-modal vocabulary but when changed in small yet important ways. To preview this idea with a metaphor, consider the game chess. We understand what moves and game states are *possible* in chess by understanding the actual rules of chess (sometimes along with information about the conditions of a game in progress). So, to make sense of many chess *impossibilities*, all we need to do is understand in what distinct ways the chess rules are being violated, either intentionally or unintentionally.

The biggest worry for the epistemic two-dimensionalist approach is that appealing to epistemic descriptions of antecedent worlds will not be sufficient to account for all useful counterpossible circumstances. For example—setting aside the argument that even epistemic two-dimensionalism is committed to impossible worlds—as far as the epistemic two-dimensionalist account has been spelled out so far, it cannot account for (1) and (4) without appealing to impossible worlds. Surely there is no world, no matter how qualitatively similar to the actual world, where Hobbes squares the circle or where there exists a 36-sided steel Platonic solid.[21] So the epistemic two-dimensionalist will fail

[21] See Brogaard and Salerno (2013) and Jago (2014).

to account for all of our intuitions about counterpossibles, and thus fails to meet at least one of the desiderata of an account of counterpossibles.[22]

1.5 Response to Williamson

Williamson (2007, 2016, 2017) raises many objections to abandoning orthodoxy. I do not have space to adequately address each of those objections. However, here I will address what I take to be his main argument against non-vacuous counterpossibles, and I will consider another of Williamson's objections in the next chapter.[23]

Williamson claims that we have strong theoretical reasons to keep orthodoxy that outweigh our intuitions that some counterpossibles are false and not true. Moreover, he offers an error theory to explain away the intuitions that there are non-vacuous counterpossibles, intuitions which have been cited as evidence against orthodoxy.[24] Williamson claims that when evaluating:

(1) If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the same time would have cared

what is prompting the intuition that (1) is false has nothing to do with the *impossibility* of squaring a circle. For surely *no matter what* Hobbes was secretly doing, sick children in South America would not have cared. Instead of relying on the impossibility of the antecedent to guide us, when we consider (1) we note that (1*) seems to be true:

---

[22] Furthermore, if the epistemic two-dimensionalist cannot account for circumstances in which God impossibly commands people to murder, then it will not be able to account for the role counterpossibles can play in debates about what grounds certain moral facts (to be discussed below). In which case, epistemic two-dimensionalism will fail to meet two of the desiderata.
[23] Also see Berto et al. (2017) for a comprehensive response to many of Williamson's objections.
[24] Williamson (2016, 2017).

(1*) If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the same time would *not* have cared.

Williamson claims that, in most cases of counterfactual reasoning, when considering the antecedent of a counterfactual we generally follow a heuristic whereby we should only accept one of two mutually exclusive consequents to follow from our consideration. So, since we accept (1*) as true we feel compelled to reject (1). But Williamson claims that this heuristic is fallible whenever the antecedent is impossible.[25] When considering a counterpossible and we realize that the antecedent is impossible, he claims that we then have independent theoretical reasons to regard that counterpossible as trivially true. So (1*) is true, and our intuition that (1) is false is simply erroneous, though understandable.

Some of the theoretical reasons Williamson cites for treating counterfactuals with necessarily false antecedents as trivially true are in fact accommodated by current accounts of non-vacuous counterpossibles. For example, one reason for wanting all counterpossibles to come out true (as opposed to all false) is that we want *all* counterfactuals of the form '$p \mathrel{\square\!\!\rightarrow} p$' to be true. But this can be accommodated by accounts of counterpossibles that rely on impossible worlds. Suppose that p is impossible. Go to the closest impossible p world and p will be true. Sure, in some cases, it might turn out that p is also false, but in some contexts this will be exactly what we want, e.g. when considering what would be the case if some fundamental particle both had and lacked some property such as charge.[26] Williamson does acknowledge that many accounts of non-vacuous counterpossibles can account for some of the theoretical reasons

[25] Williamson (2016), pp. 8-9.
[26] See Nolan (1997), pp. 554-555, 565.

he cites. But he argues that a substantial drawback to accepting non-vacuous counterpossibles is that non-vacuity conflicts with how we use counterfactuals with impossible antecedents in certain *reductio ad absurdum* proofs.

Williamson has us consider the following *reductio* proof that there is no largest prime:

(R1) If p were the largest prime, $p! + 1$ would be prime.

(R2) If p were the largest prime, $p! + 1$ would be composite.

(R3) Thus, if p were the largest prime, $p! + 1$ would be prime and would be composite.[27]

Since, according to orthodoxy, (R3) is true, and the consequent of (R3) is an *obvious* contradiction, the antecedent *must* be false; thus, there is no largest prime. However, Williamson claims that it is not clear that (R1)-(R3) come out true for accounts of counterfactuals that allow for non-vacuous counterpossibles, and so such accounts potentially count the argument as invalid.

Suppose that we accept premises (R1) and (R2). Williamson's worry is that, on an unorthodox reading of (R3), we are *supposed* to consider the most relevantly similar impossible world to be one where there is a largest prime and it is not the case that $p! + 1$ is both prime and composite. In such a case, the argument is invalid. But such an argument is supposed to be a perfectly good example of valid *reductio* reasoning often used by mathematicians. Thus, a major theoretical reason to reject non-vacuous counterpossibles is that they cannot account for the validity of *reductio* proofs.[28]

---

[27] Williamson (2016), p. 7 and Williamson (2017).
[28] Ibid.

In general, Williamson argues that any non-obvious impossibility can be shown to be obviously impossible through elaborate deductive reasoning from what is non-obviously impossible to what is obviously impossible. So, for the purposes of arguing by *reductio*, we need counterfactuals with impossible antecedents to be true. However, according to Williamson, the unorthodox view is to charitably evaluate counterfactuals with impossible antecedents as false, and so cannot explain the use of counterfactuals with impossible antecedents in *reductio* proofs.[29]

However, this not an accurate representation of accounts that treat *some* counterpossibles as non-vacuous. Most, if not all, proponents of non-vacuous counterpossibles do *not* think that *all* counterpossibles are false, just some. It just isn't true that by simply allowing for the existence of *some* non-vacuous counterpossibles there is pressure for the unorthodox to charitably read conditionals like (R1)-(R3) as false and not true. Proponents of counterpossibles do in fact allow that counterpossibles can often be true and sometimes even vacuously true. For example, Vander Laan (2004) suggests that proponents of non-vacuous counterpossibles can follow Lewis (1986a) in accepting:

> *Rule of Accommodation*: We charitably tend to select the similarity relations that make a counterfactual true, a false-making context will be harder to find.

On both the orthodox and unorthodox accounts of counterfactuals, contextual considerations play an important role in producing an appropriate similarity relation that will determine the relevant respects by which the considered worlds are more or less

[29] Ibid.

similar. For example, when considering what would be the case had Julius Caesar been in command during the Korean War, context will determine whether we consider circumstances with nuclear weapons or circumstances in which Caesar only has the technology that was actually available to him. Following this idea, Vander Laan suggests that there are contexts where the charitable interpretation of a counterpossible such as 'If there were a largest prime *p*, then pigs would fly' is indeed as vacuously true. In such a context, along with the convention of using the phrase 'pigs would fly' to mark absurdities, the counterfactual is treated as true, and so antecedent is taken to be unentertainable.[30]

Williamson acknowledges that the obvious response to his criticism is to appeal to context, and claim that in the context of a mathematical discussion conditionals like (R1)-(R3) will be interpreted as true, not false. However, Williamson (2016) thinks that appealing to context is vague and ad hoc:

> It is not controversial that counterfactuals exhibit some degree of context-dependence in which factors are held fixed; if Julius Caesar had been in command during the Korean War, would he have used catapults or nuclear weapons? But that makes it all the more striking how hard it is to hear the mathematical argument for [R3] as unsound. If context determined whether standard logic and mathematics were to be held fixed, [R3] might be expected to trigger a shift to a context in which they are not, because it brings out so clearly that the antecedent is untenable without such a shift. Yet the argument for [R3] remains compelling. (8)

But, again, this is the wrong way to think about the relationship between counterpossilbes and the context in which they are uttered. It is not the case that the antecedent of a counterfactual solely determines the context, such that the counterfactual, in particular the antecedent, *by itself* determines the appropriate context. Consider the Julius Caesar

[30] Vander Laan 2004, pp. (259-261).

example: the entire point of the example is to illustrate that *on its own* the antecedent supposition is insufficient to tell us what context is relevant for the evaluation of a counterfactual. We need much more information about the context. Likewise, it is not the case that the antecedent of a counterpossible solely determines the context, such that the counterpossible *by itself* triggers a shift to a context where the counterpossible comes out false and not true. There are other important features of a context that determine a similarity relation, which will then determine whether a counterpossible is true or false.[31] For example, salient features of the conversational context—e.g. whether the conversational setting is a philosophical setting in which deviant arithmetic is being discussed or a pedagogical setting such as a number theory course—and speaker intentions all play a role in fixing the relevant similarity relation for evaluating the counterpossible at hand. What's more, and again, according to the rule of accommodation there is in fact pressure to interpret (R3) as true, not false. So, accepting that there are some contexts in which a counterpossible is false does not mean that the context *must* shift in order to accommodate the impossibility of an antecedent and force a counterpossible to come out false.

Moreover, relying on context to account for counterpossibles is not *ad hoc*, especially if we are already inclined to rely on context in our analysis of counterfactuals with *possible* antecedents. For example, we plausibly use counterfactuals to give "*reductio*" proofs ruling out *possible* but contextually unentertainable suppositions. For example, I can easily imagine a context where I *truly* say, "if I were a member of political party X, then pigs would fly." Obviously, the idea is not to find a *possible* world where

[31] See Vander Laan (2004) and Brogaard and Salerno (2013).

the antecedent obtains and see if pigs fly—presumably, there is no such possible world. But given the rule of accommodation, the counterfactual should be heard as true, and given the convention of using the phrase 'pigs would fly' to mark absurdities, I mark the absurdity of my belonging to party X, perhaps to strike it from the conversational grounds, by expressing that the closest world in which I am a member of party X is an impossible world. So contextual considerations are important not just in evaluating counterfactuals with impossible antecedents but also counterfactuals with possible antecedents.[32] So, if we need to appeal to context and impossible worlds to make sense of these counterfactuals, then there is nothing *ad hoc* about appealing to context and impossible worlds to make sense of counterfactuals with impossible antecedents.

To be clear, proponents of non-vacuous counterpossibles agree that the heuristic Williamson discusses is fallible. They can agree that we sometimes mistakenly feel pressure to accept only one of two mutually exclusive consequents to follow from a given supposition, but proponents of non-vacuous counterfacuals need not think that this is our only, or even useful, guiding consideration when evaluating counterfactuals. Proponents of non-vacuous counterpossibles that think there is an intelligible notion of a contextually determined similarity relation used to evaluate counterfactuals with possible antecedents are free to extend that idea to counterfactuals with impossible antecedents. In either case, context can help determine whether a counterfactual is being used to give a *reductio* argument and so counted as true. So, since critics of orthodoxy can account for the

[32] This idea is related to what Nolan (1997) calls the *strangeness of impossibility condition (SIC)*. The SIC claims that every possible world is always more relevantly similar to any possible world than any impossible world. But there are plausible counterexamples to the SIC. I agree that the SIC is likely false, which means that there are true counterpossibles with possible antecedents and impossible consequents. For example, presumably, 'if I were a member of political party X, then pigs would fly' is a plausible counterexample to the SIC. See also Vander Laan (2004), Nolan (2011), and Bernstein (2016).

theoretical use of counterpossibles in *reductio* proofs, they need not feel any theoretical pressure to deny that there are some non-vacuous counterpossibles.

Moreover, I will now argue that there is a lot of philosophical work counterpossibles can do that in fact provide strong reasons in favor of treating them non-vacuously. However, I want to point out that even *if* Williamson is right about the semantics of counterpossibles, then, given the philosophical work counterpossibles can do, something still must be said about the pragmatics of counterpossibles, i.e. about what we are doing when we use counterpossibles or about the acceptance and rejection conditions of counterpossibles. As we will see, this is something that my account of counterpossibles can do independently of the semantic questions.

1.6 Theoretical Work for Counterpossibles

Reasons for accepting non-vacuous counterpossibles go beyond relying on intuitions over examples. There are theoretical considerations that come from considering the role that counterpossibles play in many philosophical debates that also weigh in favor of rejecting orthodoxy. First, counterpossibles are useful for thinking about the non-trivial consequences of philosophical theories that are either necessarily true or false. For example, compatibilists and incompatibilists might disagree about whether it is possible for an agent to possess free will while her actions are fully determined. Many philosophers involved in such disputes take the contested theories to be necessarily true, if true, or necessarily false, if false. Another example can be found in debates about mereological composition. I take *unrestricted mereological composition* roughly to be the view that for any number of objects there is an object that is the mereological sum of

those objects. Many philosophers who debate whether unrestricited mereological composition is true take the thesis to be necessarily true if true, or necessarily false, if false. In these disputes, most philosophers acknowledge that they can make sense of their interlocutor's view and reason non-trivially about what would be the case if their interlocutor were right, per impossible.[33]

In addition, non-trivial counterpossibles potentially play an important role in clarifying the apparent nature of things or relationships of dependency between certain facts. First, consider a counterpossible based on Krakauer's (2012) analysis of metaphysical dependence:

(5) If only one of Socrates or the singleton set {Socrates} were to exist without the other, then it would be Socrates.[34]

Assuming that, necessarily, Socrates exists if and only if the set {Socrates} exists, (5) is a counterpossible. What's more, if (5) is true, it provides evidence that, despite the necessary modal connection between the two, the existence of the singleton set {Socrates} is not essential to the existence of Socrates.[35]

Second, there is reason to think that certain explanations of causal or metaphysical dependence often imply a relation of counterfactual dependence between

---

[33] See Nolan (1997) and Braddon-Mitchell (2009) for more discussion. I will elaborate on this idea in chapter 5.

[34] See Fine (1994), Krakauer (2012), and Brogaard and Salerno (2013). Brogaard and Salerno (2013) argue that we can use counterpossibles to revitalize a modal account of essence, pp. 646-648. On their account, it is false to say that *if nothing had the property of being singleton Socrates, then Socrates would not exist*. However, Steward (2015) notes that the closest world where nothing has the property of being singleton Socrates is the possible world where Socrates doesn't exist, so the relevant counterfactual is actually true. However, Krakauer offers a different account of metaphysical dependence using counterpossibles such as (5) that gets around this type of objection.

[35] Krakauer (2012) argues that (5) is not merely evidence for facts about the nature of Socrates but offers it as a full account of the lack of metaphysical dependence of Socrates on the singleton set Socrates.

the relevant relata. When the relata of these causal or metaphysical dependencies involve metaphysical necessities, then very often the relevant counterfactuals will involve metaphysical impossibilities. For example, Wilson (2016) argues that grounding claims entail nontrivial counterpossible claims. Bernstein (2016) argues that causal claims involving impossibilities have wide application in philosophy.

Here are some examples of counterpossibles that can support or undermine certain philosophical explanations by apparently illustrating a lack of metaphysical dependence between the *explanandum* and the proposed *explanans*. Imagine a nominalist engaged in debate with a Platonist about the existence of numbers making the following claim:

(6) If numbers were not to exist, then the physical world would be exactly as it is, i.e. all of the microphysical particles would still exist and the regularities between them and their manifestations would be exactly the same.

The nominalist might offer (6) as reason to think that physical objects do not metaphysically depend on mathematical objects.[36] However, supposing that numbers necessarily exist, (6) is a non-trivial counterpossible. Or consider an ethicist arguing against certain versions of divine command theory by pointing out that the following seems false:

(7) If God were to command everyone to gratuitously kill and murder, then murder would be morally permissible.

[36] See Liggins (2014).

The ethicist might offer this as evidence that God's commands play no central explanatory role in what makes certain actions right and others wrong, that is, facts about God make no difference to facts about moral evaluation. However, supposing that God exists and is essentially omnibenevolent, the supposition is a metaphysical impossibility and (7) is a non-vacuous counterpossible.

So, non-vacuous counterpossibles can do work in clarifying the consequences of various philosophical views we might consider. They can help us give an account of the apparent nature of things. And, they can help us get clear on relations of metaphysical dependence and metaphysical explanations.[37]

Counterpossibles involve metaphysical impossibilities, so insofar as we need an account of what we do with counterpossibles, we need an account of what we do with metaphysical impossibilities, why we even or would want to think about metaphysical impossibilities at all. One reason, which will be an important reason (if not *the* important reason) throughout this project, is that we can often sensibly compare metaphysical impossibilities.

First, consider the comparison of possibilities. Closely related to the idea that there are ways in which possible worlds can be more (or less) similar to the actual world than others, Lewis (1973) and Kment (2006, 2014) point out that we have considered modal beliefs about comparative possibilities expressed with phrases such as "more

[37] However, I should point out that modal normativism comes from a philosophical background that looks at the project of offering metaphysical explanations or accounts of metaphysical dependence primarily as a conceptual or pragmatic project and not as a project that is looking for descriptive truths about the nature or structure of reality. So, while the normativist may in fact accept many claims such as '*q* in virtue of *p*', '*q* because of *p*', '*q* metaphysically depends on *p*', etc., such claims are not viewed as *descriptive claims* about deep features of reality on a par, say, with descriptive claims made by the empirical sciences. I will have more to say about this in chapter 3, section 4, and in chapter 4.

possible than," "less possible than", "could easily have been", "less easily could have been", etc.[38] For example, I *could more easily* have been a lawyer than I could have been the first person to land on the moon Europa. Lewis (1973) points out:

> Given the notion of comparative overall similarity of worlds, however, there is a natural comparative concept of possibility. It is more possible for a dog to talk than for a stone to talk, since some worlds with talking dogs are more like our world than is any world with talking stones. It is more possible for a stone to talk than for eighteen to be a prime number, however, since stones do talk at some worlds far from ours, but presumably eighteen is not a prime number at any world at all, no matter how remote.[39] (52)

One might take this suggestive of the idea that, in addition to possibilities, we can compare metaphysical impossibilities.

For example, Nolan (1997) notes that worlds where Hobbes squares the circle are far more similar to the actual world than the world where every proposition is true.[40] To put it in comparative terms, the world could more easily be such that Hobbes squares the circles than be such that *everything* obtains. Other examples might be:

(8) The number π could more easily have lacked six consecutive '9's (i.e. the Feynman Point) than it could have been an algebraic number

(9) Smith could more easily have had different parents than she could have been the number 5

(10) Smith could more easily have had different parents than she could have been born in different location than she actually was.

---

[38] See Lewis (1973), p. 52, Kment (2006), p. 253-54.
[39] Given that I am assuming some forms of essentialism for this paper, note that in this quoted passage Lewis is considering the comparison of impossibilities.
[40] Nolan (1997), p. 544.

Prima facie, (8) and (9) are true while (10) seems false. If (8) is true, then this is evidence that there is a distinction between the apparent necessity of the number π having a Feynman point and its being a transcendental number. Perhaps the fact the number π has a Feynman point is a sort of "metaphysical accident"[41] while the fact that π is an algebraic number is not, and this reflects something of theoretical importance. In what way? One explanation is that in most mathematical proofs involving π, the fact that π has six consecutive '9's is hardly, if ever, appealed to. Yet the fact that π is a transcendental number is often appealed to. This is some evidence that the fact π is a transcendental number is explanatorily more important than the fact π has six consecutive nines. Consider (9) and (10). On the one hand, intuitively it seems that less would need to change about the actual world in order for a human to impossibly have a different biological origin than it would need to change in order for a human to be an abstract number (or for an abstract number to be a human). Thus, (9) seems true. On the other hand, less, not more, would need to change about the actual world in order for some person to possibly be born in a different city than would need to change in order for her to impossibly have a different biological origin. This provides some reason to think that comparative claims should not all be treated as trivially true or as all false. Or, at the very least, an explanation needs to be given of our intuitions.

In Chapter 3, I will explain the connection between considerations of comparative similarity and considerations of comparative possibilities and impossibilities and how this connection relates to the evaluation of counterfactuals more generally. But it is easy to see that there a connection between these ideas by considering the following example. If

[41] Thanks to Eli Chudnoff for raising the Feynman point as an example of a metaphysical accident.

it's true that Smith could more easily have had different parents than she could have been the number 5, then there is an impossible world in which Smith has different parents but where she is not the number 5 that is more similar to the actual world than worlds in which Smith is the number 5. Therefore, it is not the case that if Smith were to have different parents, she would be the number 5. So, insofar as we can sensibly speak of something being more impossible than something else, we need to make sense of metaphysical similarity between worlds both possible and impossible.

So, we need an account of counterpossibles that can allow for our semantic intuitions about some conditionals involving impossibilities, that can do a lot of interesting philosophical work, e.g. in making sense of certain claims of metaphysical dependence. I would like to close this chapter discussing another theoretical reason in favor of providing an account of non-vacuous counterpossibles: new work for a theory of counterpossibles that comes from the literature on conceptual engineering. Recent literature on conceptual engineering has emphasized the way conceptual engineering can be used to make sense of intractable disagreements found in many metaphysical disputes—many metaphysical disputes might turn out to be more about conceptual matters than matters about the nature of things and many metaphysical disputes might in fact be normative disputes about whether and how we should use certain terms or concepts.[42] In chapter 5, I will argue that counterpossibles allow us to advocate for or against certain dependencies among our metaphysical terms or concepts while remaining in the object language. Furthermore, counterpossibles enable us to work out or negotiate the consequences of certain conceptual changes or practices without actually adopting

[42] See Burgess and Plunkett (2013), Plunkett (2016), and Thomasson (2017).

those changes or practices. Overall, instead of thinking about various conceptual practices in the metalanguage, counterpossibles allow us to think about and reason about alternative conceptual practices while remaining in the object language.

I will argue that this is related to the thought that just as there might be internal and external existence questions, there might be *internal* and *external modal questions*. Let *internal* modal claims be modal claims made against the backdrop of our actual linguistic practices, perhaps along with empirical information. Our internal modal claims capture our apparent ordinary modal beliefs, e.g. water is necessarily H_2O, 'nothing can be red all over and green all over', 'agent causation is possible', etc. Many internal modal questions are easily answered by investigating what our actual linguistic practices are. On the other hand, external modal questions go along the lines of "Is agent causation *really* impossible?" Substantive realists about certain modal questions hope to answer external modal questions by appealing to substantive facts about either modal properties, or the nature of things, or grounding relations, or relations of fundamentality, or the metaphysical laws that structure reality.

Alternatively, as I will explain, my normativist account of counterpossibles and metaphysical similarity interprets many external modal questions as practical or normative questions about which conceptual or linguistic practices are best to adopt given other theoretical and practical goals. For example, suppose that we ordinarily take theorems of classical logic to be a conceptual default, and from that point we talk about possibility or impossibility with respect to that default position. But external to that choice it would not be any more or less correct to start from an alternative logic, though there may be pragmatic reasons to avoid this. Regardless of our starting point, once a

starting point is selected, counterpossibles play a very important role in discourse: they allow us to negotiate or hypothetically weigh the consequences of adopting alternative modal practices or defaults, e.g. adopting other logics, from *within* our default position, i.e. internally, using counterpossible conditionals.[43]

1.7 Summary

In this chapter I introduced the philosophical debate over counterpossible conditionals and provided an outline of my overall project. Then I accomplished the first step of my overall argument, which was to establish: that counterpossibles are sometimes *non-vacuous*, e.g. there are some false counterpossibles; and, more importantly, that thinking about counterpossibles and metaphysical similarity between impossibilities is a useful theoretical tool for getting clear on many philosophical questions.

I overviewed the distinction between orthodox and unorthodox views of counterpossibles. Then I argued against the view that we can keep orthodoxy and account for non-vacuous counterpossibles by appealing to epistemic two-dimensionalism. I also argued against Williamson's error theory of counterpossibles, which he tried to support by claiming that there are theoretical considerations that support adopting his error theory over rejecting orthodoxy. I then argued that there are in fact stronger theoretical reasons in favor of rejecting orthodoxy: the use of counterpossibles in making sense of philosophical disputes; their use in understanding the apparent nature of things; their use in accounting for metaphysical dependence; their using in making sense of intuitions

[43] See also Nolan (1997) for an interesting discussion of how counterpossibles can help us understand alternative logical systems while keeping a classical logical consequence relation to govern our base language, pp. 545-547, 554.

regarding comparative impossibilities; and, finally, I outlined new work for a theory of counterpossibles that comes from the literature on conceptual engineering.

So, I have established the need to find an account of counterpossibles. Again, I do not mean an account of what *makes* counterpossibles true or false. Ultimately, given the pragmatist approach I will end up taking, it would be better to say we need an account of what we are doing with counterpossibles and when we properly do it. Perhaps a clearer way of putting this point is that we need to establish the non-vacuous assertibility conditions for counterpossibles. However, I am also fine with talking about establishing non-vacuous truth conditions for counterpossibles, so long as this is not taken to mean 'establishing what it is that *makes* counterpossibles true or false'. This will be become clearer in Chapter 3, especially section 2.

CHAPTER 2: EPISTEMIC READINGS AND SUBSTANTIVE REALISM

In the last chapter I established that we need to take the unorthodox view of counterpossibles seriously because the main arguments against them are unconvincing and, more importantly, there is a lot of interesting theoretical work for counterpossibles to do. So, we need an account of non-vacuous counterpossibles—now what? In this chapter I will look at three approaches to non-vacuous counterpossibles: a subjective epistemic approach, a strict epistemic approach, and a substantive realist approach. I will argue against both epistemic readings of non-vacuous counterpossibles. However, I will argue that trying to replace the epistemic readings with a substantive realist account that appeals to *substantive* metaphysical natures or *substantive* relations of metaphysical dependence or fundamentality comes with metaphysical and epistemological worries that would be best to avoid.

I take a "*substantive features of X*" to mean features of *X* that are posited to explain our *X* thought and *X* talk or explain what makes *X* claims true. Substantive features of *X* are metaphysically on a par with empirically tractable features of the world described by the physical sciences, yet we apparently cannot answer questions about *X* using either direct empirical methods or conceptual analysis.[44] The worry, roughly, is that substantive features of *X* then seem to be outside of our cognitive reach, which means that we must either be skeptics about *X* or we must accept that which of our beliefs about X are true is a matter of sheer luck, either way our theoretical reasons for talking about X to begin with will be undermined.

[44] See Sider (2011), p. 187.

So, after raising these worries, the need for an account of non-vacuous counterpossibles and metaphysical similarity will remain at the end of this chapter. I will go on to present a novel account of metaphysical similarity and counterpossibles in chapter 3.

2.1 Introducing Three Ways to Think About Similarity

The analysis of counterpossibles I will offer in chapter 3 tries to retain some of the spirit of Nolan (1997): consider an impossible world where the antecedent obtains that is most relevantly *similar* to the actual world and see if the consequent obtains as well. But, how should we think about the comparative similarity of worlds in general, let alone similarity between two impossible worlds? When evaluating any conditional in terms of similarity the context can make many different dimensions of similarity relevant to the evaluation. In fact, we have a seemingly unlimited number of options. For example, we can be fairly specific and evaluate similarity with respect to my personal biography, or with respect to the biography of Winston Churchill, etc. Or we can evaluate similarity along seemingly less natural dimensions such as similarity with respect to my personal biography *or* the biography of Winston Churchill, or we can evaluate similarity with respect to what happened on Tuesdays between the years 1645 AE and 2065 AE. More generally and more naturally, we can evaluate similarity with respect to exact matches with respect to spatiotemporal regions of the actual world, or we can evaluate similarity with respect to more general facts such as the laws of nature, or similarity with respect to the logical and mathematical truths. So, how should we think about similarity when considering a counterfactual let alone a counterpossible?

Lewis (1986b) offers some advice when responding to an potential objection to his similarity-based account of counterfactuals. Consider:

(11) If Nixon had pressed the button, there would have been a nuclear holocaust.

Intuitively, (11) is true. On the orthodox (or unorthodox) account of counterpossibles, which builds on Lewis's work, this means that there is a world where Nixon presses the button and there is a nuclear catastrophe that is more relevantly similar to the actual world than all worlds in which Nixon presses the button but there is no nuclear catastrophe. But Fine (1975) tries to raise (11) as an objection to Lewis's account of counterfactuals.[45] The worry is that worlds in which Nixon presses the button but nothing happens—perhaps because the signal fizzle out in the wires—and life goes on as pleasantly as always (e.g. a world in which I am typing away just as I actual am) are actually *more similar* to the actual world than a world where Nixon presses the button and there is a nuclear holocaust (e.g. a world in which, instead of typing as I actually am, I was likely never even born). Hence, if Fine is right, according to Lewis's account of counterfactuals, (11) is false, which is the wrong result.

Lewis responds by noting that counterfactuals and counterfactual judgments are extremely shifty and argues that context and considered modal beliefs do most of the work in determining the similarity relation used to evaluate a given counterfactual. He then goes on to provide a similarity relation that utilizes intuitively relevant dimensions of similarity, which are weighted, and which provide the correct verdict for (11):

[45] Fine (1975), discussed in Lewis (1986b).

(LSR) *Lewis's Similarity Relation for (11)*:

1. It is of the first importance to avoid big, widespread, diverse violations of law.
2. It is of the second importance to maximize the spatiotemporal region throughout which *perfect* match of particular fact prevails.
3. It is of third importance to avoid even small, localized, simple violations of the law.
4. It is of little or no importance to secure *approximate similarity* of particular fact, even in matters that concern us greatly.

The rough idea is that worlds in which Nixon presses the button but the signal fizzles out, *and* Nixon forgets what he did, *and* life continues more or less as it actually did, etc. require widespread, diverse violations of the physical laws. That is, in order for (11) to come out true, the universe would need to seriously collude against itself to erase all physical traces of Nixon's actions such as Nixon's memories of his action the next day, Nixon's mental reactions to what he did, traces of his fingerprints, any recording of the event, the signal fizzling out, etc., all of which are predetermined by the physical laws up to the event of Nixon pushing the button. Furthermore, the widespread, diverse violations of the physical laws would only result in *approximate* similarity to the actual world after the time Nixon pushed the button. Thus, LSR rules out such worlds.

LSR instead favors worlds in which only a tiny physical miracle takes place that causes some extra neurons to fire in Nixon's brain, which in turn causes him to push the button but after which there are no more miracles and things follow according to the actual physical laws resulting in a nuclear holocaust. Hence, contra Fine 1975, (11) is

true on Lewis's account.[46] However, it is important to note that Lewis is not providing LSR as an *absolute* metric for similarity once and for all.[47] Lewis ultimately thinks that similarity is determined on a case by case basis, and that, in addition to contextual factors (e.g. the rule of accommodation), we often utilize our considered modal beliefs about counterfactuals to determine the appropriate similarity relation and not the other way around.[48]

So, when determining how similar *possible* worlds are to the actual world, we can often (though perhaps not always) follow Lewis's suggestion to maximize adherence the actual physical laws while minimizing miracles and lastly consider match with particular matters of fact. However, there is good reason to think that this criterion will not always work for determining how similar *impossible* worlds are to the actual world. Imagine one impossible world in which Hobbes squares a circle, and imagine a different impossible world where, while in the process of squaring a circle, Hobbes also discovers a largest prime, but otherwise the two worlds are similar to the actual world with respect to the

---

[46] Lewis (1986b), pp. 44-46.

[47] I think the fact that Lewis devised this similarity relation in order to analyze a particular counterfactual presented as a counterexample to his view of counterfactuals is a point often lost in literature that tries to expand on Lewis's idea for metaphysical counterpossibles, e.g. Kment (2014), Krakauer (2012), Brogaard and Salerno (2013), etc. This matters since the usefulness of a "default" similarity relation diminishes if counterexamples to the SIC are more common than we think. Since I am inclined to think that counterexamples to the SIC are relatively easy to come by, I am skeptical that such a default relation has much theoretical or practical import. I follow Lewis in thinking that most counterfactual discourse is extremely sensitive to context.

[48] See Lewis (1986), p. 43. In a way, this might seem backwards, especially if you think that any analysis of counterfactuals is supposed to explain why some counterfactuals are true and others not, and you hope to provide some sort of decision procedure to help determine the truth value of counterfactuals. Of course, in general, when giving a philosophical analysis of X, it seems completely fair to start with our considered beliefs and intuitions about X, develop a theory of X, and then compare the two making adjustments where needed, which means sometimes even adjusting or explaining away our considered beliefs or intuitions about X. But Lewis seems to work our considered modal beliefs into his theory of counterfactuals in an important way by utilizing them to determine appropriate similarity relations. However, one could argue that if we are to take seriously the context sensitivity of similarity along with the rule of accommodation discussed above, the incorporation of our considered modal beliefs and intuitions into a theory of counterfactuals is a desirable aspect to Lewis's theory, not an undesirable one.

physical laws and other matters of fact. While we might be able to state some differences between these impossible worlds in terms of the comparative mismatch of the spatiotemporal regions Hobbes is working on his proofs, e.g. via the length of Hobbes's proofs and different brain states he experiences, it seems that we are after a more important difference between these two worlds. Namely, we are after something more to the effect that in the one impossible world at least one metaphysical necessity is violated and in the other impossible world at least two metaphysical necessities are violated.[49]

So, when determining similarity for impossible worlds, we need some additional criterion of metaphysical similarity between worlds. A good place to start is figuring out what is supposed to be distinctively metaphysical about the similarity between two worlds. On a more austere account, metaphysical similarity amounts to no more than similarity with respect to the physical features of the actual world and the laws of nature, perhaps along with facts about mathematics and set theory. On a much more permissive account, metaphysical similarity amounts to similarity between mathematical matters, logical matters, mereological matters, kind matters, essential matters, and, perhaps for some, many more matters such as grounding matters and fundamental matters. The account of counterfactuals and counterpossibles I will offer in the next chapter has this more permissive account in mind. Leaning more towards this more permissive account, there are three natural ways we might go when accounting for metaphysical similarity: by appealing to some sort of *epistemic criterion*, a *substantive metaphysical criterion*, or by a *conceptual criterion*.

[49] Assuming that the physical facts do not depend on the mathematical facts in any essential way.

In the remaining sections of this chapter and the next chapter, I will talk about these three options in some detail. However, before moving on, it will be useful to roughly sketch how these criteria work when evaluating counterpossibles so that we have a broad understanding of all of the criteria while considering them individually in more detail. Consider the following example of a seemingly false counterpossible:

(11) If water had not been H_2O, then water would have been a monkey.

Intuitively, impossible worlds in which water is not H_2O and water is *not* a monkey are more metaphysically similar to the actual world than impossible worlds in which water is not H_2O and water *is* a monkey.[50] Thus, (11) is false. So, by what criterion should we make this assessment?

Brogaard and Salerno (2013) suggest that when a given context requires evaluating counterpossibles we might need to maximize what they call *a priori* connections* between non-modal expressions, where these connections are given in terms of *a priori* implication relations between non-modal expressions *relevant to a subject in a given context*.[51] Since, for the right subject and in the right context, there will be no *a priori** connection between the sentences 'water is not H2O' and 'water is a monkey', (11) comes out false.[52]

Kment (2006, 2014) also thinks that there are non-vacuous counterpossibles.[53] In the development of his theory of metaphysical necessity, Kment introduces the idea of using *metaphysical laws* to rank similarity between worlds.[54] The metaphysical laws are

[50] Brogaard and Salerno (2013), p. 653.
[51] Ibid., p. 655.
[52] Ibid., pp. 654-656.
[53] Kment (2006), pp. 241, 248. See also Kment (2014), p. 62 among others.
[54] Also see Krakauer (2012) for an account of similarity based on metaphysical laws.

facts about the "deep features of the world order" and include facts about essence, metaphysical grounding, metaphysical structure.[55] Following Kment's ideas, when evaluating counterpossibles we might first maximize match to the actual world with respect to the metaphysical laws while minimizing "metaphysical miracles" and then maximize match in background facts. On this approach, in the closest impossible world in which water is not H_2O, water is not a monkey, since it seems that one natural kind, e.g. water, changing into another more complex natural kind, e.g. a social and conscious living creature composed of many substances, would involve massive and widespread metaphysical miracles compared with water simply having a different chemical composition while retaining most of its other qualities. Thus, (11) is false.

However, there is a third way of accounting for non-vacuous counterpossibles that appeals to a conceptual criterion. Moreover, as I will argue in the next chapter, this approach to non-vacuous counterpossibles helps us avoid some of the worries that I will raise for the epistemic and metaphysical approaches.[56] Roughly, for all counterfactual conditionals, i.e. those with possible or impossible antecedents, and given a particular context, what matters to determining metaphysical similarity is neither the epistemic perspective of a subject nor substantive notions of essence, grounding or metaphysical structure but rather similarity with respect to what I call *conceptual laws*. Very roughly, considering what would be the case if water were to have to have a different chemical structure does not license the application of the term 'water' to a monkey in that hypothetical, yet impossible, circumstance. However, I am not proposing that

[55] Kment (2006) Section 5 and Kment (2014) pp. 161-173. See also Schaffer (2010) and Rosen (2006).
[56] Actually, as I will explain, on my view the conceptual criterion just is a metaphysical criterion, just not a metaphysically robust criterion.

counterfactuals are about our concepts or that we reason counterfactually by explicitly considering how our language works, so care needs to be taken. I will return to, and clarify, this idea in the next chapter.

With a rough outline of these three approaches in mind, I will now start the task of critically evaluating the first two. In the remainder of this chapter, my aim is to raise sufficient worries for the epistemic and metaphysical approaches to counterpossibles to open up the argumentative space for my third option, which is a conceptual account of metaphysical similarity.

2.2 Worries for Subjective Epistemic Readings

Brogaard and Salerno (2013) offer a partially epistemic account of closeness (similarity):

(ESR) Epistemic Similarity Relation:

1. Minimize discrepancies with background facts—the laws of nature, particular matters of fact, etc.—relative to context.
2. If needed, maximize *a priori** connections, where for a speaker s in a context c, *P a priori** implies *Q* iff *Q* is a relevant *a priori* consequence of *P* for s in c.[57]

They also offer a less informal version:

> For any two impossible worlds w_1 and w_2, w_1 is closer to the base world than w_2 iff
>
> a) w_1 does not contain a greater number of sentences formally inconsistent with the relevant background facts (held fixed in the context) than w_2

[57] Brogaard and Salerno (2013), pp. 654-655.

does. And if w_1 and w_2 contain the same number of sentences formally inconsistent with the relevant background facts (held fixed in the context);

b) w_1 preserves a greater number of *a priori** implications between sentences than w_2 does. (Brogaard and Salerno, 655)

Again, *a priori** connections are given in terms of *a priori* implication relations between non-modal expressions *relevant to a given subject in a given context*. Presumably, ESR is meant to rank similarity in all contexts, though it looks as if there can be quite a bit of variance in similarity considerations depending on what *a priori* connections are relevant for the speaker s in a given context. Aside from the fact that Brogaard and Salerno call their account an 'partially epistemic' account, I take this account to be epistemic in that it is relativized to the *a priori* connections relevant for a speaker. Overall, this approach seems to be *a generalized version of epistemic two-dimensionalism*. On the standard epistemic two-dimensionalist account, '*p*' *a priori* implies '*q*' just in case for some subject '$p \wedge \sim q$' is ruled out the basis of ideal *a priori* reasoning. The generalized account offered by Brogaard and Salerno, seems to allow for *non-ideal a priori reasoning*.[58]

However, ESR is not entirely clear on some important questions. For example, when evaluating a counterpossible, can the person evaluating the counterpossible, or can the speaker in the context, ever appeal to her considered modal beliefs to help pick out a similarity relation such that she is able to have some say in which *a priori* connections are relevant? Or do they mean that what *a priori* connections are relevant will always only depend on what *a priori* connections the speaker in a context *recognizes*? The answer matters, because it will determine whether their account can block Williamson's

[58] Ibid.

objection that treating counterpossibles non-vacuously *forces* us to treat seemingly valid *reductio* arguments as invalid.

Recall that to block Williamson's objection, we want it to be the case that in *some* contexts counterpossibles such as 'if *p* were the largest prime, *p*! + 1 would be prime and would be composite' actually do come out true and not false.[59] On the one hand, on the latter subjective reading of Brogaard and Salerno's account, similarity depends on what *a priori* connections a speaker in a context recognizes. In that case, there is some pressure to think that many non-ideal speakers will not recognize an *a priori* connection between '*p* is the largest prime' and '*p*! + 1 is both prime and composite', in which case 'if *p* were the largest prime, *p*! + 1 would be prime and would be composite' comes out false. This means that the valid argument that there is no largest prime comes out invalid, and Williamson's objection stands.[60]

On the other hand, the former reading of Brogaard and Salerno's account potentially allows us to block Williamson's argument because it allows for a speaker to recognize many *a priori* connections but have some leeway in deciding which background facts or *a priori* connections are relevant. In this case, a speaker (or whoever is evaluating the counterpossible) can let her considered modal beliefs guide the selection of a similarity relation that allows (R3) to come out true and not false. But I think that this makes the account less "epistemic" since it not tied to what the speaker recognizes but to an active selection of *a priori* connections.

---

[59] Recall that I argued Williamson is incorrect in assuming that when confronted with (R3) the context shifts automatically so that we charitably read (R3) as false.

[60] I do not think that this is a conclusive objection to their view. This may not be a problem if there is a context in which a speaker does recognize the a priori connection. And maybe this needs to be done on a case by case basis.

My suspicion is that Brogaard and Salerno want to go with a subjective epistemic reading of *a priori** implications because their account seems to be largely motivated by the idea that non-vacuous counterpossibles can explain non-ideal epistemic perspectives. That is, in order to account for circumstances that are epistemically possible for a given speaker but not metaphysically possible, it is better to focus on the *a priori* connections between sentences a speaker *recognizes*. And since a speaker in a given context may fail to recognize *all* of the *a priori* connections between the sentences being considered, we can explain why a speaker might recognize some necessary truths of mathematics or logic but not others, or, e.g., the speaker might entertain the impossibility that Hesperus is not Phosphorus without thereby being committed to everything. Consider:

(12) If Hesperus were not Phosphorus, then Hesperus would not be Phosphorus.

(12*) If Hesperus were not Phosphorus, then Hesperus would not be Hesperus.

The idea is that there is a context in which (12) is true and (12*) is false even though Hesperus and Phosphorus are coreferential proper names. This means that counterpossibles are sometimes referentially opaque. On the subjective reading of *a priori**, the referential opacity demonstrated by (12) and (12*) stems from the fact that a speaker might not recognize the relevant ideal *a priori* connections between 'Hesperus' and 'Phosphorus'.[61]

However, Williamson (2007, 2017) argues that by allowing for a non-vacuous impossibility in which Hesperus is not Phosphorus we run the risk of making the following argument invalid when it should be valid:

---

[61] Of course, there is an ordinary sense of *a priori* in which 'Hesperus = Phosphorus' is not *a priori*. But I think the idea is that *given* certain background empirical facts, there is a weaker sense in which 'Hesperus = Phosphorus' is *a priori* (say for an ideal agent). See Chalmers (2002).

Rocket Argument:

1. If the rocket had continued on that course, it would have hit Hesperus.
2. Hesperus = Phosphorus
3. If the rocket had continued on that course, it would have hit Phosphorus.

Brogaard and Salerno (2013) respond:

> We agree with Williamson that this inference is valid. Moreover, our account does not require us to say otherwise. Yes, the counterfactual conditional is hyperintensional.[62] However, we need not throw out the baby with the logically ill-behaved bath water. Our logical principles may be restricted accordingly. All the typical rules governing counterfactuals are valid, when the antecedent is possible. The above argument, by contrast, does not involve an impossible antecedent.[63] (657)

So, on their view, it is the impossibility of the antecedent in (12) and (12*) that triggers considerations of what *a priori* connections are recognized by a speaker in a given context. However, since there are no impossibilities involved in the *Rocket Argument*, the validity of the argument is respected.

However, Williamson (2017) takes this to be a problem with the account. He argues that Brogaard and Salerno utilize objective, non-epistemic standards for

---

[62] A position in a sentence is *hyperintensional* just in case it does not permit the substitution of intensionally equivalent expressions. A position in a sentence is referentially opaque when it does not permit the substitution of proper names. It need not be the case that all hyperintensional positions are referentially opaque. Williamson (2017) gives an example of a sentential operator # such that 'a # b' is true whenever 'a' and 'b' express the same singular proposition. (Where a *singular proposition* is a proposition that concerns only the objects, properties, and relations that proposition is about and not their representations.) This operator is hyperintensional but not referentially opaque. For example, let a = water is H_2O, b = water is water, and c = H_2O is H_2O. Suppose that all three express the same singular proposition and all three are metaphysically necessary. Then 'a # b' will be true for a, b, and c. However, 'a # 2 = 2' will not be true. Hence, '#' is hyperintensional but not referentially opaque.

[63] One way to read this is that it is the impossibility of the antecedent that somehow *makes* the context referentially opaque. But I disagree. I will argue that it is not the modal status that really matters but the context in which a counterpossible is evaluated and which determines what criteria of similarity are needed to make sense of the counterpossible.

evaluating similarity of possible worlds and only bring in epistemic standards for impossible worlds:

> Thus, their overall relative [similarity] relation is patched together from epistemic and non-epistemic pieces. It is hard to avoid the impression that the account is being gerrymandered just to accommodate the marginal case of counterpossibles. Such a hybrid approach resembles an account of conditional probabilities on which they are purely objective when the conditioning event has a positive unconditional chance, but go epistemic or subjective when the conditioning event has zero unconditional chance: not an attractive option. (Williamson 2017, 16)

I take Williamson's objection to be that there is something *ad hoc* about adopting a hybrid account of counterfactuals merely to account for the referentially opaque contexts that allow for non-vacuous counterfactuals with impossible antecedents, which he considers to be a marginal problem. Williamson seems to take this as an objection to non-vacuous counterpossibles more generally. But I think that Williamson's objection has force: a.) only insofar as we are forced to determine similarity between impossible worlds on the basis of what a subject recognizes in a context; and b.) we make criteria for similarity solely dependent on the modal status of the antecedent so that an impossible antecedent forces us to appeal to what a subject recognizes in a context. Let me explain why I think we should reject these ideas.

On the one hand, we can think of *subjective epistemic similarity* in terms of the similarity between certain features of various epistemic states, e.g., the similarities between different bodies of evidence, or the similarities between various subjection recognize *a priori*. On the other hand, we can think of *circumstantial similarity* in terms of the similarity between the objective features of various states, e.g., the similarity of different states with respect to physical properties such as shape or mass, or the similarity of different states with respect the physical laws more generally. However, we can very

clearly set aside subjective epistemic considerations of similarity and think about the circumstantial similarity among possibilities and impossibilities. For example, an actual steel triangle is plausibly more circumstantially similar to an impossible steel Penrose triangle than it is to a possible cloud of neon gas. So, we are not forced to determine similarity between impossible worlds on the basis of what a subject recognizes in a context, we have other options.

What's more, I will argue in the next section that what sense of similarity best helps us evaluate a given counterfactual will depend on broader features of the context and not just the modal status of the antecedent. To return to Williamson's comment, we can likewise say that whether it is best to interpret the sentence 'It will probably rain today' as expressing something about credences or about circumstantial probabilities depends on broader contextual considerations and not just whether rain today has a positive unconditional chance of happening.

Let me begin to argue for these claims by pointing out an alternative way to think about the *Rocket Argument*. Again, on one way of reading Brogaard and Salerno's account, impossibilities trigger a shift to subjective epistemic considerations. In addition, on their account, whether we are permitted substitution *salva veritate* in a counterfactual depends on the modal status of the antecedent—whenever the antecedent of a counterfactual is possible, we are so permitted, otherwise we are not. But to see why an explanation in terms of the modal status of the antecedent isn't quite right, consider this model involving two worlds:

w_1	w_2
The rocket continued on that course.	The rocket continued on that course.
The rocket hit Hesperus.	The rocket hit Hesperus.
The rocket hit Phosphorus.	The rocket did not hit Phosphorus.

If the *Rocket Argument* is to be valid, then w_2 needs to be excluded so that the conclusion is true. According to Brogaard and Salerno, since the antecedent of the conclusion is possible, we only consider similarity with respect to possible worlds, so w_2 is excluded. But it is easy to see that the modal status of the antecedent is not really what is doing the work in excluding w_2. What is really doing the work in this case is the fact that 'Hesperus = Phosphorus' is a premise of the argument, which makes that particular empirical fact stunningly relevant when considering the truth of the conclusion. And that is why w_2 is excluded from consideration, not the fact that the antecedent is possible.

One might object to my response by pushing harder with the following model:

w_1	w_2
The rocket continued on that course.	The rocket continued on that course.
Hesperus = Phosphorus	Hesperus = Phosphorus
The rocket hit Hesperus.	The rocket hit Hesperus.
The rocket hit Phosphorus.	The rocket did not hit Phosphorus.

Again, in order for the argument to be valid, as it should be, we need to exclude w_2. But, again, why say that it is the modal status of the antecedent that is doing the work in excluding w_2? It seems that we can just say that some background fact regarding identity—e.g., a general rule to the effect that *whenever x = y, for all R, Rx = Ry*—is

being held fixed by the context such that worlds that obey that rule are more relevantly similar in the context than worlds that violate that principle. Again, w2 is excluded, and this explanation did not appeal to the modal status of the antecedent.[64]

Let me take a moment to clarify my points. First, I am not claiming that counterpossibles are never referentially opaque—there are contexts in which they are. What I am arguing is that whether we should read a counterpossible as being referentially opaque is not necessarily a result of the modal status of the antecedent but a result of other contextual considerations, such as we are dealing with a non-ideal epistemic perspective. Second, and importantly, I want to stress that I am not claiming that what a speaker recognizes in a context is never relevant to evaluating counterpossibles. I agree that counterpossibles and impossible worlds are very useful for clarifying epistemic modalities. Sometimes considerations of subjective epistemic similarity are best. All I am claiming is that we should not think that we must *always* think about similarity in terms of subjective epistemic features. Often the relevant sense of similarity at play in counterpossibles is not relative to any particular subjective perspective.

We sometimes need to talk of modal phenomena in a way that is not tied to the evidential or cognitive perspective of any particular agent. For example, it seems that even two *omniscient* philosophers could have substantive debates over what would or would not be the case if some impossibility were to obtain, e.g. two omniscient classical

[64] One response might be that when all of the premises of an argument are loaded into the antecedent of a counterfactual, then so long as the front-loaded antecedent is possible, the standard similarity relations apply. This would be to read the antecedent of a counterfactual as an ellipsis for everything else being assumed by the context. This is one option. But I tend to agree with Lewis (1973)—why not just take the antecedent as stated to be the actual antecedent and let context determine what additional premises are relevant to making the inference at hand.

logicians debating about alterative logics. In that case, it would not be the subjective perspective of the agents that mattered to similarity—they are, after all, omniscient. What would matter would be the similarity relation made most relevant by the conversational intentions of the speakers and other features of the discussion, such as how the alternative logics are in fact stated and what arguments count as valid or invalid according to those logics. So, if we want to allow for all kinds of counterpossible reasoning, it is better to have available a broader notion of similarity that can allow for more than *a priori* connections between sentences recognized by a subject.

2.3 Worries for Strict Epistemic Readings

However, Vetter (2016) has recently argued that counterpossibles are *best* read *epistemically*. What this means is that counterpossibles are *always* concerned with evidence, appearances, or representations.[65] She starts by noting that counterfactuals in general are sometimes epistemic insofar as they are concerned with evidence, appearances, or representations, and they are sometimes *circumstantial* insofar as they are concerned with the nonrepresentational world. One might object that clearly counterfactuals in general are never epistemic because they are in the *subjunctive mood* (if *p were* to be the case, then *q would* be the case) and not the *indicative mood* (if p is *in fact* the case, then q is *in fact* the case). But Vetter, drawing on the work of Edgington 2008, gives a number of examples of counterfactuals in the subjunctive mood that are best read as epistemic and not as circumstantial. For example,

[65] I am reading 'concerned' as 'being directly or indirectly about'. As we will see, there is a sense in which I think that linguistic features of our non-modal vocabulary matter to the evaluation of couterpossibles. But I do not think that counterpossibles are either directly or indirectly about anything linguistic.

> … 'Why did you hold Smith for questioning?' 'Because we knew the crime was committed by either Jones or Smith—if it hadn't been Jones, it would have been Smith.' (Vetter 2016, 2695)

Vetter argues that the counterfactual about Jones is best given an epistemic reading. The counterfactual is not claiming that at the closest possible world where Jones does not commit the crime Smith was waiting as a backup—after all, Smith might actually be innocent and have no inclination to criminal activities. So, read circumstantially, the conditional is false. However, given the context of the discussion, the conditional is clearly not false. Thus, the best explanation is that utterer of the conditional is making a claim regarding her available evidence.

To determine whether it is best to read a counterfactual as epistemic, Vetter takes a counterfactual giving rise to referential opacity—where a position in a sentence is *referentially opaque* just in case it does not always permit the substitution of proper names *salve veritate*—to be "conclusive evidence for the epistemic reading."[66] Vetter argues:

> The reason is simply that circumstantial modality concerns the objects, properties, and relations that a given modal claim is about, not any representational or cognitive features of the terms we use to refer to them. A failure of substitutivity for names is generally an indication that what matters are certain representational features of the names, rather than the objects to which they refer. (Ibid.)

Vetter goes on to argue that *all* non-vacuous counterpossibles are best read epistemically and not circumstantially. First, Vetter claims that "counterpossibles that are non-vacuously true or false must indeed be referentially opaque." I will show next why this is false. Second, Vetter is arguing that, since referential opacity is supposed to indicate representational or cognitive features of terms, what Vetter broadly calls 'epistemic,'

[66] Ibid., p. 2698.

counterpossibles that are non-vacuous are always epistemic.[67] But, if what Vetter means is that counterpossibles are always epistemic in that they are *about* representational or cognitive features of the terms or that we must *always* appeal to epistemic criteria for their evaluation, it isn't clear that this is true.

First, there are non-vacuous counterpossibles that have readings that are not referentially opaque. Moreover, they do not seem to be *about* the representational or cognitive features of terms. Consider the following counterpossibles:

(13) If a steel Penrose triangle were placed in a 4000 deg. F oven, it would melt.

(14) If George Eliot had been Sean Spicer's biological parent, then Sean Spicer would be a much more eloquent speaker.

First, it isn't clear to me that (13) or (14) must be referentially opaque in every context. Consider (13), for example. It isn't clear to me what any substantial connection there is between the impossibility of the antecedent and referential opacity. Suppose that 'steel Penrose triangle' is the proper name of some impossible kind and let's stipulate a name 'Oscar's tribar' such that Oscar's tribar = steel Penrose triangle. Plausibly, there is a context where this stipulation is held fixed (i.e., this one), and where if I go on to say that *if an Oscar's tribar were placed in a 4000 deg. F oven, it would melt*, I still say something true. Nothing funny is going on, despite the impossible antecedent. So, I see no reason to say that (13) is anymore referentially opaque than counterfactuals about possible tribars, so long as we hold fixed relevant information.

[67] Vetter (2016), pp. 2697-2698. I will begin the next section by looking at a general argument against Vetter' claims, which appeals to purportedly circumstantial relations such as grounding, which arguably have referentially opaque readings.

Furthermore, the (13) doesn't seem to be about how we represent impossible things but is about what would happen to an impossible thing no matter what you call that thing. But, according to Vetter's proposed connection between opacity and non-vacuous counterpossibles, if (13) isn't referentially opaque, then (13) is not a non-vacuous counterpossible. But that seems to be the wrong result. Note that there are many contexts where it would be *false* to say that *if a steel Penrose triangle were placed in a 4000 deg. F oven, it would turn into a rose*. The best explanation for this is that (13) is a *non-vacuous* counterpossible that it is about impossible hypothetical circumstances and not about how they are represented. That is, the criteria we use to evaluate (13) are circumstantial in Vetter's sense, e.g. information about the composition of steel, the shape of steel, and the melting point of steel, and these are not epistemic.

A similar story can be given for (14). Note that there is a context where it would *not* be appropriate to respond to (14) by saying (14*): "Yes, but if *Mary Ann Evans* had been Sean Spicer's biological parent, then Sean Spicer would definitely *not* have been an eloquent speaker." In such a context, the fact that George Eliot = Mary Ann Evans likely still holds, and so (14) is not referentially opaque with respect to the names—in that context, you can substitute 'Mary Ann Evans' in for 'George Eliot' and keep the same truth value.[68] Of course, there might be a context where (14*) is appropriate, say a

[68] One might object that the relevant conditional is instead 'if the actual biological parent without a Y chromosome of Sean Spicer had been George Eliot, then Sean Spicer would be a much more eloquent speaker'. Suppose that Sean Spicer's actual biological parent without a Y chromosome is Smith. Then we cannot substitute 'Smith' in for 'the actual biological parent without a Y chromosome of Sean Spicer' without changing the truth value of (14). Since they are proper names, this means that the conditional is referentially opaque and is best given an epistemic reading. I still disagree. First, in the right context, say where we are looking through a photo album of Sean Spicer's childhood, it still might be true to say, 'if Smith had been George Eliot, then Sean Spicer would have been a much more eloquent speaker'. Second, this context shows, again, that it is not the impossibility of the antecedent doing the work here but other features of the context.

context that makes relevant an impossible world where George Eliot is *not* Mary Ann Evans and the former is eloquent and the latter inarticulate. Whereas the first context seems to rule out referential opacity, in the latter context (14) seems to be referentially opaque. Furthermore, I can agree that maybe the best explanation for why an impossible world where George Eliot is not Mary Ann Evans and the former is eloquent and the latter inarticulate is made relevant in the latter context is tied to an epistemic story about some agent. Thus, it seems that any referential opacity created by (14) is not directly a result of the modal status of the antecedent but instead is a result of the similarity relation picked out by a given context. Therefore, we are neither forced to read all non-vacuous counterpossibles as referentially opaque and nor as *about* a subject's evidence nor as *about* the representational or cognitive features of the terms.[69]

2.4 Worries for Substantive Realism

There is another response to the more general claim that referential opacity in any expression indicates that the expression should be given an epistemic reading, i.e. as directly or indirectly about the representational or cognitive features of the terms that feature in the expression. This is an alternative that will especially appeal to someone who thinks that counterpossibles are ultimately made true by substantive facts about essences or substantive relations of grounding or metaphysical dependence, and so

[69] Despite my criticisms, I think that my view of counterpossibles shares in some of the spirit of Vetter's position. As we will see, I do think that representational and cognitive features of terms sometimes matter to the evaluation of counterpossibles, but I do not think this means that counterpossibles are *about* those features or are concerned with those features. Moreover, I think that there are contexts where we evaluate counterpossibles by holding fixed the representational features of a term, e.g. the actual application conditions for the name of an impossible object, and then consider what are the *circumstantial features* of an impossible world where the application conditions for the name of an impossible object are impossibly satisfied.

counterpossibles are in part about substantive metaphysical features of the nonrepresentational world.

The objection comes from the literature on grounding. Jenkins (2011) notes that claims of metaphysical dependence or grounding do not always permit substitution of necessarily coreferring expressions *salva veritate*.[70] For example, suppose it is true that there is identity between mental states and brain states, e.g. suppose that S's pain state = S's brain state B. Suppose it is true to say, 'S's pain state *depends* on S's brain state B.' But, supposing that it is always false to say, '*x* depends on *x*', it is false to say, 'S's pain state depends on S's pain state.' Thus, given the suppositions, dependence claims (and related grounding claims) are referentially opaque. Moreover, they are presumably referentially opaque in every context. However, proponents of metaphysical dependence or grounding hold that dependence and grounding claims are about the things that feature in those claims and not about our representations of those things.[71] Someone that finds this line of reasoning plausible can also take it as a reason to reject the claim that counterpossibles must be concerned with our representations simply because they are sometimes referentially opaque.[72]

---

[70] Jenkins (2011), p. 270.

[71] However, is grounding or dependence *best* characterized or explained using a referentially opaque construction? It isn't clear. Jenkins (2011) goes on to suggest that the dependence relation—a relation picked out by the two-place predicate '…depends on…'—is not a two-place relation but a four-place relation. But then it isn't clear Jenkins's proposed four place relation needs to be characterized or explained in terms of a referentially opaque two-place predicate. It likely shouldn't be since Jenkins proposes the four-place relation to get rid of the awkwardness of saying that dependence is a relation that seems to hold between two things and also *not* hold between the same two things when the two things are in fact identical, e.g. a brain state and a mental state.

[72] It is worth pointing out that the deflationary approach to metaphysics I base my account of counterpossibles on also can accept claims of grounding and metaphysical dependence, and so also accept this general objection to Vetter's argument. However, the deflationary approach I base my account on will deny that grounding claims or claims of metaphysical dependence are made true by metaphysically substantive features of the nonrepresentational world.

I think the responses to epistemic readings of counterpossibles I have looked at so far reveal that it is important to untangle three pertinent, yet different, questions regarding counterfactuals (including counterpossibles). The first question is: *what makes counterfactuals true* (*relatedly, what features of the world do counterfactuals purport to describe*)? For example, you might think that primitive subjunctive facts are what make counterfactuals true.[73] Or you might think that what makes counterfactuals true are facts about conceivability or facts about the validity of certain arguments. Or you might think that what makes counterfactuals true are facts about similarity between elements of modal space; and, further, that facts about similarity between these elements are grounded in facts about substantive metaphysical properties and relations such as essences or grounding.[74] Alternatively, you might adopt a deflationary attitude towards truth, and claim that, while counterpossibles can be true or false, nothing *makes* them true or false and they are neither describing substantive modal features of the world nor describing the validity of inferences. This is the approach I take in my account of counterfactuals and counterpossibles, which I will discuss in the next chapter, so let me set this question aside for now.

The second question is: *what are counterfactuals about*? According to Lewis (1973), his counterfactuals are about *both* similarity between possible worlds or counterparts *and* whatever actual things those are supposed to represent. However, he goes on to say that even someone who takes a *strict metalinguistic* view of counterfactuals, where counterfactuals are backed solely by the existence of valid

[73] See Lange (2009).
[74] See Kment (2006, 2014).

arguments, and not similarity relations between elements of modal space, may insist that counterfactuals are about whatever the antecedent or consequent are about.[75] For the most part, I agree, and space does not permit me to consider or argue against other suggestions. So, let us accept that counterfactuals and counterpossibles are simply about whatever the antecedents or consequents are about.[76]

The third question is: *what criteria do we use to evaluate a counterfactual*? I think this has been the pertinent question driving the literature discussed so far, and not whether a counterfactual or counterpossible is best read as being *about* something epistemic or circumstantial. Again, counterfactuals are about whatever the antecedent and consequent are about, and this is true regardless of the modal status of the antecedent (or consequent). Instead, what really matters is whether similarity of worlds should be ranked according to our evidence, or how things seem to some agent, or whether similarity should be ranked according to some other criteria independent of any agent's epistemic position. Given this, we can roughly divide criteria for evaluating counterfactuals and counterpossibles into three groups: c*ircumstantial criteria* that rely on, roughly, similarity between things, properties, relations, or events; e*pistemic criteria* that rely on, roughly, similarity between evidential states or appearances; and *conceptual criteria* that rely on,

---

[75] See Lewis (1973), p. 66.

[76] However, I will hedge somewhat from Lewis, and assume only that counterfactuals are only about whatever the antecedent and consequent are about, and not assume that counterfactuals are about similarity relations. To preview my view: I will end up granting that talk of similarity between worlds possible or otherwise play an important role in the evaluation of counterfactuals and counterpossibles. Moreover, I will even grant that possible worlds and impossible worlds exist, but only as hypostasizations of the way we speak and that speaking of similarity between these hypostasizations merely provides a formal device in the object language for illustrating or expressing rules that govern language use. I neither posit the existence of worlds possible or otherwise nor the existence of similarity facts between them as things that make counterfactual and counterpossible claims true or as things that we are attempting to describe, and which explain *why* we talk about counterpossibles and metaphysical similarity.

roughly, similarity between the conceptual rules that govern our expressions or the conceptual relations between our expressions.

Exactly what criteria are used, or should be used, is determined by the context in which the counterfactual or counterpossible is being considered. For example, presumably

(15) If water were XYZ, then water would not be H_2O

can be evaluated with respect to similarity between the evidential or phenomenological states of possible inquirers in possible environments. Alternatively, we may evaluate (15) with respect to circumstantial similarity between various impossibilities and possibilities. Regardless, contrary to the epistemic views discussed in 1.2 and 1.3, we are not constrained to reading counterpossibles epistemically simply because the antecedent is impossible.

Still, what criteria are appropriate for evaluating counterpossibles more generally will not always be uncontroversial. Consider the following counterpossibles:

(16) If intuitionist logic were correct, then $\sim\sim p \rightarrow p$ would not be a logical theorem.

There is a perfectly good sense in which (16) is not about a scenario in which if intuitionist logic is correct *for all anyone knows*, but instead it is about a scenario that is impossible assuming classical logic. Or, recall the counterpossible:

(14) If George Eliot had been Sean Spicer's biological parent, then Sean Spicer would be a much more eloquent speaker.

There is a perfectly good sense in which (14) is not about a scenario in which George Eliot is Sean Spicer's biological parent *for all anyone knows*, but instead it is about an impossible scenario in which George Eliot is indeed Sean Spicer's biological parent. Likewise, for many of the other plausibly non-vacuous counterpossibles discussed so far. Given that sense, we would not use epistemic criteria of similarity to evaluate those counterpossibles. But, focusing on (14) or (16), if the context is such that the epistemic criteria are off the table, the question is: in order to determine similarity between the relevant worlds should we or do we even need to rely on:

(a) circumstantial criteria that appeal to substantive facts about metaphysical essences, metaphysical grounding, metaphysical structure, or metaphysical fundamentality; or

(b) can we get by with more theoretically minimal conceptual criteria?

Some argue that these counterpossibles should be evaluated according to circumstantial criteria that appeal to substantive metaphysical facts. For example, some think that we should evaluate (16) with respect to similarity of non-conceptual logical structures or how various worlds, possible or otherwise, are carved at their respective logical joints.[77]

So, what exactly might a circumstantial criterion that appeals to substantive facts about metaphysical essences, metaphysical grounding, or metaphysical fundamentality look like? When offering a possible explanation of the extent of metaphysical necessity, Nolan (2011) suggests that one can take the metaphysical necessities to follow, perhaps logically, from a *distinctive set of metaphysical truths*. For our purposes, this set of

[77] See Sider (2011) and Kment (2014).

metaphysical truths is not just the set of metaphysical necessities. For example, that gold is the element with atomic number 79 is presumably metaphysically necessary, but that gold is the element with atomic number 79 or Socrates was born in Ohio is also metaphysically necessary. The former seems like a good example of the kind of distinctive truth we are looking for, while the latter seems to merely follow, perhaps logically, from the former. In addition, Fine (1994) famously argues that while it is presumably metaphysically necessary that *Socrates exists just in case the set {Socrates} exists*, when considering *what it is* to be Socrates, we don't really seem to care about the necessary connection between Socrates and sets.[78] So, likely the distinctive set of metaphysical truths we are looking for are truths that are particularly important to philosophical theorizing such as truths regarding the nature of things, set theoretic facts, certain identity claims (e.g. between mental states and brain states), facts about the relation between parts and whole, etc. Furthermore, perhaps it is this set of distinctive metaphysical truths that guides us when making judgments about metaphysical similarity between worlds.[79]

Kment (2014) takes the distinctive set of metaphysical truths to be the *metaphysical laws*, which then are used to rank similarity between worlds.[80] Recall that Kment thinks that metaphysical laws are facts about the "deep features of the world order", e.g. facts about essence, metaphysical grounding, or metaphysical structure.[81]

---

[78] See Fine (1994), p. 5.
[79] There is a sense in which I agree with this account. However, as we will see, I do not think that there is anything metaphysically substantive about these claims, but I think instead they are conceptual truths that play important theoretical roles in our cognitive lives.
[80] Also see Schaffer (2010) and Rosen (2006) for more discussion on metaphysical laws. Also see Krakauer (2012) for another account of similarity based on the idea of metaphysical laws.
[81] Kment (2006) Section 5 and Kment (2016) pp. 161-173.

According to Kment, the metaphysical laws are what connect the fundamental to the less fundamental, the grounds to the grounded.[82] Here, then, are Kment's standards of similarity used to evaluate counterfactuals:

(MSR) Metaphysical Similarity Relation:

1. It is of first importance to maximize match with respect to the metaphysical laws.
2. It is of the second importance to maximize match in the natural laws.
3. It is of the third importance to avoid large alien violations of the actual laws of nature [i.e., to avoid large, widespread violations of the laws of nature].
4. It is of the fourth importance to maximize match in particular matters of fact.
5. It is of fifth importance to avoid small alien violations of the actual laws of nature [i.e., to avoid even small, localized, simple violations of the law].[83]

This account of ranking similarity follows Lewis (1986b). However, it does not yet help us determine similarity when considering impossible worlds, so it does not yet help us evaluate counterpossibles. For example, when considering what would, per impossible, be the case were I monkey, should we look at impossible worlds where general necessary truths about species membership are false across the board or only look at impossible worlds where I am some sort of metaphysical exception, so to speak. Following Lewis (1986b), Kment notes that when we rank similarity with respect to the laws of nature, it is very important to avoid large, widespread violations of the actual laws of nature and less important to avoid small violations of the actual laws of nature, i.e. avoid big

[82] See also Schaffer (2017).
[83] See Kment (2014), p. 219.

"nomological miracles" but little "nomological miracles" are okay.[84] So, to apply Kment's account to counterpossibles, maybe we should say that when evaluating counterpossibles, it is very important to avoid large, widespread violations of the actual metaphysical laws and less important to avoid small violations of the actual metaphysical laws, i.e. avoid big "metaphysical miracles" but little "metaphysical miracles" are okay.

Recall that the question I raised was: when evaluating counterpossibles such as (14) or (16) do we need to appeal to circumstantial criteria that appeal to substantive facts about metaphysical essences, metaphysical grounding, or metaphysical fundamentality, or can we get by with more minimal conceptual criteria? We can approach an answer to this question by first asking a related question: What exactly would need to change in the actual world in order for the antecedent (or consequent) of a given counterpossible to obtain? Consider (14), for example. For someone who is not an essentialist about biological origin, the actual world merely needs to change such that George Eliot is Sean Spicer's biological parent as opposed to whoever actually is. We can say that, relative to the considered modal beliefs of our non-essentialist, such a world is a metaphysically possible world modulo small, localized nomological miracles.

However, things are more complicated if we consider an essentialist about biological origin who also thinks that essences are not contingent. For our non-contingent essentialist, then, not only does the world need to change such that George Eliot is Sean Spicer's biological parent, but *in addition* some fact about essential properties needs to change. One option, of course, is to consider an impossible world where something

[84] See the explanation of Lewis's response to Fine (1975) in 2.1 above. However, one difference between Kment and Lewis is that Kment thinks that there can be actual exceptions to the actual laws of nature.

general about essences has changed, maybe a world in which there are no essential properties or essential properties are contingent. Presumably, relative to the considered modal beliefs of our non-contingent essentialist, those impossible worlds require a massive and widespread divergence from actual facts about essences. Another option is to only change some essential fact about the *particular* people involved, which in this case is Sean Spicer. In this case, to evaluate (14), we need to not only make room for small, localized nomological miracles but also (presumably) small, localized "metaphysical miracles".

Now the question is: do we need to follow substantive realists in thinking that we need substantive features about essential properties or grounding relations—features metaphysically on a par with those features described by empirical sciences yet not discoverable though conceptual analysis or straightforward empirical means—in order to make sense of our talk and judgments of metaphysical similarity, which are used to evaluate counterpossibles like (14)? For example, maybe we can posit the existence of modal forces in the world, somehow analogous to the nuclear strong force, and think of essences as a sort of modal superglue that binds certain properties to the objects that instantiate them in a way that is stronger than mere metaphysical necessity, which is somehow more analogous to the weak force. Or maybe we can posit that reality has a synchronic metaphysical structure, and the metaphysical laws are what link the more fundamental aspects of the world, in this case essences, with the less fundamental aspects of the world, the things that possess the essential properties. So that, when evaluating (14), we are judging similarity with respect to the metaphysical laws that reflect the relations between the more fundamental with the less fundamental.

This approach does provide criteria for evaluating counterpossibles and metaphysical similarity between worlds that does not rely on the subjective perspective of any agent and does not seem to be about representations of the world. Unfortunately, I do not think that this approach is the best way to go because appealing to substantive accounts of essence, grounding, or metaphysical structure comes with metaphysical and epistemological worries that are best to avoid if we want to understand metaphysical similarity and illuminate why some counterpossibles are non-vacuous. While I do not have the space to conclusively argue that these worries cannot be addressed by the substantive realist, I will try to at least explain the worries enough to motivate looking for an easier way to think about metaphysical similarity and non-vacuous counterpossibles.

First, while interesting to think about, substantive relations akin to a modal superglue are hard to fit into a naturalistic conception of the world. In addition, though not conclusive, many have raised significant theoretical, metaphysical, and epistemological doubts concerning the substantive notions of essence, grounding, fundamentality, and structure that the robust account of similarity appeals to.[85] For example, from a more austere perspective, it seems that many competing views about what properties are essential, what grounds what, or what ultimately counts as fundamentally real will be equally adequate when it comes to our polished empirical and mathematical pictures of the world. When it comes to a metaphysical theory being empirically adequate, it doesn't seem to matter whether the non-fundamental is less real than the fundamental, or whether the grounding structure of the world bottoms out into

[85] For example, see Wilson (2014), Rayo (2015), and Warren (2016, 2017).

fundamental grounds, or whether it's turtles all the way down.[86] So, often it will be hard to see what interesting difference *substantive* metaphysical facts about grounding or fundamentality have for many of our theoretical projects.[87]

Additionally, unlike considerations of similarity based on metaphysically substantive facts, considerations of similarity based instead on empirical and conceptual features of the world will be easier to track across possibilities and impossibilities because our methods for understanding empirical and conceptual matters are, for the most part, reliable. But, if we assume a substantive realist picture about modal properties, real definitions, essential properties or grounding relations, etc., which are supposed to be matters that fall in between empirical and conceptual matters, it seems that answering many important philosophical questions will require epistemically suspect methods that go beyond more reliable empirical methods and the methods conceptual analysis (or conceptual ethics).[88] If that is the case, and we base our evaluation of counterpossibles or metaphysical similarity on metaphysically substantive facts, then it seems that we must either be skeptics about whether we know certain counterpossibles are false or we must accept that which of our beliefs that certain counterpossibles are true is a matter of sheer luck. But this seems false. Consider the following claim: impossible worlds where Hobbes squared the circle are more similar to the way things actually are than impossible

---

[86] However, Schaffer (2010) argues that grounding priority should be given to wholes and not their parts and that this is backed by considerations found in research on quantum entanglement. However, this doesn't mean that we need to posit substantive grounding facts on a par with facts about the quantum world to make sense of the research. Instead, we might be able to get by with more humble conceptual considerations about what we mean by 'whole', 'part', and what it means to say that 'parts depend on their wholes'.

[87] However, this is not to say that all talk of metaphysical relations such as 'grounding' and 'in virtue of' is not valuable. In chapter 3, section 4, I will offer a normativist account of talk of essences and metaphysical dependence.

[88] See Thomason (2017).

worlds where everything is true. That claim is true, I know it is true, and I don't think that it's a matter of luck. So, I think we should avoid appealing to substantive metaphysical features to explain our talk of metaphysical similarity.[89]

In general, it seems that whatever metaphysical and epistemological concerns there might be with making sense of how worlds are metaphysically similar to one another will merely get shifted around and not answered once talk of substantive relations of essence or ground are brought into the picture. So, it would be better if we can give an account of metaphysical similarity that does not rely on these substantive notions and that can give a clear epistemological story of our knowledge of counterpossible claims and claims about metaphysical similarity.

2.5 Summary

I started this chapter by providing a general over view of similarity between possible worlds and outlined Lewis's similarity relation used to respond to Fine's (1975) objection to his account of counterpossibles. I then suggested that such a similarity relation is not sufficient to account for similarity between impossible worlds. I then discussed Brogaard and Salerno's partially epistemic similarity relation, ESR. I argued that, while subjective epistemic considerations are sometimes important, (ESR) did not fully capture the different dimensions of similarity between impossible worlds we might be interested in looking at. I also outlined Vetter's (2016) argument that non-vacuous counterpossibles must be referentially opaque and are best read as about epistemic.

[89] This is a specific application of a more generall criticism of non-causal realism discussed in Warren (2017), which is a great synthesis and presentation of skeptical arguments against non-causal realism in various domains.

Contrary to this, I argued that there are plausible contexts in which many counterpossibles are not referentially opaque and yet should still be interpreted as non-vacuous. Moreover, I provided examples of counterpossibles that do not seem to be about evidence, appearances, or representations.

I then argued that the important question is not whether counterpossibles are about something epistemic or circumstantial, but deciding what criteria are needed to evaluate a counterpossible on the basis of metaphysical similarity: epistemic criteria, circumstantial criteria, or conceptual criteria. I also argued that the modal status of the antecedent of a counterfactual does little to help us decide. Instead, what matters are broader contextual considerations. However, I argued that trying to evaluate certain counterpossibles and metaphysical similarity by appealing to substantive metaphysical natures or substantive relations of metaphysical dependence or fundamentality comes with metaphysical and epistemological worries that are best to avoid.

CHAPTER 3: COUNTERPOSSIBLES FOR MODAL NORMATIVISTS

In this chapter, I offer a novel account of non-vacuous counterpossibles. My account builds on ideas found in *modal normativism*, an account of metaphysical necessity developed by Amie Thomasson. So far, I have discussed what non-vacuous counterpossibles are, and I have suggested that, in addition to supporting intuitions about certain conditionals, non-vacuous counterpossibles can do a lot of philosophical work.[90] All of this provides reason to think that any philosophical theory of metaphysical modality should be able to offer some kind of account of non-vacuous counterpossibles. So, in this chapter, I show how modal normativism can do just this. Conversely, I argue that by taking on the modal normativist's deflationary approach to questions of modal metaphysics, we gain a clear and plausible account of counterpossibles and metaphysical similarity that avoids the difficulties found in both the epistemic and substantive realist accounts.[91]

However, modal normativism and related accounts of modality are not widely familiar, so I begin in section 3.1 by briefly discussing the account. Then, in section 3.2, I will build on my discussion of normativism by explaining how I incorporate both expressive elements and similarity elements in to my overall account of counterpossibles. In section 3.3, I will explain in more detail how the modal normativist thinks about metaphysical similarity. After that, in section 3.4, I will explain the advantages my

[90] See chapter 1.
[91] See chapter 2.

account has over the accounts discussed in the last chapter. Finally, in section 3.5, I will briefly look at an alternative deflationary account of modality, which I call strong modal classificationism. I will argue that strong modal classificationism fails to account for the role and function of modal discourse, and that it should not be confused with the account I am offering here.

3.1 Modal Normativism

Modal normativism is an account of metaphysical necessity developed primarily in Thomasson (2007a, 2007b, 2009, 2013).[92] Modal normativism is motivated largely by general epistemological and metaphysical concerns that come with positing substantive modal facts and properties or substantive possible worlds as *truthmakers* for modal claims.[93] According to modal normativism, viewing modal claims as descriptive, quasi-scientific claims in need of truthmakers, such as possible worlds or modal facts and properties on par with the facts and properties described the empirical sciences, is misguided. Instead, modal normativists argue that by looking at the *function* of modal thought and discourse, we can see that modal claims primarily have an *expressive* or *pragmatic* role as opposed to a descriptive role. The modal normativist attempts to:

> … make sense of claims about metaphysical necessity by way of understanding the normative rules for using our [non-modal] terms. (Thomasson 2010, 96)

Moreover, on this view, claims of metaphysical necessity:

---

[92] However, more work is forthcoming. Modal normativism is a part of a family of modal theories that include modal expressivist and inferentialist accounts found in the work of Gilbert Ryle (2000), Simon Blackburn (1993), Wilfrid Sellars (1958), and Robert Brandom (2000, 2007, 2008, 2014) as well as modal conventionalist accounts found in the work of A.J. Ayer, certain readings of Wittgenstein, and Alan Sidelle (see Sidelle 1989). See Thomasson (2009) for a more detailed history and defense of this approach.
[93] Thomasson (2010), p. 136.

> …primarily serve the *prescriptive* function of expressing [but *not* describing] semantic rules for the terms used in them, or their consequences, while remaining in the object-language. (Ibid.)

I will now elaborate on this account and outline some of its virtues. In the next section, I will show how to apply elements of this account of modality, along with related work found in Brandom (2008) and Sellars (1958), to give an account of counterfactuals and counterpossibles. In section 3.3 I will develop an account of metaphysical similarity using metaphysical laws, but metaphysical laws suitable for modal normativism.

Modal normativism presupposes that *competent use* of non-modal vocabulary is *governed* by constitutive rules. Some of these constitutive rules come in the form of the *application* and *coapplication conditions* of a term, which can, but need not be, be expressed in a metalanguage and constitute the *frame-level content* of non-modal vocabulary.[94] Some *object language analogues* of the frame-level content of a term, e.g. a name or sortal, are object language statements of the existence, identity, and persistence conditions of the items to which the term purports to refer. Indeed, according to modal normativism, the function of basic modal terms is to illustrate, endorse, or express actual application conditions, which govern the use of non-modal vocabulary, in the object language, e.g. 'All bachelors are *necessarily* unmarried'. Likewise, certain consequences of the application and coapplication conditions yield object language persistence conditions, which are also usually stated in modal terms as the conditions under which an entity *could* or *could not* survive. Since, on this view, basic truths about the identity and

[94] For example, see Thomasson (2007a), pp. 38-45 and Thomasson (2015), pp. 89-94. Thomasson takes the frame-level content of a term, such as its application conditions, stated in a metalanguage, to be necessary for establishing *determinate* reference when the meaning of a term is grounded, e.g. through a christening. However, mastery of frame-level rules is not necessary for a subject to talk *about* whatever the term refers to, but it is necessary to count as a *competent* user of the term.

persistence conditions are analytic, all modal truths are analytic or derived from analytic truths with an empirical truth.[95]

Two brief points of clarification are required. First, it *might* be that the rules that constitute the frame-level content of expressions can be explicitly stated in a metalanguage, but *they need not be*. Sometimes the rules are only tacitly followed by language users, and competent language users need not be able to ascend to a metalanguage and recite the rules.[96] Second, the rules are *constitutive* in the sense that, for a speaker to count as using any term, the speaker must be subject to the rules that govern the use of that term (though not necessarily have mastered the rules).[97]

So, in general, basic modal claims, for claims of metaphysical necessity, simply reflect the frame-level content of the relevant terms involved in those claims, or what follows from that frame-level content, perhaps combined with empirical facts. One worry often cited for past theories of modality that treated modal claims as analytic was that such theories cannot account of *a posteriori* necessities. However, this is not a problem for modal normativism. For example, suppose that there is a schematic linguistic rule for the term 'water' along the lines of "Apply 'water' to the watery stuff in the environment just in case it has chemical composition *C* (where *C* is determined by the chemical composition of whatever sample was used in the 'baptism' of the term 'water')". According to modal normativism, this rule, along with the empirical information that the chemical composition of the watery stuff in our environment is H_2O, can all be expressed with the object level sentence 'Water is *necessarily* H_2O', where the modal term makes

[95] Thomasson (2007a), pp. 62-63.
[96] See Thomasson (2007b), p. 140, and Thomasson (2015), pp. 91-93.
[97] See Thomasson (2007b), p. 138.

explicit that we are dealing with a constitutive rule governing our application of the term 'water'.[98]

What advantages might there be to illustrating, endorsing, or expressing metalanguage rules in the object language? Thomasson highlights three reasons. First, conveying rules while remaining in the object language doesn't require ascending into the meta-language. This matters because the constitutive rules that govern language use may only be tacitly followed and difficult to express in the meta-language, if they are expressible at all. Thomasson (2015) claims that:

> Application conditions should be thought of as semantic rules analogous to grammatical rules; just as competent speakers must be masters of following grammatical rules, but need not be capable of stating them (although it is plausibly those speakers' normative practices in speaking, correcting the speech of others, etc., that fixes the grammatical rules for a particular language), so must competent speakers be masters at following the semantic rules—but need not be capable of stating them. (92)

Furthermore, plausibly, some terms may be semantically basic insofar as they cannot be learned by learning definitions stated in other terms, and may have been learned by other means such as ostention.[99]

Second, by expressing rules in an indicative mood, e.g. "Necessarily, bachelors are unmarried", as opposed to an imperative mood, e.g. as the linguistic command "Only apply 'bachelor' to unmarried people!", means that the rules can be used straightforwardly in reasoning, e.g. "Bachelors are necessarily unmarried; Jones is a bachelor; thus, Jones must be unmarried". Finally, modals allow us to express permissions, e.g. "It is *possible* that someone from Miami is a bachelor" as well as

---

[98] Ibid., p. 145.
[99] Ibid.

requirements, e.g. "Bachelors *must* be unmarried".[100] As we will see in Chapter 5, we can also use object level modal claims to *metalinguistically* press for changes in the rules that govern our expressions or how we apply terms, e.g. "Some women *are possibly* bachelors".

Modal normativism goes along well with a deflationary theory of truth, e.g. a deflationary theory in which the truth predicate is not something that requires a substantive analysis, e.g. positing a truth-making relation between propositions and substantive facts, because it primarily functions as a device of generalization and not as something that picks out a substantive property.[101] So, it is important to stress that, while modal claims primarily play a prescriptive, and not descriptive, role in discourse, they are still truth apt.[102] Thus, given a deflationary view of truth and the prescriptive (as opposed to descriptive) function of modal thought and discourse, there is no need to posit the existence of substantive possible worlds, substantive modal properties, or substantive properties of essence, or any other substantive relations of grounding or fundamentality to make modal claims true.

What's more, as a result, the burden of giving a plausible story for how we come to know what is metaphysically necessary is considerably lightened. Given the normativist explanation of metaphysical necessity, we come to know modal truths primarily through conceptual analysis, perhaps along with making empirical discoveries. For example, on the normativist account, coming to know that 'water is *necessarily* H_2O'

---

[100] See Thomasson (2007b), pp. 138-140. See also Thomasson (2013).
[101] For example, see Horwich (1990).
[102] It is also important to stress that modal claims are made in the object language and so are not descriptions of, nor do they report, the rules that govern the use of non-modal terms. See Thomasson (2007b), pp. 140-143.

does not require that we have some insight into some substantive modal property of water. All it requires is a tacit understanding of the frame-level rule governing the application of 'water' along with certain empirical information obtained via straightforward empirical methods, e.g. that water has the chemical structure H_2O.[103]

Before moving on, I need to make two more important points of clarification. First, the *object level* expressions of constitutive rules that govern our concepts, i.e. the statements of metaphysical necessities, are about the world and not about our conceptual conventions. According to modal noramtivism, modal claims such as 'all bachelors are necessarily unmarried' or 'the number π is necessarily a transcendental number' are *absolutely not* purporting to *describe* the rules that govern the terms involved in those claims. The rules of use for non-modal expressions are stated, if ever explicitly stated at all, in a *metalanguage*. Modal claims, on the other hand, are stated in the *object language*, and, insofar as the terms are being used and not mentioned, are still in that sense about the world. I will suggest in section 3.4 that the same goes for many other metaphysical claims, e.g. claims about the nature of things, metaphysical grounding, or metaphysical similarity.

Finally, it is important to note that the modal normativist can still freely talk about modal facts, modal properties, and even possible worlds. Thomasson (2007b) suggests that even though the normativist is suspicious of claims about substantive modal features, one can nonetheless make trivial inferences from modal claims to the existence of modal facts, modal properties, and even possible worlds—but these are not things that explain

[103] See Sidelle (1989) and Thomasson (2007b).

what makes our modal claims true.[104] I will build on this idea, and the work of Schiffer (2003) and Steinberg (2013), in the next section and in chapter 6, where I will briefly propose that talk of metaphysically impossible circumstances conceptually falls out of talk of what is metaphysically impossible.

3.2 General Talk of Worlds and Similarity for Modal Normativists

My normativist account of counterpossibles relies on two different elements, an expressivist element and an element using worlds and similarity, which are two different ways of thinking about counterfactuals that seem to be in tension. So, in this section I am going give a general account of counterfactuals in expressivist terms and then work my way into an account introducing talk of worlds and similarity while respecting the non-descriptivist spirit of expressivism.

On the one hand, I take an expressivist approach to modal claims:

> A counterfactual 'If *p* were the case, then q *would* be the case' *expresses* a *requirement* to accept that *q* given *p* as a hypothetical supposition and a range of *relevant auxiliary assumptions* determined by the context. (*Expressive Element*)

On this approach, a counterfactual is an object language expression of constitutive inferential connections between '*p*' and '*q*' for a given a range of relevant auxiliary assumptions. However, I also stick with certain elements of the extended Lewis-Stalnaker presentation of counterfactuals and say that:

[104] See also Thomasson (2015).

> A counterfactual 'If *p* were the case, then q *would* be the case' is *assertible* in *w* just in case there is some *p q* world that more relevantly similar to w than every *p* non-*q* world. (*Similarity Element*)

A couple of clarificatory points. I take the *Similarity Element* to apply to both possible and impossible worlds. In this way, I am staying within the extended Lewis-Stalnaker framework. By 'assertible', I only mean that *it is acceptable to introduce the counterfactual into discourse*, to add it to one's scorecard, or group scorecard, so to speak. Given a deflationary view of truth, I am even fine with saying that counterfactuals are truth apt, and that a counterfactual '$p \mathrel{\square\!\!\rightarrow} q$' is true just in case certain worlds are more similar to the actual world than certain others.

What I will not say is that facts about similarity and worlds are what *make* counterfactuals true. This is because I take the two elements to be related in a trivial way. I will argue that righthand side of *Similarity Element*—a claim seemingly about worlds and similarity between them—to be just another claim made in the object language that illustrates, endorses, or expresses a metalanguage requirement that governs the use of '*p*' and '*q*'. For example, it follows from *Expressivist Element* that someone fails to express a requirement to accept *q* given *p* when there is some relevant auxiliary assumption determined by the context, perhaps *p* itself, that infirms the move from *p* to *q*. However, the arguments I give in this section imply that another perfectly acceptable way to use the object language to convey when someone in *w* fails to express a requirement to accept *q* given *p*, is with a claim that there exists some *p* non-*q* world that is more relevantly similar (or just as similar) to *w* than some *p q* world. In other words, I will argue the following:

> 'There is some *p q* world that is more relevantly similar to *w* than every *p* non-*q* world' *expresses* a *requirement* to accept that *q* given *p* as a hypothetical supposition and a range of *relevant auxiliary assumptions* determined by the context. (*Worldwide Expressive Element*)

So, combining *Worldwide Expressivist Element* with *Similarity Element*, the latter says no more than a counterfactual is assertible just in case it illustrates, endorses, or expresses a requirement, which will depend on the actual rules and permissions governing the terms used in the counterfactual (though it will not depend on anyone being able to explicitly state those rules and permissions). An implication of my view is that evaluating counterfactuals and counterpossibles merely comes down to using straightforward empirical methods and conceptual analysis.

The final point is that I am going to talk about inferences and inference tickets in what follows; so, I want to briefly say something about inferences in order to avoid confusion about my view. In some sense of the word, inferences matter to the evaluation of counterfactuals and counterpossibles on my account, so my account has something in common with other metalinguistic approaches.[105] However, there are two crucial clarifications that need to be made. On my view, counterfactuals and counterpossibles are not made true by, and are not describing, the existence of inferences. Most importantly, I am *absolutely not* talking about *formal inferences*, as inferences that are good in virtue of the logical form and the meaning of the premises and conclusion.

[105] See Lewis (1973), pp. 65-76.

Overall, since I am arguing that we sometimes need to hypothetically reason about impossibilities, we have to be careful if we want to say that counterfactuals illustrate, endorse, or express certain inferential requirements. In particular, the inferential relationships need to be fine-grained enough to prevent *explosion*, i.e. the inferential relationships need to be such that not anything follows from an impossibility. This means that the inferential relationships need to be *hyperintensional inferences*, i.e. inferences that do not necessarily respect substitution of intensionally equivalent expressions, e.g. two claims that are both metaphysically necessary. In fact, what this means is that counterpossibles are actually quite at home with the use-based semantic assumptions made by the normativist (and inferentialist). It certainly does not conceptually follow from a red ball's being colored that $2 + 2 = 4$. This means that it is color-conceptually possible for a red ball to be colored but for $2 + 2 \neq 4$. This can be expressed with the counterpossible 'if it were the case that $2 + 2 \neq 4$, then the red ball would still be colored'. Moreover, the counterpossible needs to be non-vacuous insofar we do not want to say 'if it were the case that $2 + 2 \neq 4$, then the red ball would also *not* be colored'. So, just as counterfactuals allow us to express requirements and permissions pertinent to certain non-monotonic inferences (which I will review in the next section), counterpossibles allow us to express requirements and permissions pertinent to hyperintensional inferences.

To summarize the point just made: if counterpossibles were thought of as *formal* inference tickets, assuming classical logic, then we would basically just have the orthodox account and fail at meeting most of the *desiderata* outlined in the introduction. So, I am not talking about formal inferences.

Why do I rely on both *Expressive Element* and *Similarity Element*? I rely on both elements because I think that they highlight two important insights about counterfactuals. The first insight comes from the expressive element, which is that if we do not treat modal statements as *descriptive* statements about counterfactual features of the world, we do not need to posit the existence of substantive subjunctive properties or substantive facts about metaphysical similarity between worlds to make counterfactuals true. This in turn allows us to use counterfactuals and counterpossibles with taking on substantive metaphysical commitments, which in turn give us a clear and plausible picture about what we do with counterfactuals and counterpossibles and how we come to know counterfactual and counterpossible claims—though straightforward empirical methods and conceptual analysis.

The second insight comes from the similarity element, which is that talk about worlds and similarity between them is important to evaluating counterfactual claims. First, talk of worlds and circumstances gives us a more fine-grained way of talking about possible and impossible hypothetical suppositions. For example, when considering what would be the case if Tabby the cat were on a mat, talk of worlds and circumstances as *distinct* ways that Tabby could be on a mat allows us to capture the thought that not every way Tabby could be on a mat will be relevant to the context. Secondly, talk of similarity between worlds and circumstances gives us another resource for expressing or illustrating the frame-level content that governs the use of our non-modal vocabulary. For example, in a relatively normal context, if I think that hypothetical circumstances in which Tabby is on a flying mat are more similar to how things actually are than circumstances in which Tabby is on a non-flying mat, then I presumably either have strange empirical beliefs or

conceptual beliefs about the world, e.g. I think that 'mat' applies to sorts of things that can fly.

I will say much more below, but, for now, I just want to say enough to point out that one might think that there is a certain amount of awkwardness in giving a presentation of counterfactuals in expressivist terms *and* in terms of worlds and similarity.[106] On the one hand, according to expressivist accounts, counterfactuals are *not* descriptive claims about substantive entailment relations, substantive modal facts, or about substantive relations of similarity between substantive worlds, which would then in turn serve as truthmakers for counterfactual claims. Instead, counterfactuals are thought of as 'inference tickets', that is, they are object language claims that serve to *license*, or *endorse*, certain inferences between the antecedent and consequent. For example, on expressivist readings of counterfactuals found in the work of Ryle and Sellars, the counterfactual 'if a sugar cube were stirred into a normal cup of coffee, the sugar cube would dissolve' licenses, or endorses, the inference from 'a sugar cube was stirred into a normal cup of coffee' to 'that sugar cube dissolved'.[107] Given the expressivist element,

---

[106] Traditionally, talk of truth in worlds has been associated with representationalist views in philosophy of language while expressivist and inferentialist views tend to stick to talk of appropriate moves in reasoning. For example, on the former approach, the validity of an argument is cashed out in terms of there being no world in which the premises are true and the conclusion false, while on the latter approach the validity of an argument is cashed out in terms of legitimate, gap free moves from the premises to the conclusion (e.g. see Tennant (1990)). Of course, for certain languages, soundness and completeness arguments show an equivalence between the arguments these two ways of thinking find acceptable.

[107] See Ryle (2002), Sellars (1958). While on this view the counterfactual does not describe substantive modal properties of sugar cubes, I must again emphasize that, on this view, the counterfactual absolutely does not describe the existence of an inference and so is *made true* by the existence of an inference. And, again, the inferences are not necessarily *formal* inferences, e.g. inferences that are intended to be good in virtue of form and the meaning of logical connectives. Instead, it is better to think of the inferences as *material inferences* that are good in virtue of the meaning of the concepts *because* those inferences are meaning constituting. For example, the inference from "A is to the West of B" to "B is to the East of A" is a good inference because the inference partly constitutes the meaning of the expressions 'East' and 'West'. See Brandom (2000), pp. 52-55 and Brandom (2007), pp. 657-658. Furthermore, given that I am giving an account of counterpossibles, the inference absolutely cannot be thought of *formally*.

there seemingly is no need to talk about worlds or similarity, so why bother? I adopt the similarity element, in part, for dialectical reasons since the alternative views of counterpossibles I am engaging with work within the extended Lewis-Stalnaker framework. But I also adopt it because, on the other hand, I just sketched reasons in the last paragraph to think that talk of worlds and similarity is important to evaluating counterfactual claims and useful for conveying rules that govern the use of ordinary non-modal vocabulary. However, adopting both elements is not so awkward as one might initially expect, and, in this section, I will take some time to explain why. I will do this by first giving a general account of counterfactuals in expressivist terms and then work my way into an account introducing talk of worlds and similarity while respecting the deflationary spirit of expressivism.[108]

In the last section I reviewed modal normativism, according to which basic modal expressions such as 'Necessarily, *p*' and 'Possibly, *p*' serve to illustrate, endorse, or express, but not to describe, constitutive rules and permissions, or their consequences, that govern our non-modal expressions while remaining in the object language. 'Necessarily, *p*', often symbolized with a "box operator", i.e., '□*p*', serves to correct or condemn conflicting uses of the terms involved in '*p*', and claims such as 'Possibly, *p*',

[108] I will not rely much on the expressivist view of modality developed by Simon Blackburn (1993). Instead, in addition to modal normativism, I will focus on the modal expressivism as presented in the work of Brandom (2000, 2007). This brings with it certain commitments in the philosophy of language, roughly captured by the phrase *meaning is use*. However, there are various differences in how modal normativism, as developed by Thomasson, and modal expressivism, as developed by Brandom, think about how ordinary vocabulary gets its content through use and how the content might be constituted by rules of use, e.g. Thomasson relies on an analytic/synthetic distinction while Brandom does not. I cannot hope to adjudicate these issues in this dissertation, nor can I offer a full-blown defense of any of these use-based approaches to content nor offer an original use-based account. So, I will try to stay neutral on certain issues, e.g. on whether the normative force of rules of use is primitive, as in the rationalist approach endorsed by Brandom, or is instead minimal in a way that is more consistent with use-theories found in Horwich (1990).

often symbolized with a "diamond operator", i.e., '$\Diamond p$', convey that the terms in 'p' are used in a permissible way. Now note that the counterfactual 'if p were the case, then q *would* be the case' is often symbolized with a "box conditional" operator '$p \Box\!\!\rightarrow q$', and the counterfactual 'if p were the case, then q *might* be the case' is often symbolized with a "diamond conditional" operator '$p \Diamond\!\!\rightarrow q$'. Indeed, for a long time a common, if not standard, analysis of *would*-counterfactuals was as strict conditionals of the form $\Box(p \supset q)$. Even Lewis's 1973 treatment of counterfactuals retains some of this spirit, though he thinks of counterfactuals as variably strict conditionals. For Lewis, counterfactuals are *variably* strict because they are extremely sensitive to context, as demonstrated, for example, by their non-monotonicity. In order to account for both of these features, Lewis relies on a similarity ordering of worlds to evaluate counterfactuals. However, in the end, this puts pressure on the idea that strictness is really doing anything of theoretical interest in the evaluation of counterfactuals. Still, given a normativist account of basic modal expressions, it is natural to think that the modal expressions 'would' and 'might' play some *normative* role in conditional claims correlative to the role that basic ordinary modal expressions 'necessity' and 'possibility' play in many non-conditional claims. Moreover, this normative role sometimes has to do with conveying how 'p' and 'q' are properly used, i.e. with the constitutive rules and permissions governing uses of 'p' and 'q'.

Given this normativist idea, I want to show how talk of worlds and similarity harmlessly falls out of expressivist talk of counterfactuals. I will start with an incomplete presentation of the expressivist view and work my way towards a complete presentation. Here is the first incomplete expressivist presentation:

(E1) A counterfactual '$p \boxminus\!\!\rightarrow q$' *expresses* a *requirement* to accept that q on condition of accepting that p as a hypothetical supposition.

Three points of clarification are in order. First, let me emphasize that counterfactuals are *neither* describing requirements or inferences *nor* made true by any requirements or inferences. Second, note that (E1) is given in terms of a requirement of *acceptance*. By 'acceptance' I do *not* mean *believe*—we might accept p as a hypothetical supposition without thereby believing p or committing ourselves to p other than as a hypothetical consideration. In fact, given my assumption of modal normativism, it might be better to put acceptance in terms of the application of 'p' and 'q'. In other words, it might be better to state (E1) as follows:

(E1*) A counterfactual '$p \boxminus\!\!\rightarrow q$' *expresses* a *requirement* to apply 'q' on condition of the application of 'p' to a hypothetical scenario.

On this reading, a counterfactual expresses a rule governing the use of 'p' and 'q' to the effect that in circumstances in which a person is entitled to apply 'p' can be rebuked for refusing to also apply 'q'. I do not think that what follows depends on choosing one reading over the other, so I will stick to starting with (E1). I think this may be more natural given that counterfactuals are closely connected to hypothetical reasoning where we think about what else we might have to accept given that we accept a hypothetical assumption.

Third, and more importantly, sometimes the literature on inference tickets talks about counterfactuals as expressing *entitlements* as opposed to requirements. For example, here is Sellars (1958):

> …the distinction between 'A is constantly conjoined with, but does not cause, B' and 'A causes B' is to be interpreted in terms of the idea that statements of the second form imply that *simply from the fact that something is A* one is entitled to infer that it is B, whereas statements of the first form, whatever else they may do, imply that one is not entitled to infer *simply from the fact that something is A* that it is also B.[109] (185)

But I do not think that talk of entitlements is quite the right way to put the normative point of counterfactual discourse. If I accept 'if *p* were the case, then *q would* be the case' it is not that I am entitled to accept *q* if I also accept *p*, but that I *ought* to accept *q* if I accept *p*. Just saying that I am *entitled* to accept *q* leaves open the possibility I might not accept *q*—e.g. I am entitled to a free beverage with a purchased meal, but I am still *allowed* not to take the free beverage and might not take it. So, to keep the analogy with the normativist normative interpretation of basic modal operators such as 'necessity', it is best to take the expressive force of *would*-counterfactuals as expressing a requirement rather than an entitlement. Brandom (2000) also argues that, in "the game of giving and asking for reasons", there is a distinction between what assertions we are *entitled* to versus what assertions we are *committed* to, and he phrases the latter in terms of *oughts* and *obligations*.[110] For example, when describing a game of linguistic scorekeeping, where one's score is determined by those sentences she is entitled or committed to, Brandom (2000) claims:

> …for such a game…to be recognizable as involving assertions, it must be the case that playing one [sentence], or otherwise adding it to one's score, can *commit* one to playing others, or adding them to one's score. If one asserts, "The swatch is red," one *ought* to add to one's score also "The swatch is colored." Making one move *obliges* one to be prepared to make the other as well.[111] (190-191)

---

[109] In the context of Sellars's discussion he is using 'imply' more as 'conversationally implicate' and not a more substantial relation of implication.

[110] In the next section I will briefly discuss that we need not take the normative force of modal expressions to commit us to substantive facts about normativity.

[111] Brandom (2000), pp. 190-191. Emphasis is in the original. Brandom goes on to note that it is possible to not always act in the way a player is obliged without necessarily being expelled from the game.

On the one hand, questions of entitlement are "upstream" from assertions, and come up when we are asked for reasons for making an assertion. So, e.g., if I am asked why I am entitled to claim, "The swatch is red," I can respond, "Because the swatch is crimson." On the other hand, commitments are "downstream" from assertions. Counterfactuals are one way of expressing what commitments are undertaken by making certain assertions, or accepting certain claims, without actually making those assertions, or without actually accepting those claims and thereby actually incurring the commitment.

We can ask two questions about the *requirement* expressed by a counterfactual. The first question is: where does the requirement come from? Given the discussion of modal normativism in the previous section, the requirement might *sometimes* stem from the conceptual frame-level content of '*q*' and '*p*'. For example, the satisfaction of the application conditions for '*p*' may *analytically entail* the satisfaction of the application conditions for '*q*', e.g. 'if I were in a house, then I would be in a building'.[112] Another normativist example comes from the thought that there are certain rules that govern the use of '*p*' that are jointly determined by both the application and coapplciation conditions of '*p*', which can be expressed in the object language in the form of persistence conditions, i.e. the conditions under which certain objects could or could not survive, e.g. 'if I were to flatten this clay statue, the lump of clay would survive but the statue would not'. However, the requirement might sometimes stem from the inferential relationship between *noninferential circumstances* in which it is appropriate to introduce '*p*' and those in which it is appropriate to introduce '*q*'.[113] That is, it may be that being in a

[112] Thomasson (2007b), p. 44.
[113] Brandom (2007), p. 658. By 'noninferential circumstance', I take Brandom to mean circumstances outside of the domain of logic or conceptual analysis, circumstances that involve observations,

circumstance of application for 'p' *circumstantially entails* the application of 'q', e.g. 'if the lioness were hungry, then nearby prey would be in danger' or 'if it were to rain, then the streets would be wet'.[114] What frame-level application conditions, circumstantial application conditions, or other considerations are at play will depend on the content of the antecedent and consequent as well as other features of the context.

The second question is: when does 'if *p* were the case, then *q would* be the case' fail to express a requirement? On a certain strict conditional reading of counterfactuals, the answer is: *never*. Earlier work on the idea that counterfactuals express inference tickets focused on the lawlike nature of certain counterfactual claims, so that lawlike dispositional statements or lawlike counterfactuals actually express *season* inference tickets, inference tickets that are good any time of year so to speak.[115] The less metaphorical way of making this point is that a "season" inference between the antecedent, '*p*', and the consequent, '*q*',—an inference which for expressivists and normativists partly constitutes the meaning of '*p*' and '*q*'—is an inference that is *maximally* counterfactually robust insofar as the addition of *any* range of auxiliary hypotheses will not infirm the inference. In the context of nomological counterfactuals, one way of reading Sellars (1958) is as claiming that the only purpose of introducing talk of causal necessity into our discourse, as we might with '$\Box_C(p \rightarrow q)$', is to express that the inference from '*p*' to '*q*' is a meaning constituting inference that is maximally counterfactually robust. This idea is also reflected in the modal normativist view of

circumstances that involve the recognition of certain cultural practices, among other things. Importantly, I will take these circumstances are tractable using empirical methods.

[114] Brandom (2007), p. 658.

[115] See Ryle (2002), Sellars (1958).

modal operators of metaphysical necessity insofar as '□($p \rightarrow q$)' can be taken as a way of illustrating, endorsing, or expressing that '$p \rightarrow q$' reflects some frame-level rule, or consequence of some rule, that governs the use of 'p' and 'q'.[116] But, when talking about the normative force of 'would' in *counterfactuals more generally*, we need to be careful, because a counterfactual claim '$p \boxrightarrow q$' is not necessarily equivalent to a strict conditional '□($p \rightarrow q$)'. For the normativist, the latter may indeed express a sort of *everything*-considered counterfactual robustness of some meaning constituting relation between the antecedent and the consequent, i.e., a connection between 'p' and 'q' with "full-stop" normative force—but not every useful or interesting counterfactual needs to be *everything*-considered robust. Very often contextually limited robustness is useful to think about, especially when dealing with empirical counterfactuals. Furthermore, as I will argue below, having a grasp on even these weaker counterfactuals, i.e. an understanding of even *some*-things-considered robustness, can be equally important for being able to properly use ordinary empirical terms such as 'cat'.

So, we need to adjust (E1) to recognize that the requirements expressed with counterfactual claims often depend on *contextual* considerations—a requirement expressed by a counterfactual may not hold "full-stop" insofar as it may hold in one context but not hold in others. This gives us:

(E2) A counterfactual '$p \boxrightarrow q$' expresses a requirement to accept that q on condition of accepting p as a hypothetical supposition given *auxiliary assumptions* determined by the context.

[116] Of course, '$p \rightarrow q$' is itself a conditional claim of the object language, e.g. if someone is a bachelor, then that person is unmarried.

For example, suppose that in some context 'if the match were struck, it would ignite' expresses a requirement to accept that a match ignites on condition of the acceptance that the match is struck. Suppose that in this context facts such as the matches are dry, that the matches are not struck in a vacuum, etc. are held fixed. To begin to show how the expressive and similarity elements can work together, the counterfactual uttered in this context, expresses a requirement to the effect that when hypothetically reasoning about what to expect in the presence of struck matches, or when considering possible explanations for a lit match, the only *hypothetical circumstances* we are *allowed* to appeal to are those circumstances where struck matches ignite as opposed to circumstances where struck matches do not ignite (but perhaps we are still allowed to appeal to circumstances where a match is not struck but still ignites, e.g. one where a match is brought into sufficient proximity to a flame).

However, in different contexts, the requirement expressed by the counterfactual in the previous context might not hold. For example, there are contexts where we might bring in certain auxiliary assumptions—for example, that the match is underwater, that we are in a vacuum, etc.—that can easily undermine the requirement expressed in the first context. There are at least two ways this undermining can happen. It can happen *explicitly*, as when I say, 'If the match were struck *in a vacuum*, it would *not* ignite', but that need not be the case. It could be that we are in a context where we are already talking about vacuums, say a context where we are talking about space walking astronauts. In this case, 'if the match were struck, it would ignite' fails to express a requirement to accept ‘the match ignites’ given the acceptance of ‘the match is struck’ because we are

making auxiliary hypothetical assumptions that infirm the inference from antecedent to the consequent.

Of course, it could be that certain auxiliary hypothetical assumptions are *irrelevant* to the evaluation of a counterfactual in a given context. For example, suppose that while explaining to a hiker that campfires are not permitted in the forest during a long-lasting drought, a ranger says, "if the winds were to blow embers from the fire, part of the forest would catch fire and burn down". Then it would not be appropriate for the hiker to disagree by saying, "But there could be sudden rain storm that would put the fire out." The hiker is trying to introduce an auxiliary assumption about rain storms. However, since the ranger is making the claim during a long-lasting drought, the hiker's response is inappropriate because it is trying to raise a possibility that is *irrelevant* to the ranger's claim. This gives us the following adjustment:

(E3) A counterfactual '$p \,\square\!\!\rightarrow q$' expresses a requirement to accept that q on condition of accepting p as a hypothetical supposition given a *range of relevant* auxiliary assumptions determined by the context.[117]

The range of relevant auxiliary assumptions are those assumptions we hold fixed and are needed to accommodate p without deviating *too far* from what is the case in the circumstance in which we are evaluating the counterfactual.[118]

---

[117] Alternatively, (E3*): A counterfactual '$p \,\square\!\!\rightarrow q$' expresses a requirement to apply 'q' on condition of the application of 'p' to a hypothetical scenario given a *range of relevant* auxiliary assumptions determined by the context.

[118] See Stalnaker (1969) p. 104 and Lewis (1973) pp. 8-9. Often determining relevant auxiliary hypothetical assumptions is a matter of trading off between similarity and differences with the circumstance relevant for evaluating a counterfactual.

What auxiliary assumptions the context picks out as relevant will sometimes be a matter of the intentions of those considering a counterfactual, or it will sometimes be a matter of what information is salient to those considering a counterfactuals, or it will sometimes be a matter of making charitable sense of a counterfactual, or sometimes it will be a matter of consistency with what is true in the circumstance a counterfactual is being considered (as in the hiker example just outlined).[119] But it is worth noting that sometimes the context will pick out auxiliary assumptions that are true but unknown in the circumstance in which a counterfactual is being evaluated.[120] To use an example from Lewis (1973), in a context where we are considering someone's utterance, "if I were to look in my wallet, I would find a penny", facts about what is in the speaker's wallet at that moment are held fixed even though we might not yet know what is in her wallet.[121] So, in order to see if the speaker succeeded in expressing a requirement to accept there is a penny in her wallet given an acceptance that she has looked in her wallet, we actually need to open her wallet and find out. If there is not a penny in the speaker's wallet, then the speaker failed to express a requirement. The point is, sometimes how the world in fact is determines whether a requirement is expressed with a counterfactual. In this case, what will matter to what auxiliary assumptions are relevant, and so held fixed, will be

---

[119] Sometimes it is a matter of *making charitable sense of the use of a counterfactual in an argument*, i.e. as charitably reading a counterfactual so that an argument being presented comes out as valid. So, there may be a context that determines that 'if p were the largest prime, $p! + 1$ would be prime and would be composite' is false. But in the presentation of an argument, as in Williamson's example discussed above, in order to charitably take someone to making a valid argument, the charitable reading of 'if p were the largest prime, $p! + 1$ would be prime and would be composite' is as true.

[120] I mean this in a way similar to the way that there may in fact be evidence available to support my belief without my being aware of that evidence.

[121] See Lewis (1973), p. 68.

empirically tractable features of the circumstances we are considering, including the one we are actually in.

I will now begin to show how to move from talk about requirements to accept *q* on condition of accepting *p* given a range of relevant auxiliary assumptions determined by the context *to talk about worlds and similarity* while retaining the expressivist element. Basically, I start with the idea that similarity judgments about a range of hypothetical circumstances help us determine whether we have deviated too far from what is the case when considering a counterfactual. However, I will argue that these judgments of comparative similarity reflect our judgments of comparative possibilities. For example, if I could more easily inhabit some other star system in the Milky Way than inhabit a star system in the Andromeda galaxy, then the former *possibility* is more similar to actuality than the latter.[122] This goes in three steps.

For the first step, recall that the expressivist element presupposes that *talking about what is possible* is not a way of describing modal features of the world but a way of conveying rules and permissions that govern the use of ordinary non-modal vocabulary while staying in the object language, which means all we need to talk about what is possible is a tacit understanding of how to properly use ordinary non-modal vocabulary. But, for the second step, I will show that talking about *possibilities* is simply a more fine-grained way of talking about what is possible. So, once we know how to make the linguistic transformation from talk of what is possible to the introduction of talk of possibilities and worlds, we can then start talking about similarity between them. Finally,

[122] See Lewis (1973), pp. 52-53.

I will argue that this talk of comparative *similarity* of possibilities, and possible worlds, is simply another way of conveying how to properly use ordinary non-modal vocabulary or rebuking improper use all while staying in the object language. For example, if I ask you to imagine two hypothetical circumstances in which all the physical facts about cats are the same but in one of them cats teleport at random, and you tell me that the circumstance in which cats teleport at random is more similar to the actual circumstances, then your false object language claim about comparative similarity reflects that you are not a competent user of the term 'cat'.

To begin clarifying these thoughts, I will use Brandom's illustration of keeping score in a game of giving and asking for reasons as an example. Suppose that a player's total score in this linguistic game is given by those sentences that she is entitled to as well as those sentences she is committed to.[123] Some of the sentences in a player's score will reflect what other sentences are even up for play at all, and these sentences will often involve modal expressions. For example, imagine a context in which I attempt to endorse a claim to the effect of '*necessarily*, *p*', i.e. I attempt to add the claim '*necessarily*, *p*' to my *linguistic scorecard*, and this reflects that '*not*, *p*' is *not* up for play. So, for example, I might say, "I *must* leave work early if I am to make it to the bank on time and deposit my check", which reflects that 'I will stay at work until 5 and deposit my check on time' is not up for play. However, someone can challenge my score, by saying something to the effect of '*possibly*, *not-p*', e.g. they might say, "You *could* stay at work until 5 PM and still deposit your check on time."[124] Now, I may ask what entitles them to that claim by

[123] See Brandom (2000), pp. 190-191.
[124] See Lewis (1983) for more discussion of relative modalities in language games.

asking the following question: "But *how* could I do that?" They may say, "You *could* deposit your check online *or* you *could* have your partner come pick up the check and deposit it for you *or* you *could* hire a certified courier to deposit the check for you *or* you *could* live in world with teletransportation *or* you *could* live in a society such that full-service banks are open 24/7 *or* you *could* …etc." Of course, I will not find just any of the disjuncts as entitling my interlocutor her more general claim about what is possible; but, suppose that I accept at least the first one. If that is the case, then I will not be able to add my original modal claim to my scorecard but, instead, add to my scorecard the claim 'I *can* stay at work until 5 and deposit my check on time'. Furthermore, I will argue below, this addition in turn justifies adding to my scorecard the claim '*There is a possibility* that I will stay at work until 5 PM and still deposit my check on time, say, by depositing my check online'.

But first, let me point out a number of important things going on in this example. More generally, I attempted to add a claim about what is necessary (and, by implication, what is not possible, i.e. what is off limits) to my scorecard, and that was challenged by my interlocutor with a claim about what is possible (what is up for grabs). I then challenged her entitlement to that claim by asking her *how* it was possible, to which she offered a number of *distinct possibilities* to justify her more general claim about what is possible. What's more, note that it seems that not all of the distinct possibilities she raised were acceptable, likely because not all of the possibilities were relevant given the context. So, we started with more coarse-grained claims about what is possible, then moved to more fine-grained claims about distinct possibilities. Then, if we press on why some of the possibilities offered in the example were relevant but not others, it seems that

we rely on judgments about the ways in which the possibilities listed are more or less similar to actuality, e.g. "A society with teletransportation (much less 24/7 full-service banking!) would be nice, but our society is *nothing like that*!"

In general, one reason for the normativist to introduce talk of distinct possibilities and distinct possible circumstances is that following. The modal normativist should not, and need not, hesitate to agree with Lewis (1986) in that if *p* is possible, there are many *ways* in which it is possible that *p*, and that it seems that we can *count* the many ways in which it is possible that *p* as well as qualitatively *compare* them.[125] For example, suppose it is true that: *possibly, Tabby the cat is on a mat*. Now, Tabby could be on a mat in the living room, Tabby could be on a mat in the bathroom, Tabby could be on a mat in the Oval Office in the White House, Tabby could be on a mat in a space station on Mars, Tabby could be on a flying mat, Tabby could be on a blue mat, a red mat, etc. So, that's at least *six* different possible ways Tabby could be on a mat (and maybe one impossible way). And all of these distinct ways can be interpreted as each expressing a potential permission of the application of 'Tabby' while remaining in the object language.

Since counting and, as I will explain in more detail below, comparing possibilities is useful, it only seems natural to go ahead and start quantifying over possibilities. However, this need not mean we are making substantial ontological commitments—the sort of ontological commitments that might scare away those inclined to accept the expressivist element—either because we are not making any *genuine* ontological

[125] Lewis (1986). See also Yablo (1996) for how possible world talk falls out of *how* questions, like the one asked in the previous example, and certain pretenses.

commitments[126] or because we are not making *substantial* ontological commitments.[127] I will look at both of these in a bit more detail in Chapter 6. However, let me briefly explain the latter option to motivate the thought that the modal normativist need not fear talk of worlds, and so there is no tension in my appealing to worlds when giving a normativist account of counterfactuals and counterpossibles.

We can take talk of the existence of a *possibility*, talk of the existence of a *way* things actually are or *could* be, or even talk of the existence of a *possible world,* which is a possibility plus a complete range of other facts, to be a hypostatization out of talk of what is possible. In other words, talk of the existence of possibilities and the like is underwritten by taking it to be a conceptual truth that *if p is possible, there is a possibility in which p*.[128] I would even go so far as to say that someone who says, "*p* is possible but there is no possibility at which *p*" is guilty of a massive conceptual confusion. So, talk of possibilities that allows us to count and compare them, trivially falls out of our talk about what is possible. Thus, so long as we do not then take these possibilities to be causally efficacious or take them to serve as truthmakers for basic modal claims or counterfactuals, there is no harm talking about the existence of possibilities.[129]

What's more, given talk of distinct possibilities, we can ask which of a number of considered possibilities are more or less similar to how things actually are. For example, perhaps the possibility of Tabby being on a mat in the living room is more similar to how things actually are than the possibility of Tabby being on a mat in the Oval Office of the

---

[126] See Rosen (1990) and Yablo (1996).
[127] See Thomasson (2007b, 2015) and Steinberg (2012).
[128] See Thomasson (2007b) and Steinberg (2012) for a more detailed defense.
[129] Alternatively, if one wanted one could say that if *p* is possible then, according to a fiction of possible worlds, there is a possible world in which *p*. See Rosen (1990), Yablo (1996) and Kim and Waslen (2006).

White House because Tabby is actually on a sofa in the living room of a house in Canberra, Australia. These comparison judgments should mirror our judgments about just how different the way things actually are *would* have to be, just how much the way things actually are *would* need to change, in order for the different possibilities to come about.[130]

The reason similarity judgments reflect our counterfactual judgments about what *would* happen *if*, is that both kinds of modal claims reflect our competency with ordinary non-modal vocabulary. Recall from above the normativist view that object language persistence conditions, stated in terms of what would happen if, illustrate, endorse, or express the frame-level application and coapplication conditions of non-modal terms like 'Tabby'. On related inferentialist views, these counterfactual judgments reflect the meaning constituting material inferential content of ordinary empirical claims about cats like Tabby. For example, Brandom (2008) summarizes an aspect of what he calls the Kant-Sellars thesis as follows:

> One has not grasped the concept cat unless one knows that it would still be possible for the cat to be on the mat if the lighting had been *slightly different*, but not if all life on earth had been extinguished by an asteroid-strike.[131] (97)

The way things actually are, with Tabby comfortably on a couch in Canberra, Australia, would need to change considerably more in order for Tabby to be on a mat in the Oval Office in the White House than for Tabby to be on a mat in the living room. But we can

---

[130] In Chapter 6, I will argue that similar considerations discussed in the previous paragraph also apply to talking about distinct impossibilities. Since, I think that what really matters to giving an account of counterpossibles is giving an account of metaphysical similarity, I will do that in the next chapter.

[131] My emphasis. Brandom also seems to suggest that the practice of updating our linguistic scorecards requires associating with each claim a *range* of counterfactual robustness, which requires the capacity to ignore some collateral information that is not relevant for the given range determined by a context. He then discusses this ability as involving the more general ability to find some aspects of *similarity* between objects as more, or less, relevant than others. He seems to think that these are qualitative judgments about relevant similarity that provide limitations on what sort of discursive abilities we can reasonably hope to attribute to Artificial Intelligence. See Brandom (2008), pp. 74-81.

make the same point in terms of worlds and similarity: the *possibility* of Tabby being on a mat in the Oval Office in the White House is *less similar* to how things actually are than the *possibility* of Tabby being on a mat in the living room.

So, these similarity judgments reflect our linguistic competency with the empirical claim 'Tabby is on the couch' as well as reflects the various commitments and entitlements we take on when we undertake or make the empirical claim 'Tabby is on the couch'. Either way, if we endorse the empirical claim 'Tabby is on the couch' we are committed to the comparative claim 'The possibility that Tabby is on a mat in the living room is more relevantly similar to how things actually are than the possibility of Tabby being on a mat in the Oval Office in the White House'. Presumably, someone who thinks differently would have strange beliefs about the world underwritten by either empirically false beliefs about cats or a deviant understanding of the concept 'cat', e.g., maybe they think that cats are *sorts* of creatures intrinsically capable of teleportation.[132] In that case, we might think that someone is not competently using, or entitled to, the claim 'Tabby is on the couch' because they fail to recognize their contextual commitment to Tabby being on a mat *in the living room* given the supposition that Tabby is on a mat at all and not much else has changed about the world, i.e., their commitment to the counterfactual 'if Tabby were on a mat, Tabby *would* be on a mat in the living room'.

So, similarity judgments play an important role in properly using ordinary empirical expressions, and similarity claims are simply another way of conveying rules and permissions regarding the use of ordinary empirical expressions while staying in the

[132] In a metalanguage they associate the term 'cat' with the sortal term 'things that teleport'.

object language. This is because these claims mirror a range of counterfactual robustness of certain inferences, which, according to the expressivist element, reflects the ability to properly use ordinary non-modal expressions. Brandom (2008) puts the point as follows:

> …we cannot intelligibly describe someone as deploying a concept … unless he treats the [meaning constituting] material inferences he takes to be good as having a certain *range of counterfactual robustness*, that is, as remaining good under various merely hypothetical circumstances. One grasps the claim ‘‘the lioness is hungry’’ only insofar as one takes it to have various consequences (which *would* be true if it *were* true) and to rule out some others (which *would not* be true if it *were* true). And it is not intelligible that one should endorse as materially good an inference involving it, such as the inference from ‘‘the lioness is hungry’’ to ‘‘nearby prey animals visible to and accessible by the lioness are in danger of being eaten,’’ but be disposed to make no distinction at all between collateral premises that would, and those that would not, if true infirm the inference. (105)

So, understanding the statement ‘the lioness is hungry’ is partly constituted by grasping the meaning constituting inferential relationships the statement shares with other statements, e.g. by endorsing an inference from ‘the lioness is hungry’ to ‘nearby prey are in danger’. But, *more importantly*, properly deploying a concept is *also* constituted by grasping what kinds of collateral hypothetical assumptions will respect or infirm those meaning constituting inferential relationships. For example, understanding ‘the lioness is hungry’ is constituted by understanding that the inference from ‘the lioness is hungry’ to ‘nearby prey are in danger’ will still go through whether or not we assume that the day is Tuesday, that Hillary is running for President, or that I ate oatmeal for breakfast, but the inference will *not* go through on the assumptions that the lioness has been shot with a tranquilizer, or is in a cage, or that lions are vegetarians.[133] So, while according to the expressvist element, counterfactuals such as:

[133] Ibid.

(Lioness) If the lioness were hungry, nearby prey would be in danger

are object language expressions of a requirement to accept 'nearby prey are in danger' on condition of 'the lioness is hungry', the expressivist element also captures the thought that grasping ordinary empirical expressions such as 'the lioness is hungry' *also* importantly involves understanding the *contexts* in which (Lioness) is assertible, which is to associate the counterfactual with a range of robustness, and this I have argued can safely illustrated, endorsed, or expressed in terms of similarity judgments about possible worlds.

I have just argued that considerations of similarity and worlds are just object language resources used to convey the proper use of ordinary non-modal vocabulary. So, there is no awkwardness in my relying on the expressivist element:

> A counterfactual 'If *p* were the case, then q *would* be the case' *expresses* a *requirement* to accept that *q* given *p* as a hypothetical supposition and a range of *relevant auxiliary assumptions* determined by the context. (*Expressive Element*)

and the similarity element:

> A counterfactual 'If *p* were the case, then *q would* be the case' is *assertible* in *w* just in case there is some *p q* world that more relevantly similar to *w* than every *p* non-*q* world. (*Similarity Element*).

This is because, while not used as truthmakers for modal claims, talk of worlds and similarity still naturally falls out of a use-based account of meaning found in the work of Thomasson, Sellars, and Brandom. In turn, this is because talk of possibilities falls naturally out of talk of what is possible, and because similarity judgments reflect aspects

of the counterfactual robustness of certain inferences in a given context, which reflects the ability to properly use ordinary non-modal expressions. Moreover, we need not worry about taking on substantive ontological commitments by talking about and quantifying over possibilities or possible worlds.

In Chapter 6, I will argue that talk about impossibilities naturally falls out of talk about what is impossible, which is just to talk about what is not possible. However, while it is not possible for me to be a non-human animal, there are many distinct ways in which it is not possible: I cannot be an orangutan, I cannot be a dolphin, I cannot be an anhinga, etc. However, what I want to focus on in the remainder of this chapter, and what is a critical part of my account of counterpossibles for modal normativism, is how to go about comparing these distinct impossibilities with the actual world in terms of *metaphysical similarity*. Given my account of metaphysical similarity, I will then have completed my normativist account of counterpossibles.

3.3 Counterpossibles for Modal Normativists

In what follows, I could rely on *Similarity Element* for the presentation of my account of counterfactuals and counterpossibles:

> A counterfactual 'If *p* were the case, then *q would* be the case' is *assertible* in *w* just in case there is some *p q* world that more *relevantly similar* to w than every *p* non-*q* world. (*Similarity Element*)

However, this would not be the most ideal presentation because it is potentially misleading. For one, in the last section I just argued that:

> 'There is some *p q* world that is more relevantly similar to *w* than every *p* non-*q* world' *expresses* a *requirement* to accept that *q* given *p* and a range of *relevant auxiliary assumptions* determined by the context. (*Worldwide Expressive Element*)

So, *Similarity Element* really just comes down to:

> A counterfactual 'If *p* were the case, then q *would* be the case' is *assertible* in a context just in case 'If *p* were the case, then q *would* be the case' *expresses* a *requirement* to accept that *q* given *p* as a hypothetical supposition and a range of *relevant auxiliary assumptions* determined by the context.

In some ways, this is exactly what I want since my overall claim is that we are not describing anything by asserting counterfactuals and counterpossibles, we are expressing requirements. But, I prefer to work with the language of the extended Lewis-Stalnaker framework. So instead, so I will use this presentation:

> A counterfactual 'If *p* were the case, then *q* would be the case' *expresses a requirement* in *w* to accept *q* given *p* as a hypothetical supposition just in case there is some *p q* world that more relevantly similar to *w* than every *p* non-*q* world. (*Counterpossibles for Modal Normativists*)[134]

Given *Worldwide Expressive Element*, this latter presentation comes down to:

> A counterfactual 'If *p* were the case, then *q* would be the case' expresses a requirement, in a context, to accept *q* given *p* as a hypothetical supposition just in

[134] See chapter 2. For the details of these views, see especially Nolan (1997), and also Kim & Maslen (2006), Krakauer (2012), Brogaard and Salerno (2013), and Kment (2014).

> case *there is* a requirement to accept that *q* given *p* as a hypothetical supposition and a range of relevant auxiliary assumptions determined by that context.

The right side of the biconditional states that there is a requirement to infer '*q*' from '*p*', i.e. where '*p*' and '*q*' are mentioned. This presentation is an explicit metalinguistic and prescriptive (rather than descriptive) account of counterfactuals.

However, since I take talk of similarity between worlds in *Counterpossibles for Modal Normativists* to be object language expressions of the normative force of inferential relations between *p* and *q* given a range of relevant auxiliary assumptions, I will stick with talk of worlds and similarity. Very roughly, considering both possible and impossible worlds, if we are going to *hypothetically* accept that the world is some way other than it actually is, then we can think of *Counterpossibles for Modal Normativists* as outlining a requirement about what kinds of worlds we ought to accept: if you're going to hypothetically accept a *p* world, it *ought* to also be a *q* world. But, given the arguments of the previous section, *Counterpossibles for Modal Normativists* grants that another way of making the same point is to say that *p q* worlds are more relevantly similar to our world than every *p* non-*q* world, but where claims about worlds and similarity are not primarily descriptive but, instead, are expressive claims that illustrate the actual constitutive rules that govern the use of '*p*' and '*q*' along with additional relevant conceptual or empirical information being held fixed by the context. So, my account respects other extended Lewis-Stalnaker accounts of counterfactuals and counterpossibles by keeping talk of similarity and worlds *simpliciter*, which allows for talk of both possible and impossible worlds.

However, it is important to stress three important ways in which this account is significantly different from alternatives. The first difference is that I am giving my account in *expressivist* terms. Also, my theory of counterpossibles respects the normative aspect of modal normativism, that is, my theory is consistent with the claim that what we are expressing with object language modal claims are rules and permissions that govern the use of non-modal vocabulary.[135]

Second, I adopt a deflationary view of truth and reject truthmakers for modal claims. So, I am not claiming that facts about worlds and similarity make counterfactuals and counterpossibles true. However, I am not claiming that counterpossibles describe, or are made true by, facts about the inferential, conceptual, or circumstantial relationships between *p* and *q*. Still, since I am adopting a deflationary view of truth, it is important to emphasize that nothing prevents me from saying that counterfactuals and counterpossibles are truth apt.[136] In particular, according to my theory, some counterpossibles are false. Thus, I will continue to speak of some counterpossibles as true and others as false.

---

[135] It is important to note that the prescriptive element of normativism need not commit us to any substantive views about normativity itself or the normativity of concepts. For example, it need not commit us to the view that people must only apply 'bachelor' to unmarried persons *because* the concept BACHELOR has intrinsic normative properties, or that people regularly apply 'bachelor' to unmarried persons *because* the concept BACHELOR has some rational properties that somehow compel them to do so. Alternatively, we can simply say that the rules of use that constitute the meaning of terms are simply generalizations of the regular patterns of use of those terms in real or imagined cases. Expressing the regular patterns of use in terms of normative rules merely functions to further reify the current patterns of use or their use in imagined cases, and overall this serves to coordinate linguistic usage (cf. Lewis 1969). This latter option only requires a very thin sense of normativity and does not require a substantive theory of concepts. Regardless of which option we go with, the prescriptive element in my account of counterfactuals keeps the account consistent with the normativist view of metaphysical modality and makes it more than a mere expressivist account. See Kim & Maslen (2006) for another prescriptive account of counterpossibles.

[136] See Section 3 above.

The third, and main, difference between my account of counterfactuals and counterpossibles and alternatives is in how I account for metaphysical similarity between worlds, which appeals to a *conceptual criterion*.[137] Roughly, for all counterfactual conditionals, i.e. those with possible or impossible antecedents, and given a particular context, what matters to determining metaphysical similarity is neither the epistemic perspective of an agent nor substantive notions of essence, grounding or metaphysical structure but tacit understanding of what I call *conceptual laws*. As I will argue in the next section, this approach to metaphysical similarity and non-vacuous counterpossibles comes with clear advantages over the epistemic and metaphysical approaches discussed in the previous chapter.

3.4 Metaphysical Laws for Modal Normativists

On my account, metaphysical similarity is often determined by the metaphysical laws. However, I will now argue that *metaphysical laws are just object language expressions of conceptual laws*. As such, talk of metaphysical similarity between impossible worlds is just an object language resource used to convey the proper use of ordinary non-modal vocabulary when the rules governing the use of ordinary non-modal vocabulary are changed in small yet important ways. I will argue for this claim by way of introducing the ideas of metaphysical explanation and laws. Not only is this needed to complete my account of counterpossibles by giving an account of metaphysical similarity, but it will set the context needed for understanding the worry for my view I consider in the next chapter.

[137] See the end of section 2.1 above.

First, let me also briefly argue why my overall account of counterpossibles and metaphysical laws matters. This overall project is important because there is reason to think that metaphysics has moved from relying on primarily *intensional resources*—e.g. talk of metaphysical necessity, supervenience, counterfactuals, possible worlds, etc.—to relying heavily on *hyperintensional resources*—e.g. talk of essences, grounding, counterpossibles, impossible worlds, etc.[138] By 'hyperintensional resources' I mean sentence constructions that do not guarantee the substitution of necessarily coextensive expressions. For example: while it is true that, *necessarily*, 0 is divisible by 2 without remainder just in case 25 is divisible by 5 without remainder; and it is true that 0 is an even number *in virtue of* 0 being divisible by 2 without remainder; it seems *false* to say that 0 is an even number *in virtue of* 25 being divisible by 5 without remainder.

Of course, the use of intensional resources raises questions about what ontological commitments we are required to make in order accommodate our talk of metaphysical necessity, possibility, possible worlds, etc., as well as epistemological questions about how we know these modal claims. So, the same questions are raised when we start using hyperintensional resources, i.e. when we start talking about essences, real definitions, grounding, counterpossibles, etc. So, one accomplishment of this overall project is showing how modal normativism can accommodate the hyperintensional turn, and I do this by giving a general normativist strategy for accommodating talk of counterpossibles in terms of essences, real definitions, and metaphysical dependence.[139] It follows from

---

[138] See Nolan 2014.

[139] I will not work with any particular conception of essence, real definition, or metaphysical dependence. For example, Fine (1994b) outlines a number of alternative ways to think about essence. See Rosen (2015) for an interesting defense of real definition, the relation between real definition and essence, and an account of real definition in terms of grounding. The literature on grounding and metaphysical dependence is vast. My aim here is to provide a *general* normativist strategy for using these hyperintensional resources that

my view that we come to know about counterpossibles, essences, real definitions, and metaphysical dependence primarily through conceptual analysis: (i) by analyzing the concepts expressed in our "home" language; and (ii) sometimes analyzing languages that are relevantly similar to our "home" language. So, a major payoff of my normativist account of these hyperintensional resources is that it yields a plausible epistemological story about how we come to know certain counterpossibles, metaphysical laws, and understanding metaphysical explanations.

Schaffer (2017) nicely outlines metaphysical explanations as follows. *Metaphysical explanations* are *explanations* in the sense that they do not merely attempt to describe what is metaphysically case but explain *why* something is metaphysically the case; metaphysical explanations are *metaphysical* insofar as they seem to be about the non-causal "constitutive generation of a dependent outcome".[140] I take common explanatory locutions to include 'A *because* B' and 'B *explains* A'. Here is an example of a potential metaphysical explanation:

> The singleton set {Socrates} exists *because of* the existence of Socrates. (SINGLETON SOCRATES)

Schaffer (2017) takes *metaphysical laws* are general principles that seemingly provide a non-causal link between the *explanans* of a metaphysical explanation and the *explanandum*, and which supports a range of counterfactual robustness between the *explanans* and *explanandum*.[141]

---

does not incur substantive ontological commitments. Further work will need to be done to give more detailed accounts of each hyperintensional resource and how they are related.

[140] Schaffer (2017), p. 303.

[141] Ibid., p. 305.

Some of the metaphysical laws will include *general* principles about mereological composition and set formation, among others. For example,

> For any entities found up to stage *n*, there is a set at stage *n* + 1 that contains all an only the entities found up to stage *n*. (SET LAW)

However, Kment (2014) takes the metaphysical laws that play a role in metaphysical explanations to not only include general principles, but also truths about the essence or real definition of an entity, given the existence of that entity, or truths about particular relations of metaphysical dependence.[142] So, in addition to general principles, I will take common law locutions to include: A *depends on* B, A *is grounded by* B, B *is essential to* A, *what it is to be* A *is to be* B, etc.[143] Given this, here are some more examples of potential metaphysical laws:

> Given the existence of the set {Socrates}, *what it is* for the set {Socrates} to exist is for Socrates to exist. ({SOCRATES} LAW)
>
> Given the existence of Socrates, *what it is* for Socrates to exist is for Socrates to originate from his actual biological parents. (SOCRATES LAW)

---

[142] See Kment (2014), pp. 159-173. See also Kment (2006).

[143] Alternatively, I am fine with saying that I am here giving an account of *instances of metaphysical laws*, e.g. 'Socrates is essentially human'. The explanation I give for that particular claim can be generalized to explain potential general principles such as 'For all x, if x is a person then x is essentially human', which I would argue is an object language schema that serves the prescriptive, rather than descriptive, function of endorsing general rules, possibly found in a metalanguage, that governs the use of names. Also, I will not address Schaffer's (2017) particular argument for a general view about metaphysical laws that represents them with mathematical functions and entails that the notion of essence is explanatorily inert (though he still thinks that the notion of metaphysical dependence matters to metaphysical laws). Instead, I will focus on examples found in Fine (1994, 2012), Rosen (2006, 2015), and Kment (2006, 2014), which tend to involve essence, real definition, or metaphysical dependence, and I will provide a general normativist strategy for accommodating the use of these claims that does not take on substantive metaphysical commitments.

So, when someone asks for a metaphysical explanation for the existence of the set {Socrates}, we first provide them with the grounds, the existence of Socrates, and then connect the grounds to the existence of the set {Socrates} using ({SOCRATES} LAW) and (SET LAW).[144]

While I will argue that the metaphysical laws are important when evaluating counterpossibles, they also play an important role in metaphysical explanations. This is not the place to offer a full account of metaphysical explanations for modal normativists but let me offer the following sketch of a non-descriptivist account of metaphysical explanations such as (SINGLETON SOCRATES). Basically, metaphysical explanations are object language expressions of our *entitlements* to make or endorse certain claims. To be entitled to endorse a claim *q* is to be able to produce an adequate reason for making that claim. So, for example, when I claim, "The set {Socrates} exists" I must be prepared to say what entitles me to make that claim. For example, someone may ask, "Why?", to which I will respond "The set {Socrates} exists *because* Socrates exists." What this metaphysical explanation reflects is that, given the existence of Socrates, what entitles me to use the expression 'the set {Socrates} exists' is that the application condition 'the set {Socrates} exists' consists in the application conditions for 'Socrates exists' being met, along with general claims about set formation. So, when I make the claim "The set {Socrates} exists *because* Socrates exists" I am not making a claim in any way metaphysically analogous with the claim "The window is broken *because* someone

[144] Ibid. See also Rosen (2006).

smashed it with a brick." Instead, I am using an object language claim to express an entitlement to use the expression in the way I am using it given the context we are in.

Unfortunately, basic modal vocabulary is not enough to do all of the interesting metaphysics we want to do. Fine (1994) famously argues that essence cannot be reduced to modality. For example, granting that, *necessarily*, the set {Socrates} exists iff Socrates exists, it seems *false* to say that, while the existence of Socrates is essential to the existence of the set {Socrates}, the existence of the set {Socrates} is essential to the existence of Socrates. If Fine's arguments are correct, and we want claims of essence to serve as metaphysical laws that feature in metaphysical explanations like (SINGLETON SOCRATES), we need an account of essence that is more fine-grained than metaphysical necessity. I will now argue that, for the normativist, the metaphysical laws just are hyperintensional object-language expressions of certain conceptual rules and reflect important asymmetries in the application and coapplication conditions of non-modal vocabulary.

Thomasson (2007b) develops modal normativism from the thought that modal expressions such as '*must*' are effective ways of stating the constitutive rules of a game in the indicative instead of the imperative mood.[145] For example, often a more effective way of stating the constitutive rules of chess is to use an indicative claim such as 'the pawn *must* never move backwards' as opposed to 'do not move the pawn backwards!'. So, let me motivate my general normativist strategy to account for metaphysical laws by giving a toy illustration of the role claims of essence and dependence can play in explanations

[145] Thomasson (2007b), p. 138.

about a game such as chess. In chess, c*astling* is a move where a player moves the king two squares towards a rook and then moves the rook to the square over which the king has crossed. According to a standard rule of chess, you are not allowed to castle after you have already moved the king or rook. According to another standard rule, you are not allowed to move a pawn backwards. When talking about chess in general, one might appropriately say, "You *must neither* move a pawn backwards *nor* castle after you move the king." However, even supposing that both rules are constitutive rules, i.e. are both *necessary* in a game of chess, we cannot freely substitute one statement of rule for another when offering certain explanations about the game. While it is true to say both, "*Part of what it is to be* a pawn is that you not move a pawn backwards" and "Castling *depends* on not moving your king beforehand", it is *false* to say, "*Part of what it is to be* a pawn is that you not move the king before you castle" and *false* to say "Castling *depends* on not moving your pawn backwards".

One reason to think that the former claim is false is the following. Imagine that two chess players are playing either a chess game in which castling is not permitted or a deviant game where a player can more her king before castling. Either way, these changes in the standard rules of chess make no difference to how a player is permitted to move the pawn. Thus, not moving the king before castling is not part of what it is to be a pawn. What this example shows is that linguistic constructions such as 'A *depends on* B' and '*what it is to be* A *is to be* B' can be used to make important distinctions amongst the rules of chess when offering certain explanations.

Let the *conceptual laws* be the constitutive rules of ordinary non-modal vocabulary, which may be stateable in a metalanguage. The conceptual laws governing

the use of an expression 'S' at least include: the actual *application conditions* for 'S', which are constitutive rules that state the conditions under which a term is successfully applied; the actual *coapplciaiton conditions* for 'S', which are constitutive rules that state when a term is applied to one and the same thing, given that the term has been successfully applied twice; and the actual *analytic entailment relations* between 'S' and other expressions of the language, where *p* analytically entails *q* just in case the satisfaction of the application conditions for *p* guarantees the satisfaction of the application conditions of *q*, e.g. 'the ball is red' analytically entails 'the ball is colored'.[146] Again, conforming to the conceptual laws does not require ascending to a metalanguage and explicitly stating them—neither does endorsing, illustrating, or expressing the conceptual laws using the object language. In fact, it may in fact be much more useful to endorse conceptual laws using object level claims.

Some conceptual laws, e.g. 'Apply 'Socrates' to *x* only when 'human' applies', can be endorsed or illustrated in the object language using basic modal claims, e.g. 'Socrates is *necessarily* human'. But I claim that *certain conceptual laws*, or their consequences, found in a metalanguage, *are best expressed* in the object language with statements using the terms with certain semantic and formal properties. For example, sentences constructed with '*essential*' or '*metaphysically depends*' are each generally taken to be hyperintensional insofar as they do not permit the substitution of intensionally equivalent expressions. 'Metaphysically depends' is generally taken to be *asymmetric* insofar as if '*x* metaphysically depends on *y*' is true then '*y* metaphysically depends on *x*' is false. I am arguing that the primary reason we need modal terms with these formal

[146] See Thomasson (2015) chapter 2 and (2007a), especially chapters 2 &3, for more on these conditions

properties in our object language is *not* because we need to describe substantive essential properties or substantive relations of metaphysical dependence. Instead, we need these terms because there are often important conceptual asymmetries or conceptual independencies between expressions in our language that bear on the competent and appropriate use of those expressions, and these conceptual asymmetries or conceptual independencies are best illustrated or endorsed in the object language using hyperintensional modal expressions.

For example, suppose that a conceptual law governing the use of the expression 'the set {Socrates}' is the following rule stated in a metalanguage: Apply 'the set {Socrates}' when and only when 'Socrates' applies. This prescriptive rule of use can be expressed in the object language using the expression: *necessarily*, the set {Socrates} exists iff Socrates exists. However, this latter modal claim fails to make an important distinction amongst the actual rules of use that govern the use of the expression 'Socrates' and 'the set {Socrates}'—namely, that the application conditions that govern 'Socrates' must be satisfied in order for the application conditions for 'the set {Socrates}' to be satisfied, but not vice versa. Long before the development of set theory, people were able to competently use the term 'Socrates' to talk about the teacher of Plato. So, the claims 'the set {Socrates} *metaphysically depends* on Socrates' and 'Socrates does not *metaphysically depend* on the set {Socrates}' are object level claims about metaphysical dependence and independence that illustrate this important conceptual asymmetry.[147]

[147] One might worry about the difference in force between the metalinguistic claim 'Apply 'Socrates' to x only when 'human' applies' and the metalinguistic claim 'The rules of use that govern 'Socrates' play an important role in the rules of use that govern 'the set {Socrates}' but not vice versa'. The former sounds

Overall, this is a very natural approach to claims of essence, real definition, metaphysical dependence given that modal normativism already takes all claims of metaphysical necessity to be analytic or derived from an analytic truth with an empirical truth.[148] Fine (1994) argues that, just as necessity is not a good guide to essence, analyticity is *not* a good guide to meaning. For example, Fine argues 'something is a bachelor iff it is an unmarried man, and all triangles are three-sided' is analytic *simpliciter*—i.e., true in virtue of the meaning of *all* the terms involved—but plays no part in explaining the meaning of 'bachelor'.[149] Fine (1994) takes arguments like this to

> suggest that even when all questions of analyticity have been resolved, real issues as to their source will still remain. The study of semantics is no more exhausted by claims of analyticity than is the metaphysics of identity exhausted by claims of necessity. (12)

Instead, we need to look to more fine-grained aspects of definition to explain meaning, and *these* aspects are what *explain* the analyticity of certain truths, e.g. 'all bachelors are unmarried males'. For the normativist, these fine-grained aspects of definition will be connected to the central rules that govern the use of ordinary language.

However, Fine's overall suggestion is that there is an analogous sense of *real definition* at play in metaphysics, and that modal claims are not a good guide to the real definition of an object. While 'necessarily, something is a bachelor iff it is an unmarried man and all triangles are three-sided' is true, triangles have nothing to do with *what it is*

---

like a command, and hence prescriptive, while the latter sounds descriptive. So, the worry continues, why think that the object level claim of metaphysical dependence expresses anything prescriptive. One way to argue that the object level claim of dependence still has normative force is to assume that, in general, claims of metaphysical dependence entail claims of necessity. If a particular ball is colored in virtue of its being red, then the ball is necessarily colored if it is red. Now assume that we cannot derive prescriptive claims strictly from descriptive claims. Then we can say that the object level claim of metaphysical dependence has implicit normative force that comes out explicitly in the entailed modal claim.

[148] Thomasson (2007a), p. 64.

[149] See Fine (1994), p. 11.

to be a bachelor.[150] Instead, to find the real definition of the property of being a bachelor we must look to the essence of bachelorhood or what it metaphysically depends on. Rosen (2015) claims that

> when we ask what it is for a thing to be a person or for a creature to be conscious or for a fact to be a law of nature or for two expressions to be synonymous or for an object to be colored or for an action to be free or for an artifact to be an artwork, we are best understood as seeking real definitions of the properties, kinds, and relations that figure in our questions, rather than semantic or conceptual equivalents, even when the correctness of the account is meant to be recognizable a priori. (189)

But, I do think that non-metaphorical talk of *things* having definitions is making a category mistake. More importantly, given that conceptual analysis and straightforward empirical inquiry are reliable sources of knowledge, thinking that there are *non-conceptual* and *non-causal* essences or relations of metaphysical dependence in the world raises serious worries about how we can reliably know anything about them. If we cannot appeal to conceptual analysis and straightforward empirical inquiry to answer questions about the metaphysical laws, then we have good reason to think that the metaphysical laws are outside of our cognitive reach. But in that case, we must either be skeptics about the metaphysical laws or we must accept that which of our beliefs about the metaphysical laws are true is a matter of sheer luck.[151] Either way our theoretical reasons for talking about the metaphysical laws to begin with will be undermined if we take the metaphysical laws to be non-conceptual and non-causal/empirical.

Alternatively, on my account of metaphysical laws, all that is meant by 'real' in '*real* definitions' is that we are *using*, as opposed to *mentioning*, terms to express fine-

[150] Ibid. Also see Rosen (2015) and Kment (2014), pp. 157-158.
[151] See Rayo (2015) and Warren (2016).

grained requirements of language use found in a metalanguage. On my account, all it is to know that facts about the set {Socrates} are in no way part of what it is to be Socrates is simply to be a competent user of the term 'Socrates'. Likewise, for claims of metaphysical dependence. So, on my account, the claim 'the ball being colored *metaphysically depends on* its being red but not its being spherical' is not describing something about the way reality is non-causally structured, but is expressing a normative ordering of rules governing the terms 'color' and 'blue', e.g. that one is not entitled to infer 'the ball is colored' from 'the ball is spherical' or that one is not entitled to infer 'the ball is blue' from 'the ball is colored'. Since claims about essence and metaphysical dependence, the so-called metaphysical laws, just are object language expressions of certain conceptual laws, this account more easily avoids worries raised by taking talk of real definitions too seriously. However, before moving on, I need to make some clarifications, which should also alleviate certain worries one might have about my view.

First, I am not claiming that claims about essence, real definitions, grounding and the like are in any way *about* our definitional conventions instead of being about the world. On my view, claims of essence and metaphysical dependence are stated in the object language where terms are used and not mentioned—so, these claims are about the world. For example, the claim 'what it is to be Socrates in no way depends on the existence of the set {Socrates}' is a claim about Socrates and the set {Socrates}. On my view, the claim is not about the rules of use that govern the use of the relevant terms since those rules are stated in a metalanguage where the terms are mentioned and not used. However, on my view, the claim does endorse or illustrate a rule of use governing the use of 'Socrates' by *using* the term. So, there is a sense in which real definitions and

the like can be seen as *defining the term by using the term*. What's more, for my normativist account, these object language definitions have normative force to the effect that one ought to only make object language claims about Socrates that are consistent with the real definition given in the object language.

Second, on my view: it is not the case that on my view the metaphysical laws are made true by conventions or our mental activities; so, it is not the case that the metaphysical laws *could* have been different had we adopted different conventions. On my view, the metaphysical laws are not contingent—they are indeed metaphysically necessary. What is true on my view is that our linguistic practices do play some role in fixing the meaning of object language claims that use terms such as 'Socrates' and 'the set {Socrates}', including claims of essence, real definition, or metaphysical dependence. But that the use of our terms plays a role in determining the meaning of those terms is relatively uncontroversial and is in no way the same as claiming that our linguistic practices make the propositions expressed by sentences true.[152] What's more, once the meanings of the various terms that appear in the metaphysical laws are fixed by our actual practices, then the metaphysical laws will be true at every *possible* world, including possible worlds with no people, possible worlds with no language, or *possible* worlds inhabited by people but that speak deviant languages.

Thus, in effect, for the normativist, *the actual metaphysical laws just are hyperintensional object-language expressions of the actual conceptual laws*. Furthermore, it is important to stress that the normativist need not take every

[152] See Thomasson (2007a, 2015) for a more thorough defense against this objection. Like Thomasson, I am inclined to accept a deflationary view of truth that rejects the idea that our any of our claims are made true by anything at all.

metaphysical law to be *purely* conceptual. For the normativist, the metaphysical laws—in particular, those that are related to *a posteriori* necessities—often consist in conceptual laws along with what follows from them given relevant empirical information that is "plugged in" to a framework of conceptual laws. For example, for the normativist it is a metaphysical law that *Ruby is essentially human*, which depends both on the frame-level content of the term 'Ruby' as well as certain empirical facts about Ruby, namely that she is a living, conscious, and intelligent animal as opposed to an automaton.[153] However, to avoid confusion with the alternative heavyweight metaphysical view, I will continue to use the expression 'conceptual laws' to refer to the normativist conception of metaphysical laws.

I think that considerations of metaphysical similarity between worlds, of the distinctive kind outlined above, is a matter of making sure we are not deviating too far from the *actual* conceptual laws. When dealing with possible worlds, maximizing conceptual laws is easy, since the actual conceptual laws play a large role in determining what is possible in the first place; so, the object language expressions of the actual conceptual laws are true in all possible worlds. But when dealing with metaphysically impossible worlds, such as when we consider counterpossibles involving the violation of certain essential properties or the existence of impossible things or circumstances, the modal normativist will still try to maximize overall adherence to the actual conceptual laws while minimizing those violations that will trivialize the conditional.

[153] See also section 3 above.

The plausibility of this account of metaphysical similarity and counterpossibles relies on two ideas. First, the idea that an expression in one language, with its meaning constitutive frame-level content or conceptual role in that language, can have semantic or conceptual counterparts in alternative languages or conceptual schemes. What's more, the counterparts can be more or less similar to one another.[154] This way, small changes in the frame-level content or conceptual role of an expression neither result in an entirely distinct expression nor do they result in a complete change of subject. Second, that metalanguage frame-level relations of semantic or conceptual similarity between various expressions in various languages can sometimes be expressed with claims that *use* the expressions and are purportedly about the metaphysical similarity between two impossible things, e.g. similarity with respect to essential properties between two modal counterparts.

Let me further explain my account of metaphysical similarity by working through some examples, and, in Chapter 4.3, I will give a more detailed example. But, before doing so, I want to emphasize an important point in order to avoid confusion about my view. In what follows, I will often explain how particular counterpossibles are evaluated on my view by appealing to an explicit statement of some potential conceptual law stated in a metalanguage. In each case, I am not defending the claim that my example of a conceptual law is an actual conceptual law. Also, I am not claiming that the evaluation of a counterpossible ever takes place in a metalanguage. The evaluation and discussion of counterpossibles generally takes place in the object language the counterpossible is stated

[154] For an interesting and more detailed development of the idea of conceptual or semantic counterparts see Warren (2015). Warren argues that that semantic counterparts "allow us more flexibility and generality when talking of meaning and concepts than a simplistic picture that allows for only semantic identity and non-identity with nothing in between (Warren 2015, p.1366)."

in. Again, in order for a speaker to competently use a term, she need not be able to state the rules that govern the use of that term in a metalanguage. Furthermore, the rules may not be stateable in a metalanguage at all. However, since I am claiming that the object language metaphysical laws just are endorsements of the conceptual laws, it will be more apt to illustrate my view by working through examples in a metalanguage.

The first example I will use to show how this account works is (11) from the end of section 2.1:

(11) If water had not been H_2O, then water would have been a monkey.

Like the epistemic and robust metaphysical options discussed in chapter 2, on the normativist account, (11) comes out false. Assuming an accommodating context, we set aside the background empirical fact that water has the chemical structure H_2O. We tacitly note that even in the impossible situation where (actual) water has the chemical structure of XYZ, the concept 'water' would have no relevant conceptual connection with the concept 'monkey'. In supposing the impossible circumstance where water is XYZ, though the application and coapplication conditions of 'water' would no longer include the factual empirical information that water is actually H_2O, we would not then be free to treat the term 'water' as something that picks out a countable thing that falls under the sortal 'monkey'—monkeys are countable animals while water is a non-living fluid. Furthermore, this discussion and consideration of (11) need not occur in the metalanguage. All of this can be considered and stated in the object language using the relevant expressions as well as talk of worlds and similarity: impossible worlds where water has the chemical structure XYZ are more similar to the actual world than worlds where water is a monkey. If a person were to claim otherwise, they would subject to

correction by competent speakers of the language on the basis of the frame-level content of 'water', which governs its use, i.e. on the basis of the actual conceptual laws.

3.5 Advantages

My account of counterpossibles avoids the worries I raised for the epistemic approach discussed in chapter 2. Recall (13) from section 2.3:

(13) If a steel Penrose triangle were placed in a 4000 deg. F oven, it would melt.

The concern raised against the strict epistemic reading was that (13) doesn't seem to be about how we represent impossible things but is about what would happen to an impossible thing no matter what you call that thing.

On my account of counterpossibles for modal normativists, (13) is indeed about impossible steel things and their melting points. It is true that some conceptual information is important to evaluating (13), namely the application conditions of 'steel Penrose triangle', but this has nothing to do with the fact that steel Penrose triangles are impossible. Consider the following counterfactual:

(13') If a huasadoe were placed in were placed in a 4000 deg. F oven, it would melt.

Unless we have a tacit or explicit understanding of the rules that govern the use of 'huasadoe', which tell us under what circumstances the term would be successfully applied, (13') is a pseudo claim at best or meaningless at worst.[155] Furthermore, this is true whether I stipulate that 'huasadoe' means 'steel tribar' or 'steel Penrose triangle', so

[155] Thomasson (2015), p. 40.

the modal status of the antecedent has no significant bearing on whether (13') is best evaluated using epistemic or circumstantial criteria.

Therefore, on my view, we can allow that whether the application conditions for 'steel Penrose triangle' are met in the relevant impossibility is a circumstantial matter, i.e. empirical matter. Likewise, supposing that the application conditions for 'steel Penrose triangle' are satisfied in the relevant impossibility, whether a given steel Penrose triangle melts in a 4000 deg. F oven is a circumstantial matter. When evaluating (13), we will likely rely primarily on empirical considerations, e.g. the composition of steel, the shape of steel, and the melting point of steel, which are *not* epistemic but circumstantial. So, (13) is about impossible circumstances and not impossible representations.

Furthermore, suppose that 'steel Penrose triangle' and 'Oscar's tribar' are governed by the same rules of use, e.g. have the same application and coapplication conditions, and are involved in the same analytic entailments. Then, on my view, we can freely substitute one term for the other in contexts where the relevant object level expressions of the conceptual laws are held fixed. It does not matter that the antecedent is impossible. However, there may indeed be a context where (13) is referentially opaque, e.g. a context where it is unknown by some agent(s) that 'steel Penrose triangle' and 'Oscar's tribar' mean the same thing (compare 'Hesperus' and 'Phosphorus'). In that case, in order to make sense of the conceptual position of those who are not fully competent with the terms, we may evaluate (13) with respect to their relative position by holding epistemic features of their subjective position fixed. But nothing in my view *requires* that we treat counterpossibles this way, as tracking phenomena about how an

agent represents impossible things. Again, on my view, (13) is easily read as being about steel Penrose triangles and their melting points.

However, my account is different from other non-epistemic approaches in that it thinks something prescriptive, not descriptive, is being expressed by the counterfactual mood of (13). Namely, that in the right context, impossible circumstances in which the application conditions for steel Penrose triangles are impossibly satisfied and those things melt at 4000 deg. F are *required* in that context over circumstances in which the application conditions for steel Penrose triangles are impossibly satisfied and those things don't melt. In other worlds, on my view, if someone were to claim that impossible circumstances in which some existing Penrose triangle is placed in a 4000 deg. F oven but does not melt and, say, instead turns into a rose are more relevantly similar to the actual world than impossible circumstances in which some existing Penrose triangle is placed in a 4000 deg. F oven and does melt, then they would be subject to correction by competent speakers of the language (13) is stated in.[156]

My account of counterpossibles and metaphysical similarity avoids the difficulties encountered by the substantive realist approach. Recall the counterpossible:

[156] Since my account is mostly concerned with accounting for the pragmatic function of counterfactuals, one might wonder what my account has to say about why anyone might care about counterfactuals statements about strange, impossible things, such as (13). I do not think that the normativist must provide a pragmatic explanation for any and every possible utterance of a counterfactual. But I will say that, even in the case of (13), it is entirely plausible to think that someone might utter (13) to illustrate a contextual norm to the effect that facts about the shape of a steel object never entitle us to infer facts about the melting point of steel, that even given a piece of steel shaped like a Penrose triangle, we are required to accept that it will melt when placed in a 4000 deg. F oven. This would also require assuming that a piece of steel being shaped like a Penrose triangle neither requires any changes in the forces that connect of the elements that molecularly compose a steel object, nor the shape of steel molecules, nor atomic composition of the molecules that compose the steel object.

(14) If George Eliot had been Sean Spicer's biological parent, then Sean Spicer would be a much more eloquent speaker.

In line with a substantive realist reading of (14), I pointed out that there is a perfectly good sense in which (14) is not about a scenario in which George Eliot is Sean Spicer's biological parent *for all anyone knows* but is instead about an impossible scenario in which George Eliot is indeed Sean Spicer's biological parent. So, we should not use epistemic criteria of similarity to evaluate (14). But, I then asked if in order to determine similarity between the relevant worlds should we or do we even need to rely on:

(a) circumstantial criteria that appeal to substantive facts about metaphysical essences, metaphysical grounding, metaphysical structure, or metaphysical fundamentality; or

(b) can we get by with more theoretically minimal conceptual criteria?

I then argued that (a) comes with metaphysical and epistemological worries that are best to avoid. My account of counterpossibles and metaphysical similarity, which is more in line with (b), can avoid these worries. Let me explain.

In order to make sense of considerations of metaphysical similarity needed to evaluate (14), we can simply rely on the role object language claims of essence play in coordinating our discourse by illustrating, endorsing, or expressing semantic features of the terms in those essentialist claims, such as rules and permissions that govern the use of 'Sean Spicer'. Recall that, on the normativist account of similarity I am offering, so-called metaphysical laws just are hyperintensional object-language expressions of the actual conceptual laws. The normativist agrees that, in order for the antecedent of (14) to

hold, there needs to be a change in some fact about actual essential properties, and, assuming non-contingent essentialism, this means we are dealing with a metaphysical impossibility. However, for the normativist, object level talk of changes in actual essential properties merely reflect a metalinguistic change in the ordinary rules that implicitly govern the relevant expressions. For example, an object level consideration of a change in the essential property of Sean Spicer's biological origin reflects a change in some constitutive rule that governs the name 'Sean Spicer'. Let me explain this in terms of a metalanguage mentioning the name 'Sean Spicer'.[157]

Suppose that one metalinguistic rule governing use of the name 'Sean Spicer' is: if 'Sean Spicer' applies to x, then the name is to be applied to y only when y traces to a common genetic origin with x.[158] Object language claims that violate this rule count as impossible against the back drop of the non-contingent essentialist's ordinary linguistic and conceptual practices. However, according to the normativist account of conceptual similarity, violations of this rule alone do not necessarily trivialize (14) or make it such that (14) is no longer talking about Sean Spicer but some other person, so long as the violation is "manageable", e.g. other speakers still roughly understand who in the actual world the utterer intends (14) to be about. Of course, what counts as a manageable violation of the rules that govern the use of a term will depend on the context. But intuitively, a circumstance in which the application of the name 'Sean Spicer' to a biological progeny of George Eliot is allowed, but where all other relevant actual features

[157] Again, this is a *model* of how counterpossible thought and discourse works with respect to (14). I am not saying that anyone ever explicitly reasons or talks through counterpossibles in a metalanguage.
[158] Thomasson (2007), p. 144. Not that the normativist is committed to non-contingent essentialism. Thomasson merely presents the rule as one example of the ways modal normativism can accommodate certain beliefs about *de re* modal properties.

of Sean Spicer are held fixed (e.g. certain facts about Sean Spicer's appearances, psychological dispositions, interests in politics, being born in the United States, being a human being, or whatever), does not mean that the application of 'Sean Spicer' to members of other species, or inanimate objects, or abstract objects will also be thereby *permitted*. Of course, all of this is much more easily and naturally communicated in an object language that uses the term 'Sean Spicer' and utilizes talk of impossible worlds or counterparts: impossible counterparts of Sean Spicer that have George Eliot as a biological parent are still more similar to Sean Spicer than other impossible counterparts, e.g. Sean Spicer the empty set, and so should be considered given the context.

This normativist approach to essences or so-called metaphysical laws avoids many of the difficulties discussed above. Recall that on the normativist account of modality in general, claims about what is metaphysically necessary or what properties are essential are not descriptive claims in need of truthmakers but have an expressive and prescriptive function. Therefore, modal claims are not in competition with descriptive claims of the empirical sciences, so we do not need to solve a placement problem and find a way to fit so-called modal facts in with non-modal empirical facts. Moreover, as Thomasson (2007b) notes, the normativist approach promises a more plausible epistemology:

> The normativist doesn't think of acquiring modal knowledge as a matter of coming to see new, different features of the world or perceiving possible worlds or a platonic world of essences. (150)

Instead, we come to understand claims of essence and metaphysical necessity by mastering our concepts and their relations, perhaps along with discovering additional relevant empirical information, e.g. as when we come to understand that *water is*

necessarily H_2O. These advantages also hold for my account of counterpossibles and metaphysical similarity for modal normativists.

On the account I have just offered, counterpossible discourse, or talk of impossible worlds and metaphysical similarity, is not used to make descriptive claims about substantive modal properties but is playing a prescriptive role. In general, talk of counterpossibles and talk of metaphysical similarity between impossible worlds allows us to express rules and permissions of languages that are different, yet *relevantly similar* to our "home" language. This general use, in turn, has two important applications. First, talk of counterpossibles and talk of metaphysical similarity between impossible worlds allows us to express rules and permissions *that continue to govern* the use of counterpart non-modal vocabulary when small, yet important, changes are made to the semantic normative profile (i.e. the set of application and coapplication conditions, the analytic entailments, etc.) of vocabulary in our "home" language (e.g. 'if George Eliot had been Sean Spicer's biological parent, Sean Spicer would still be human'). However, talk of counterpossibles and metaphysical similarity between impossible worlds can also be used to express the consequences of making changes to the current semantic normative profile of vocabulary in our "home" language but while remaining in the object language and without actually adopting those changes (e.g. 'if negation were governed by intuitionist logic, then $\neg\neg p \rightarrow p$ would not be a logical theorem').[159]

So, on my view, all we need to properly use and evaluate counterpossible claims and claims of metaphysical similarity are: a tacit grasp of, or ability to follow, the rules

[159] I will argue in Chapter 5 that this ability allows us to use counterpossibles to explore semantic normative consequences of adopting new linguistic practices by changing our current practices but without actually adopting those changes.

and permissions that govern the terms of our "home" language, perhaps along with other relevant empirical information; and a tacit grasp of, or ability to follow, the rules governing a language that is different, yet relevantly similar to our "home" language. So, there is no need to posit substantive relations of grounding and essence to make sense of our talk and judgments of counterpossibles and metaphysical similarity. Thus, we avoid the metaphysical and epistemological worries that come with substantive accounts of metaphysical similarity.

3.6 Strong Modal Classificationism?

In this section, I want to briefly look at an alternative, extreme deflationary approach to counterpossibles and metaphysical similarity. The purpose of this is to provide a contrast view of metaphysical laws that should not be confused with the views I have been arguing for in this chapter. I will argue that there are reasons to think that this alternative deflationary approach is *too* deflationary because it fails to capture the prescriptive function and role of counterpossible and metaphysical similarity discourse.

Sider (2011) and Cameron (2010) both argue that the metaphysically possible/impossible distinction is not a natural distinction, i.e. that it does not carve at the joints of reality. For example, Cameron (2010) suggests that the distinction between what is metaphysically possible and impossible is similar to what a B-theorist about time might say about the distinction between the past and the present. While Sider is skeptical of claims that certain metaphysical and mathematical truths, e.g. *wholes depend on their parts* or *2 + 2 = 4*, are made true by convention, he thinks it *is* a conventional matter that

such truths count as necessary.[160] According to Sider (2011), all it is to be a metaphysical necessity is to be true and to be classified as a necessity, and the selection is more or less arbitrary.[161]

What does this classificationist view of modality have to say about counterpossibles and metaphysical similarity? Let's call Sider's view of modality *weak modal classificationism*. While Sider is a conventionalist about basic modality, Sider also thinks that counterfactuals are objective in that what matters to the objectivity of counterfactuals are the similarity relations used to evaluate them.[162] Moreover, Sider seems to think that there are substantive metaphysical questions that can neither be answered by direct empirical methods nor conceptual analysis.[163] Furthermore, while skeptical of grounding relations between the fundamental and non-fundamental, Sider does think that there is a fundamental reality and that there is a privileged language that best fits the fundamental nature of reality.[164] So it seems that the natural option for this version of weak modal classificationism is to adopt a limited version of the substantive account of metaphysical similarity between worlds that takes facts about fundamentality to determine relevant similarity between worlds.

However, we might take the spirit of Sider's view of modality and try to generalize it to the notion of metaphysical laws themselves. On such a generalization, to be a metaphysical law comes down to nothing more than being classified as a

---

[160] However, see Thomasson (2007a, 2009, 2015) for responses to general objections against conventionalism about analytic and mathematical truths.
[161] Sider (2011), p. 271, 275.
[162] Ibid., p. 291.
[163] Ibid., p. 187.
[164] Ibid.

metaphysical law (and perhaps also being a truth). Call this view *strong modal classificationism*. On this view, we might say that the space of metaphysical necessity is determined by the classified laws and what follows from them. So, on a strong classificationist account of the talk and consideration of counterpossibles and metaphysical similarity, all we are doing is considering what would follow in different modal spaces such that the metaphysical laws are chosen differently, which is arbitrary.

Strong modal classificationism is not the view I have been arguing for. The main reason I think that strong modal classificationism is a bad way to go is that, while it is a deflationary view of metaphysical laws, it is not deflationary in the right way. Both weak and strong modal classificationism provide a plausible epistemological story of how we come to know basic modal claims such as '*p* is necessary'—we classified them as such. And strong modal classificationism likely provides a clear epistemological story for how we come to know what the metaphysical laws are—we classified them as such. By that fact, strong modal classificationism likely provides a clear epistemological story of how we know certain counterpossibles and claims of metaphysical similarity. However, both views fail to account for the function and role basic modal discourse plays in certain of our ordinary and theoretical practices, and strong modal classificationism fails to account for the function and role of counterpossible and metaphysical similarity discourse.

Again, many have noted the variety of ways modal thought is useful and important. For example: we care very much about the survival conditions of many things, including ordinary objects; modal thought is important to evaluating explanatory and causal claims; modal thought helps us predict so that we can make plans; it helps us in determining who or what is responsible for certain events; etc. In addition, many have

argued that members of the same linguistic community generally manage to use words to reliably coordinate their beliefs and actions, and speakers have similar dispositions in how they classify things as falling into the extension of terms in their language; further, they argue that this is all best explained by the existence of implicit linguistic conventions.[165] On the normativist view, object level talk of metaphysical necessities, essences, metaphysical dependencies, and metaphysical similarity helps us coordinate and make part of common knowledge our implicit conventions by serving the prescriptive function or correcting or condemning illicit uses of ordinary language. So, while this normativist account gives a plausible metaphysical and epistemological story of such object level modal talk, more importantly, this account explains *why* we even bother with such discourse to begin with and *what* we are *doing* with such discourse. In chapter 6, I will argue that the normativist account of modality, which now includes an account of counterpossibles and metaphysical laws, also plays an important role in negotiating for linguistic and conceptual changes while remaining in the object language. Thus, modal normativism should not be confused with strong modal classificationism.

3.7 Summary

At the start of this chapter, we seemed to be stuck in a dilemma. I argued in chapter 1: there are non-vacuous counterpossibles and that many non-vacuous counterpossibles can do a lot of interesting philosophical work; so, we need a theory of counterpossibles that accounts for their non-vacuity while allowing us to do the philosophical work they can be used for. However, in chapter 2, I argued that there are

[165] See Lewis 1969, Jackson 1998, and Schroeter 2017 section 4.1.

serious concerns with giving non-vacuous counterpossibles an epistemic reading and serious concerns with giving non-vacuous counterpossibles a substantive realist reading. So, given the difficulties, it seemed we had no way to account for non-vacuous counterpossibles and metaphysical similarity without taking on some difficulty or another. In this chapter I provided a plausible account of non-vacuous counterpossibles and metaphysical similarity. On my account, the considerations of metaphysical similarity needed to evaluate counterpossibles need not come down to epistemic matters, such as the subjective position of some agent, and it need not come down to substantive metaphysical matters, such as similarity with respect to substantive grounding relations. Instead, on my account, the considerations of metaphysical similarity needed to evaluate counterpossibles often comes down to conceptual matters. However, such conceptual matters are often expressed in object level claims about essences, metaphysical dependence, etc., which are claims that taken to be substantive metaphysical laws by the substantive realists. But, on my view, insofar as they make sense, expressions of metaphysical laws just are hyperintensional object language expressions of the actual conceptual laws, or their consequences, perhaps along with empirical information. Likewise, object language talk of a distinctive sort of metaphysical similarity between worlds plays a prescriptive function of illustrating, endorsing, or expressing rules that govern the use of ordinary vocabulary.

In the introduction to chapter 1, I established three desiderata for any theory of counterpossibles:

1. An account of counterpossibles should be consistent with our semantic intuitions, i.e. provide for the non-vacuity of counterpossibles—in particular, allow that there are some false counterpossibles;
2. An account of counterpossibles needs to enable counterpossibles to do the theoretical work we want them to;
3. An account of counterpossibles should avoid problematic ontological commitments and provide a clear story about how we come know counterpossible claims as well as related claims about metaphysical similarity.

My account of counterpossibles is consistent with our semantic intuitions in that it allows for some false counterpossibles. More importantly, my account provides a clear and plausible story of the metaphysics and epistemology of counterpossibles and metaphysical similarity. Thus, we are not forced to choose between epistemic and substantive realist accounts of counterpossibles. However, in the next two chapters I will look at some worries that my account fails to satisfy the second *desideratum*, that my account of counterpossibles can't actually do the work I outlined in chapter 1, which is the use of counterpossibles to clarify relations of metaphysical dependence and comparative impossibilities. However, I think that these worries can be addressed. Furthermore, in chapter 5, I will argue that my account of counterpossibles for modal normativists is bolstered by the new work for a theory of counterpossibles found in the literature on conceptual engineering.

CHAPTER 4: OBJECTION: MODAL NORMATIVISM CAN'T DO THE WORK

On my normativist account of counterfactuals, 'if *p* were the case, then *q* would be the case' expresses a requirement to accept that *q* given *p* as a hypothetical supposition just in case there is some *p q* world more relevantly similar to the actual world than all *p* non-*q* worlds. Roughly, considering both possible and impossible worlds, if we are going to *hypothetically* accept that the world is some way other than it actually is, then counterfactuals and counterpossibles convey a requirement about what kinds of worlds we ought to accept: if you're going to hypothetically accept a *p* world, it *ought* to also be a *q* world. But I also grant that another way of making the same point is to say that *p q* worlds are more relevantly similar to our world than every *p* non-*q* world, but where claims about worlds and similarity are not primarily descriptive but, instead, are expressive claims that illustrate the actual constitutive rules that govern the use of '*p*' and '*q*' along with additional relevant conceptual or empirical information being held fixed by the context.

In general, considerations of similarity reflect a norm of suppositional reasoning that we should not suppose more than is determined by the context when considering whether the consequent follows from the antecedent. Moreover, considerations of similarity enforce this norm by constraining the domain of considered circumstances to those that are contextually relevant. There are cases where considerations of similarity will be driven by epistemic features, e.g. the features of a range of phenomenological circumstances, and other cases where considerations of similarity will be driven by physical features, e.g. the empirically tractable features of a range of physical

circumstances. I argued that when confronted with consideration of a distinctive kind of metaphysical similarity, e.g. similarity with respect to essential properties, it was best to appeal to conceptual criteria, rather than substantive metaphysical criteria, to rank similarity.[166]

My account of counterpossibles, metaphysical similarity, and metaphysical laws does require questionable ontological commitments and delivers a clear and plausible epistemological story about our talk and consideration of counterpossibles, metaphysical similarity, and metaphysical laws. This is because the account rests on three plausible claims presented in the last chapter. First, that many expressions of metaphysical laws, e.g. claims of essence and metaphysical dependence, just are object language expressions of the actual conceptual laws, where the conceptual laws are the constitutive rules found in the frame-level content of expressions. Second, that expressions in our language can have semantic counterparts, i.e. that expressions in our language can be more or less similar to expressions in alternative languages. Finally, talk of counterpossibles and metaphysical similarity between impossible worlds allow us to express rules and permissions of languages that are different, yet *relevantly similar* to our "home" language.[167]

In this chapter, I look at three worries for my account. All of these worries fall under a more general worry: namely, that my account can't actually do the work I set out

[166] The same can be said for other considerations metaphysical similarity such as logical and mathematical similarity, especially given a plausible inferentialist account of these things. For an interesting example of recent work on conventionalism in mathematics and logic, see Warren (2015, 2016).
[167] I provided an explanation and argued for the first and third claims in the last chapter. The second claim about semantic counterparts is developed and defended in Warren (2015, 2016), though also see Jackson (1998) p. 45, Thomasson (2007a) p. 37, and Sellars (1958) pp. 287-288 for closely related suggestions.

for counterpossibles in section 1.6. If my account cannot do the work I set out for it at the beginning of this dissertation, then I might be stuck with two options. On the one hand, I might have to reconsider the theoretical reasons to accept that there are non-vacuous counterpossibles, which might mean I should reconsider Williamson's arguments. On the other hand, I might have to reconsider the plausibility of the substantive approach to counterpossibles since it presumably can do the work. Neither of these is an attractive option given what I have set out to do.

In the next section, I will explain an intuitive distinction between indicative conditionals and their counterfactual counterparts. I will suggest how those inclined to substantive metaphysical realist views about grounding and fundamentality might use this distinction to argue against my account of counterfactuals and counterpossibles. Then, in section 4.2, I will develop this worry further by discussing various conditional claims that fans of metaphysical grounding and fundamentality might use to motivate the idea that we really do need to appeal to substantive metaphysical properties and relations to explain the apparent distinction between indicative and counterfactual conditionals seemingly about metaphysical matters. The worry is that because I reject substantive relations of grounding or fundamentality, I cannot adequately account for this distinction. In response, I will show how the normativist can do just this. In section 4.3, I will look at complications that arise in certain a famous example about Socrates that requires impossible worlds and seems to bolster the substantive realist's worries and show how my account can easily handle this case as well. Finally, in section 4.4, I will summarize a worry that the that the best explanation of the comparative impossibilities I discussed in section 1.6 is that, just as certain physical properties seem to present varying degrees of

resistance to changes in reality, substantive essential properties and grounding relations likewise present varying degrees of resistance to changes in reality. So, the worry is that since I reject substantive metaphysical properties and relations, I cannot adequately account for comparative impossibilities. However, I will show how my normative account can do just this.

4.1 Indicative Conditionals Vs. Counterfactuals

There is an intuitive distinction between certain indicative conditionals and counterfactuals: namely, that epistemic modal concepts seem more relevant to indicatives, while non-epistemic modal concepts seem more relevant to counterfactuals.[168] Some evidence for this distinction comes from what Lycan (2001) calls 'Adams Pairs', in which an indicative conditional and it's corresponding counterfactual differ in truth value—e.g., note the contrast between "If Oswald did not shoot Kennedy, then someone else did" and "If Oswald had not shot Kennedy, then someone else would have."

One way to explain the difference between indicative conditionals and counterfactuals is in terms of different kinds of information we are supposed to hold fixed when evaluating a given conditional. For example, when evaluating indicatives, Stalnaker (1975) suggests that we hold fixed those presuppositions that are common ground given the context. Nolan (2003) gives a possible world account for indicative conditionals that requires we hold fixed what we know. Lycan (2001) suggests that:

> A conditional is lexicalized [indicatively] when its utterer holds fixed some salient fact that is looming large in his or her epistemic field… A conditional is

[168] See Lycan (2001).

> lexicalized [subjunctively] when its utterer means to prescind from contextually salient facts and consider a wider range of alternative possibilities constrained only by broader and perhaps more idealized epistemic considerations. (145)

For many cases, the overall idea is that, on the one hand, when evaluating certain indicative conditionals we hold fixed salient epistemic matters, such as how things appear to us, and conceptual matters, such as certain semantic features of our expressions.[169] On the other hand, when evaluating counterfactuals, we will not limit ourselves to common presuppositions, beliefs, or knowledge, but we still might hold fixed other relevant information, e.g. certain empirical matters of fact, certain causal relations, or the physical laws more generally.[170]

Let me clarify the purported difference between the two conditionals by returning to the well-known Adams Pair:

(17) If Oswald did not shoot Kennedy, then somebody else did.

(17*) If Oswald had not shot Kennedy, then somebody else would have.

It seems that (17) is true but (17*) is false. On the one hand, (17) suggests something about our preconsidered beliefs or evidence. Given everything else that we currently believe germane to Kennedy's death is true—in particular the fact that Kennedy was shot and killed on November 22, 1963—if it turns out that we are wrong about who killed Kennedy, then it certainly does follow that somebody else shot Kennedy. On the other hand, (17*), abstracts away from our common presuppositions and beliefs. What's more, causal connections between Oswald's actions and Kennedy's death seem to have some bearing on the evaluation of (17*), and perhaps explains why (17*) is false: if Oswald's

[169] See also Lycan (2001), Chalmers (2002), and Yablo (2002).
[170] See Lewis (1973, 1986) and Kment (2014).

gun had jammed, there would have been no gunman (e.g. a gunman in the grassy knoll) nearby to act as a causal backup for Kennedy's death.

So, there is a difference in content and force between the indicative and counterfactual conditionals seemingly about empirical matters such as the assassination of Kennedy. The difference in content is reflected in the difference in truth values between (17) and (17*). The difference in force can be explained by which conditional you would appeal to when offering certain explanations—an explanation regarding what we know vs. an explanation of what caused what. More generally, what we have are two contingent events E and C (in the above example C = *Oswald shot Kennedy* and E = *Kennedy died from a gunshot*). We also have that C sometimes bears an evidential connection to E such that if we have no reason to change our belief that E occurred, then if we are actually wrong about C, we need to find a different hypothesis to explain E. Given this, the indicative conditional can be used to convey that C is a relevant part of our current *evidential explanation* for E. However, E sometimes counterfactually depends on C such that if C had not occurred, E would not have occurred. Given this, the counterfactual can sometimes help communicate that C is a relevant part of our current *causal explanation* of E. One thing that the Kennedy example shows is that our evidential and causal explanations of E can come apart.[171]

However, many counterfactuals pertinent to a metaphysical point of view involve or seem to be about non-empirical matters regarding mathematics, logic, composition,

[171] I am not claiming that all counterfactuals are about non-representational matters such as the causal dependence of one event on another. I agree with Vetter (2016) that some counterfactuals are concerned with representational matters such as what the available evidence entails. See section 2.3 above.

determinate/determinable relations, essence, etc.[172] One metaphysical picture of these counterfactuals is that they are on a par with counterfactuals involving primarily empirical claims. For example, for some, it is tempting to treat the counterfactual *if Socrates had not existed, then the set {Socrates} would not have existed* as a counterfactual claim about substantive metaphysical matters that is somehow on a par with a counterfactual claim about empirical matters such as *if Oswald had not shot Kennedy, then Kennedy would have gone on to unify prominent Democratic leaders in Texas for the upcoming 1964 elections*. As I just explained, the latter claim can be seen as expressing some empirical or causal relationship between the content of the antecedent and the consequent. But in the Socrates example, it seems that we are likely not expressing any empirical or causal relationship between living beings and abstract sets. Some substantive realists respond by arguing that, while the relationship between Socrates and the singleton set {Socrates} is not a diachronic relation of causal dependence, there is some kind of synchronic relationship of metaphysical dependence or metaphysical fundamentality between the two. For example, here is Kment (2014) giving an analogy between the laws of metaphysics with the laws of nature:

> That is to say, facts about more fundamental things give rise to facts about less fundamental ones in accordance with the laws of metaphysics, just as in a deterministic universe earlier events bring about later ones in accordance with the laws of nature. (147)

Call this picture the *metaphysical fundamentalist* picture.[173]

---

[172] I will not offer or commit to an explanation of the exact difference between a conditional that is purportedly about metaphysical matters and one that is purportedly about empirical matters. So, I'll leave it at an intuitive level, e.g. conditional claims about mathematics, logic, and essences versus conditional claims about, say, observable physical events.

[173] Something like this view is endorsed in Fine (2001), Schaffer (2009, 2017), Rosen (2010), Krakauer (2012), Kment (2006, 2014), and Wilson (2016).

I think that this perceived parallel with causation is misleading and theoretically burdensome since it requires us to posit substantive metaphysical relations that come with their own conceptual, metaphysical, and epistemological difficulties. However, the fundamentalist can now try to raise the following worry for my view of counterpossibles and metaphysical laws. Since I think we should avoid positing any world-oriented metaphysical information analogous with world-oriented causal information to hold fixed when evaluating the relevant counterfactuals, I undermine my ability to offer a plausible explanation of the intuitive distinctions between certain indicative and counterfactual conditionals seemingly about metaphysical matters. Moreover, this means that the normativist cannot account for the role counterfactuals, counterpossibles, and considerations of metaphysical laws and similarity play in offering explanations about metaphysical matters, such as the relationships between urelements and sets, which is a job description I outlined for counterpossibles in section 1.6.

4.2 Evidential Indicatives Vs. Counterfactual Dependence

In this section, I will clarify the worry just outlined by looking at examples found in the literature on grounding and fundamentality. Then, I will begin to formulate a normativist response based on the examples presented. I will argue that the normativist can still account for the intuitive differences between indicative and counterfactual conditionals and do so without appealing to substantive metaphysical relations that are analogous to empirical relations such as causation. Then I will clarify the response in section 4.3 when I look at an argument presented by Wilson (2016) that grounding claims entail the non-vacuity of counterpossibles.

Overall, the metaphysical fundamentalist is trying to argue that the discussion of (17) and (17*) shows that indicative conditionals seem more related to representational matters—e.g., our evidence, what we believe, or how the world is presented to us—and *some* counterfactuals seem more related to nonrepresentational matters—e.g., worldly causal relationships. Therefore, an analogous explanation can be given for the apparent differences between indicative and counterfactual (as well as counterpossible) conditionals primarily about metaphysical matters. In particular, many counterfactual conditionals primarily about metaphysical matters reflect something about the nonrepresentational world, perhaps the existence, or sometimes even a lack of existence, of a substantive metaphysical relationship between whatever the antecedents and consequents are about, e.g. relations of metaphysical fundamentality or metaphysical grounding that obtain between urelements and sets.[174]

Of course, while the metaphysical fundamentalist thinks that certain counterfactuals are used in giving substantive *metaphysical explanations*, e.g. explanations about the grounding structure of reality, the fundamentalist still thinks that indicatives can be used in evidential or conceptual explanations.[175] Thus, the fundamentalist is able retain the apparent difference between the content and force of the relevant indicative and counterfactual conditionals even when such conditionals are primarily about metaphysical matters. The fundamentalist will now raise the following view for my account of counterfactuals and counterpossibles: I can't account for the

[174] For these examples, I'll mostly set aside the complications in the literature that arise from talking about sentences versus talking about facts, e.g. in the latter case we treat grounding as an operator and in the latter as a relation, cf. Fine (2012). Nothing in this discussion really hinges on the distinction, so, I will sometimes use fact talk freely and loosely.
[175] Cf. Kment (2006, 2014).

distinction between indicative and counterfactual conditionals seemingly about metaphysical matters because I think *all* metaphysical matters reduce to conceptual matters (perhaps along with empirical information). We can develop the fundamentalist's concern with my normativist approach more clearly by looking directly at examples often discussed in the literature on metaphysical grounding and fundamentality. Overall, this puts pressure on my normativist account of counterfactuals to offer a plausible explanation of the apparent differences the moods of the relevant conditionals, but, of course, to do so without having to appeal to substantive metaphysical relations analogous to the causal relations found in the Kennedy example.

The first example that apparently illustrates a difference between evidential explanations and substantive metaphysical explanations involves determinates and determinables. Note again that, in all of the examples, we are looking for two features:

i. For the indicative conditional, if it turns out that our ostensible evidence or explanation for E, which is C, is false, then we don't change our belief that E is the case, but instead find different evidence or a different explanation for E; and
ii. For the counterfactual, that E *counterfactually depends* on C such that if C were not the case, then E would not be the case.

Consider the following. Let E = *the ball is red* and C = *the ball is crimson*. It is generally held by fans of dependence and grounding talk that E metaphysically depends on (or is grounded by) C. In (17), which is in the indicative mood, we hold the effect of Oswald's actions, namely Kennedy's death, fixed. Analogous to (17), in this case suppose that we hold the "effect", namely that the ball is red, fixed. Also suppose that we find out that all along we *might* have been confusing crimson with some other shade of red that

phenomenally presents itself to us as crimson but is in fact a distinct shade of red.[176] Then the following indicative conditional seems true: *if the ball is not crimson, then the ball is some other shade of red*. So, like (17), this indicative conditional can play a role in epistemic explanations involving C and E.

Now consider the same conditional as a counterfactual: *if the ball were not crimson, then the ball would be some other shade of red*. This counterfactual is arguably false. Here's how. Unlike (17), when we evaluate (17*), i.e., the Kennedy *counterfactual*, we do not hold fixed whatever is *causally downstream* from Oswald's actions. Here is a picturesque way of thinking about it: if God were to rewind the world to right before Oswald pulled the trigger and intervened by jamming Oswald's gun, then Kennedy would not have died (assuming that God did not put a gunman in the grassy knoll). Similarly, one can argue that if God were to rewind the world to right before she made the ball crimson and intervened by changing the color, then she might have made the ball *any other shade of any other color*, e.g. perhaps cerulean and so the ball would be blue. So, just as Kennedy's death counterfactually depends on Oswald shooting him, a ball's being red counterfactually depends on it being some particular shade of red. Now, the fundamentalist will go on to explain that just as the counterfactual dependence between Oswald's actions and Kennedy's death is somehow connected to a causal relationship between the two events, the counterfactual dependence between the ball's being crimson and the ball's being red is somehow related to a substantive relation of metaphysical dependence between the two facts.[177] Some fundamentalists will go on to argue that the

[176] Cf. Jackson (1982).
[177] The fundamentalist I am considering thinks that the relation, which is likely a big-G Grounding relation, is distinct from say the determinate/determinable relationship, which is likely a little-g grounding relation. See Wilson 2014 for more and a criticism of big-G grounding relations.

counterfactual dependence reflected by the Kennedy counterfactual suggests that just as Oswald's actions are *temporally or causally prior* to Kennedy's death, the ball's being a particular shade of red is *metaphysically prior* to it's being red, i.e. is *more fundamental than*.[178] So, analogous to (17*), the counterfactual just discussed shows that counterfactuals can play a role in substantive metaphysical explanations involving C and E. Thus, the fundamentalist thinks that different explanatory relationships between C and E are highlighted depending on whether we express the conditional as an indicative or as a counterfactual.

For the second example, let E = *the fact that A or B* and C = *the fact that A*.[179] Again, it is generally held by fans of dependence and grounding talk that E metaphysically depends on (or is grounded by) C. The corresponding indicative seems to be: *if the fact that A is not the case, then B makes the fact that A or B the case*. So far, so good: this seems to fit the analogy with the Kennedy indicative because we are holding the effect, which in this example is the fact that *A* or *B*, fixed while entertaining the possibility we are wrong about *A*; so, if *A* or *B* is the case and *A* is not the case, then *B* is the case.

The corresponding counterfactual is: *if the fact that A were not the case, then B would have made the fact that A or B the case*. There is a plausible way to read this counterfactual as false, and so keep the analogy with the Kennedy case, but we need to be careful. On a less plausible reading, the idea is that: if God were to rewind the world to right before she brought the fact that *A* into existence and intervened on *A* coming into

[178] Cf. Kment (2014).
[179] Where 'A' and 'B' are names of well-formed sentences.

existence, then she would not thereby have intervened on the fact that *A* or *B* coming into existence because she had already brought the (backup) fact that *B* into existence. But, on this reading, the counterfactual is true, which deviates from the Kennedy example. However, there is a more plausible way of reading the counterfactual so that it comes out false. The idea is that: if God were to rewind the world to right before she brought the fact that *A* into existence or she brought the fact that *B* into existence and intervened on *A* coming into existence, then she would have thereby intervened on the fact that *A* or *B* coming into existence. On this way of reading the counterfactual, God hasn't necessarily decided if the fact that *B* will be the case, so when she intervenes on the fact that *A*, she intervenes on the fact that *A* or *B*, at least for the time being. This way of reading the counterfactual fits with the Kennedy case. Again, the metaphysical fundamentalist takes the counterfactual dependence between the fact that *A* and the fact that *A* or *B* to suggest a distinct relation of metaphysical dependence between the two facts or suggest that the disjunctive fact is less fundamental than the simple fact.

Now let E = *the stuff in this glass is water* and C = *the stuff is in this glass is H_2O*. Again, it is generally held by fans of dependence and grounding talk that E metaphysically depends on (or is grounded by) C. Hold fixed that the stuff in this glass is water. The corresponding indicative seems to be: *if the stuff in this glass is not H_2O, then the water in this glass has a chemical composition other than H_2O*. This seems true: if our chemistry is wrong, then we need some other evidential or chemical explanation for the stuff in this glass being water. So far, this fits with the Kennedy example.

Now consider the relevant counterfactual: *if the stuff in this glass were not H_2O, then the water in this glass would have had a chemical composition other than H_2O*.

There is a way to read this as false and so get the right result: if God were to rewind the *actual* world to right before she brought pairs of hydrogen atoms into covalent bonds with single oxygen atoms and intervened such that H_2O never came into existence, then she would have thereby intervened on the chemical composition of water, and so she would have intervened on the *water* in the glass coming into existence at all. Thus, there would be no water in the glass to have a different chemical composition.

Let me summarize where we are before I offer the start of a normativist explanation of the apparent differences between the indicative and counterfactual versions of these conditionals. In the Kennedy conditionals we follow something like the following heuristic. When evaluating the *evidential relationships* between C = *Oswald shot Kennedy* and E = *Kennedy died from a gunshot*, we hold fixed relevant information, namely Kennedy's death, which also happens to be causally downstream from C. But then we suppose that we are actually wrong about C, in which case somebody else did cause Kennedy's death. When evaluating the *causal relationships* between C and E, we do not hold fixed what is causally downstream from C, so when we rewind the world to a moment in time before Oswald's actions and intervene on those actions, we "stop the causal flow" by removing the cause of Kennedy's death, so Kennedy's death does not result so long as there is no nearby causal backup.[180]

One feature of the Kennedy case that explains why the two conditional moods can come apart is that there is a "causal gap" so to speak between the two events, which can open up certain epistemic possibilities in the right context. To clarify, there are a

[180] I am not suggesting that the best way to evaluate subjunctive conditionals is in terms of similarity of world histories up to the antecedent time, which is discussed in Lycan (2001), especially pp. 142, 153-154. I am only using this language for illustrative purposes.

multitude of ways Kennedy dies from a gunshot in the total space of physical possibilities. But, (17*) relies on context and particular considerations of similarity with the actual world up to the time of Oswald's actions that significantly constrains the range of possibilities we need to consider, namely it removes all nearby possible scenarios in which Kennedy dies from a gunshot. The indicative mood does not require the same considerations, and so keeps the space of relevant possibilities open.

The metaphysical fundamentalist wants us to believe that something similar is going on in the three conditionals just discussed. In particular, the fundamentalist thinks that there are certain non-causal connections between whatever the antecedent and consequent claims are about that matter to evaluating the conditionals. Sometimes there is a "metaphysical gap" between two facts, e.g. there are lots of possible ways for a ball to be red, that can open up certain possibilities in the one context, e.g. an epistemic context, but which are precluded in others. What conditional mood we use to talk about the two facts can matter. The indicative construction is useful for expressing that we are in an epistemic context, the counterfactual is useful for expressing that we are in the latter, more restricted context. So, perhaps when we encounter an indicative conditional involving C and E, we hold fixed whatever is metaphysically downstream from C but not when we evaluate the corresponding counterfactual.

Now, the normativist agrees that there are non-causal considerations at play when evaluating the conditionals just reviewed, which also seems to account for the differences between the indicative and counterfactual moods. However, for reasons already discussed, the normativist is skeptical of bringing in substantive metaphysical relations such as fundamentality or grounding, to account for the differences. But, the normativist

notes that we are not required to explain the apparent distinction in content and force between the indicative conditionals and counterfactuals just discussed by positing such relations. For the normativist, an alternative and less theoretically burdensome explanation for the apparent differences revealed in the above examples is that we are simply dealing with important conceptual connections between the antecedents and consequents. On the normativist account, sometimes there is a "conceptual gap" between two sentences that can open up certain possibilities in one context that are precluded in another. So, the normativist agrees that what conditional mood we use to talk about the two facts expressed by the sentences can matter.

On the one hand, when evaluating the above indicative conditionals, we hold fixed what is *conceptually downstream* from C, and note that there is enough of a conceptual gap between the antecedent and consequent to allow for some other possible conceptual fact to explain the consequent.[181] This gap can open up certain possibilities, specifically possibilities involving alternative conceptual explanations of E. The indicative mood is useful for expressing that we are in an epistemic context where those possibilities are available. On the other hand, when evaluating the above counterfactuals, we do not hold fixed everything that is conceptually downstream from the antecedent, and so when we "intervene" on the antecedent we also thereby "intervene" on whatever is conceptually downstream from the antecedent. This results in constraining the possibilities to be considered in a given context. The normativist agrees that the counterfactual mood reflects that E counterfactually depends on C in all of the above

[181] Though the normativist is free to sometimes talk about phenomenological gaps as in the indicative involving colors.

cases. But, the normativist does not think that the counterfactuals reflect that C is a relevant part of a *substantive* metaphysical explanation of E. Instead, the normativist thinks that the counterfactual dependence expressed by the counterfactuals is a relevant part of the *conceptual explanation* of E.[182]

Understanding relations of conceptual dependence involved in the above examples and the role they can play in conceptual explanations requires no more than understanding the truth preserving conceptual connections between C and E, perhaps along with certain empirical facts. Further, those conceptual connections are determined by the frame-level contents of the expressions involved in the linguistic analogues of C and E, e.g. the inferential roles of those expressions. For example, 'something is red' conceptually follows from 'something is crimson', and '*A* or *B*' conceptually follows from '*A*'. The water example is a bit more complicated because whether the stuff in a glass is water partly depends on empirical facts about the chemical composition of what is in the glass. But explaining the counterfactual dependence between the water in a given glass and the chemical composition of the stuff in that glass also seems to rely heavily on the conceptual truth that whatever the chemical composition of the stuff in the glass is, that stuff has its chemical composition necessarily.[183] So, all in all, the normativist can account for the apparent differences between indicative conditionals and counterfactuals seemingly about metaphysical matters without appealing to substantive metaphystical relations of grounding or fundamentality. The mood of a conditional can influence what sorts of information we hold fixed when evaluating a conditional, but this does not

[182] As I discussed in 3.4, what conceptually entitles us to E.
[183] Cf. Thomasson (2007) and Sidelle (1989).

require that we posit substantive metaphysical information to hold fixed. Instead, we can simply rely on conceptual information.

4.3 The Asymmetry of Metaphysical Explanation

I now want to look at an objection to my view to the effect that it cannot explain the asymmetry of metaphysical dependence and metaphysical explanations.

A claim about ordinary constant conjunction between *x* and *y* can fail to reflect causal asymmetries between *x* and *y*. Similarly, certain modal generalizations, e.g. 'necessarily, *x* iff *y*', may not fully reflect metaphysical asymmetries between *x* and *y*. Metaphysical laws are supposed to reflect these asymmetries as well as support a certain degree of counterfactual robustness.[184] For example:

> Given the existence of the set {Socrates}, *part of what it is* for the set {Socrates} to exist is for Socrates to exist. ({SOCRATES} LAW)

should support the counterfactual:

(18) If an intervention had prevented Socrates from existing, then the set {Socrates} would not have existed.

However, since the existence of Socrates seemingly does not depend on the existence of the set {Socrates}, the following counterfactual should be *false*:

(18') If an intervention had prevented the set {Socrates} from existing, then Socrates would not have existed.

[184] Schaffer 2017, pp. 305-309.

Moreover, the asymmetry and differences in counterfactual dependence should hold despite the fact that 'necessarily, Socrates exists iff the set {Socrates} exists' is true.[185]

Wilson (2016) suggests that (18') is false because it does not respect the *metaphysical asymmetry* between Socrates and the set {Socrates}. Given Wilson's interventionist account of grounding, which takes there to be a close connection between counterfactuals claims and grounding claims, I take him to be suggesting something like the following. Lewis (1986b) argues that just as we ordinarily take the past to be independent of the present and the present independent of the future, we ordinarily take causes to be independent of their effects. The cause of the glass breaking is my dropping it. So, *if I had not dropped the glass, it would not have shattered*. But, in ordinary contexts, it is false to say that *if the glass had not shattered, then I would not have dropped it*—in nearby scenarios, someone could have caught the glass right after I dropped it or maybe I was standing two feet to the left so that the glass landed on some thick carpet. Moreover, given a connection between causation and counterfactual dependence, if this second counterfactual were true, it would imply backward causation, which in *standard* contexts is not possible.

Applying this idea to metaphysics, the fundamentalist claims that in standard contexts we take more fundamental facts to be metaphysically independent of less fundamental facts. Intuitively, the existence of Socrates is somehow metaphysically upstream from the existence of the set {Socrates}, e.g., Socrates is more fundamental

[185] Actually, one could reject that the necessary covariance holds; see Rayo (2015). Assuming that the covariance holds makes defending my view of metaphysical laws more difficult (as it does for even realist account of laws), but it also allows me to show how my view works in more detail, so I will keep the assumption.

than the set {Socrates}. So, while an intervention on the existence of Socrates has metaphysical consequences for the existence of {Socrates}, an intervention on the existence of {Socrates} does not have the same consequences. As such, counterfactuals that imply that the fundamental sometimes depend on the less fundamental, such as (18'), are guilty of a sort of metaphysical backtracking or backward metaphysical causation. Hence, (18') is false on a standard interpretation which does not allow for backtracking.

Wilson argues that the truth of (18) and the falsity of (18') require that there are nontrivial counterpossibles since if, necessarily, Socrates exists iff {Socrates} exists, you cannot intervene on one without intervening on the other. Again, the fundamentalist takes this to mean that some counterpossibles reflect a substantive relation of grounding or metaphysical dependence between two facts. Some fundamentalists go further by suggesting that such relations are what make certain counterpossibles true while the absence of such relations make certain counterpossibles false.[186] All in all, on the fundamentalist picture, substantive metaphysical relations support counterfactuals and counterpossibles about seemingly metaphysical matters and explain the asymmetry of metaphysical explanations. These connections are exactly what explains *how* counterpossibles are able to do work in clarifying dependence relationships, and they explain exactly *why* philosophical explanations of causal or metaphysical dependence can be supported or undermined by appealing to non-vacuous counterpossibles. The worry is that, without appealing to substantive metaphysical relations, it isn't clear that my theory of counterpossibles can do the needed work. So, I need to show how my normativist

[186] See Kment (2006).

account of metaphysical laws respects this asymmetry and supports the relevant counterpossibles.

I agree that (18) seems true while (18') seems false. Moreover, I agree that this reflects something of theoretical interest and not just something about our intuitions about certain conditional claims. However, I do not think that we must appeal to substantive metaphysical relations such as grounding in order to explain the data. The normativist can offer an alternative but parallel explanation, and, moreover, one that is less theoretically burdensome. Roughly, the normativist explanation is that the asymmetry between Socrates and the set {Socrates} exists because set concepts play no important or central definitional or conceptual role in our concepts of particular concrete objects such as Socrates. In other words, while the urelement Socrates plays an important role in the application and coapplication conditions for 'the set {Socrates}', the converse does not hold. Basically, singleton sets of urelements are *conceptually downstream* from the urelements themselves. For example, suppose that certain externalist views about language use and meaning is true. If all expert mathematicians, logicians, and other relevant inquirers had decided that the practice of set theory was useless, and all set-theoretic concepts should be forever abandoned, then they would have intervened on the concept or meaning of 'the set {Socrates}' but would not have significantly (if at all) have intervened on the concept or meaning of 'Socrates'.

The *one-way* dependence between Socrates and the set {Socrates} implies that the following is false:

(18'') If only one of Socrates or the set {Socrates} were to exist, then it would be the set {Socrates}.[187]

Since (18'') is false, there is an impossible world where the set {Socrates} fails to exist but Socrates still exists that is more relevantly similar to the actual world than some world where Socrates fails to exist but the set {Socrates} exists anyway. That is, given the following two impossible circumstances:

(IC-1) An impossible circumstance in which Socrates exists but the set {Socrates} does not exist;

(IC-2) An impossible circumstance in which the set {Socrates} exists but Socrates does not exist.

IC-1 should come out more relevantly similar to the actual circumstances than IC-2.

So, how do we go about determining whether IC-1 is more similar to the actual circumstances that IC-2? Recall the metaphysical laws from above:

> For any entities found up to stage *n*, there is a set at stage *n* + 1 that contains all an only the entities found up to stage *n*. (SET LAW)
>
> Given the existence of the set {Socrates}, *what it is* for the set {Socrates} to exist is for Socrates to exist. ({SOCRATES} LAW)
>
> Given the existence of Socrates, *what it is* for Socrates to exist is for Socrates to originate from his actual biological parents. (SOCRATES LAW)

[187] Steward (2015) notes that the closest world where nothing has the property of being singleton Socrates is the *possible* world where Socrates doesn't exist, so the (previously stated) conditional 'if the set {Socrates} had not existed, then Socrates would not have existed' is true, not false. However, Krakauer (2012) suggests framing the issue of metaphysical dependence using counterpossibles like (18''), which gets around Steward's worry.

One way can we decide whether one impossible circumstance is more similar to the actual circumstances than some other impossible circumstance is by deciding which circumstance requires less deviation from the actual metaphysical laws while holding the natural laws and other matters of fact fixed.[188] Given something like this similarity metric, my account of metaphysical laws—according to which the metaphysical laws are object language expressions of the actual conceptual laws—should be able to support an evaluation of similarity in which IC-1 is more relevantly similar to the actual circumstances than IC-2. If that is the case, then my account of metaphysical laws can support counterpossibles about seemingly metaphysical matters and explain the asymmetry of metaphysical explanations without positing the existence of substantive metaphysical relations.

Suppose that all it is for Ks to exist in a hypothetical circumstance is that the application conditions *actually* associated with 'K' are fulfilled.[189] Now, IC-1 does require a violation of (SET LAW), which presumably is an object language expression of the general application conditions for sets, including singleton sets. One possibility is that (SET LAW) mostly holds in IC-1 but has an exception when it comes to Socrates only. Another possibility is that singleton sets or sets in general simply do not exist in IC-1. Supposing that some standard theory of sets is actually true, this second possibility would require a widespread violation of (SET LAW) since *every* urelement that exists in IC-1 would be without a singleton. I think that there are considerations in favor of either option. One reason in favor of the first option, where the lack of existence of the set

[188] Given a Lewisian approach to similarity, this will be just one metric of similarity among many, and which is similarity relation matters will be determined by context. See Lewis (1973).
[189]See Thomasson (2015) for a more thorough presentation and defense of this idea.

{Socrates} is an exception to (SET LAW), is that there would still be many claims involving sets that are actually true, e.g. 'Ø exists' and '{Ø} exists', that will be true at IC-1. One reason in favor of the second option, say where IC-1 is a world without sets, is that it isn't really difficult to imagine a world with urelements but no sets because it isn't clear what significant difference the existence of sets makes to the concrete world. So, why not let IC-1 be an impossible circumstance in which no sets exist as opposed to an impossible circumstance in which, for some strange reason, every set but the singleton set {Socrates} exists? Regardless, I don't think that it's clear on *any* account of metaphysical laws which option is better. Furthermore, there is nothing in my account of metaphysical laws, where (SET LAW) is an object language conceptual truth that illustrates the application conditions of the term 'set' and *not* the name of any urelement, that forces me to choose one option over the other. So, I will set this complication aside.

What is important is showing how my account respects the intuition that the existence of Socrates does not depend on the existence of any set. Suppose that (SOCRATES LAW) is an object language expression of the following metalinguistic rule: Apply 'Socrates' to *x* only when *x* biologically originates from *O*.[190] (SOCRATES LAW) implies that, given the existence of Socrates, the existence of the set {Socrates} is *not* part of what it is for Socrates to exist. On my view of laws, this object language implication regarding the ontological independence of Socrates from the existence of sets

[190] Where '*O*' specifies whatever actual biological process that produced Socrates. This means that in order to discover under what conditions we may appropriately apply 'Socrates', which in the object language is discovering *what it is* to be Socrates, we do need to use straightforward empirical methods to discover who Socrates's actual biological parents are or find some DNA profile of Socrates. However, when we do discover this empirical information, this does not mean that we have also discovered some substantive metaphysical property of Socrates. Instead, on my view, the real definition of Socrates is the combination of a conceptual truth, in this case (SOCRATES LAW), with that empirical discovery whatever it may be.

is no more than a tacit endorsement of application conditions for the term ‘Socrates’ that do not require the application conditions for any set term, including ‘the set {Socrates}’, to be fulfilled. Thus, on my account, we are still entitled to apply ‘Socrates’ at IC-1 even though the application conditions given by ‘the set {Socrates}’ are not fulfilled. Thus, IC-1 is relevantly similar to the actual world with respect to the actual metaphysical laws regarding the existence of Socrates, among other things.

However, given the context, this will *not* be the case for IC-2. The actual application conditions for ‘the set {Socrates}’, by definition, requires the existence of Socrates, so the actual application conditions for ‘the set {Socrates}’ are not met by IC-2. But, the set {Socrates} still exists according to IC-2. Thus, ({SOCRATES} LAW) is violated at IC-2. For similar reasons, IC-2 also requires a violation of (SET LAW). According to (SET LAW), the set {Socrates} should contain *only* those entities that are found at the urelement stage, i.e. Socrates. But Socrates doesn’t exist in IC-2. Thus, (SET LAW) is violated in a circumstance in which at least one set exists (and so cannot be a relevant circumstance in which no sets exist). Therefore, in order to accommodate IC-2, we would need to change the conceptual laws governing our language, such as the application conditions of ‘the set {Socrates}’ and perhaps even (SET LAW), and this is in addition to changing other empirical matters of fact, such as Socrates not existing and related matters of fact. So, IC-2 deviates too far from the actual metaphysical laws regarding the set {Socrates}, among other things. So, IC-2 is not a relevant circumstance. Thus, IC-1 is more relevantly similar to the actual world than IC-2.

Let me emphasize that my primary aim here is neither to establish once and for all that (18’’) is false nor to argue for essentialist or dependence claims about singleton sets

and Socrates. Like Fine (1994), I am only using Socrates and the singleton as an example of the intelligibility of a position regarding the connection between basic modal claims and claims about essence and metaphysical dependence.[191] In particular, I am using the example of Socrates and the singleton set {Socrates} to show that my account of counterpossibles and metaphysical laws can support the apparent asymmetry between the two without appealing to substantive metaphysical relations. And, indeed, the explanation I just gave did not require appealing to any substantive metaphysical relations, e.g. a grounding relation, between sets and urelements. The explanation I gave only appealed to conceptual matters such as the application conditions for certain terms and how those application conditions compare. These conceptual matters can be expressed with object language claims such as ({SOCRATES} LAW), (SOCRATES LAW), (SET LAW) or object language implications of these claims. Moreover, we need not ever be able or required to state the metaphysical laws in a metalanguage using explicit formulations of application and coapplication conditions. However, the consideration most relevant to the explanation I just gave might be summarized in a metalanguage as follows. The application conditions for 'Socrates' do not involve antecedent association of the term 'Socrates' with the sortal 'singleton set'; so, while the urelement Socrates plays an important role in appropriate application of 'the set {Socrates}', the converse does not hold. This shows that we can explain the apparent metaphysical asymmetry between Socrates and the singleton set {Socrates} without appealing to substantive metaphysical relations.

[191] Fine (1994), p. 5.

Let me further explain by addressing a worry with the explanation I've given so far. Recall that, according to modal normativism, claims about what is metaphysically necessary are object-language expressions of constitutive rules, or their consequences, governing nonmodal terms. This means that 'necessarily, Socrates exists iff the set {Socrates} exists' either expresses some constitutive rules governing the nonmodal terms 'Socrates' and 'the set {Socrates}' or some consequence of those rules. Doesn't this mean that the meaning of 'Socrates' changes in the *impossible* circumstance that Socrates exists but not the set {Socrates}? So, by the normativist's own lights, in such a context, we are not even talking about Socrates at all?

My response is that competently evaluating counterpossibles like (18'') comes down to having a tacit ability to follow the rules of a language that is relevantly similar to the "home" language (18'') is stated in.[192] So, even if the meaning of 'Socrates' is altered by accommodating an impossible circumstance in which one of Socrates or the set {Socrates} exists without the other, the alteration is relevant and contextually negligible, and so has no bearing on the *appropriateness* of talking about the existence of Socrates in relevant hypothetical impossible circumstances. In other words, terms in our home language can have contextually determined *semantic counterparts* with their own governing rules that are relevantly similar to the rules that govern the terms in our home language.[193] A concise presentation of the idea of semantic counterparts can be found in Warren (2015) whereby two expressions α and β are semantic counterparts whenever the semantic role of an expression α in a language *L* is sufficiently similar (in the relevant

[192] Presumably not just similarity with respect to the semantic rules, but also a language homophonic to the home language and with the same rules of grammar.
[193] Warren 2015, pp. 1366, see also Warren 2016.

context) to the semantic role of an expression β in a language *K*. Warren claims that semantic counterparts "allows us more flexibility and generality when talking of meaning and concepts than a simplistic picture allowing for only semantic identity and non-identity with nothing in between."[194]

Let language *K* be a language devoid of all set terms and so devoid of the 'the set {Socrates}' and such that 'Socrates exists iff the set {Socrates} exists' is false or meaningless. Let language *L* be our home language not devoid of set terms and a language in which 'Socrates exists iff the set {Socrates} exists' is true. The idea is that, supposing that 'Socrates' in language *K* is a semantic counterpart of the 'Socrates' in *L*, then, all else being equal, the semantic role of both expressions, determined by their respective application and coapplciation conditions, will be relevantly similar. So, for any arbitrary sentence involving 'Socrates' but no set theoretic expressions, that sentence will be true in *L* iff it is true in *K*. For example, 'Socrates is a philosopher', 'Socrates is the teacher of Plato', 'Socrates has a snub nose', and '*Socrates exists*' will be true in both languages. So, contextually relevant similarities between the terms will hold, and we can still appropriately talk about the existence of Socrates in relevant hypothetical impossible circumstances. Furthermore, this explanation needn't be given in a metalanguage by home language users evaluating (18''). This can all be communicated in the object language utilizing talk of impossible worlds and counterparts: actual Socrates has a counterpart even in impossible worlds where Socrates exists but the set {Socrates} does

[194] Warren (2015), pp. 1366. See also Warren (2016).

not exist; so, if the set {Socrates} were not to exist, Socrates would still exist, but not vice versa.

Thus, I can account for the metaphysical asymmetry implied by (SOCRATES LAW) and ({SOCRATES} LAW) even when the laws are treated as object language expressions of conceptual laws and not as descriptive claims about essences or relations of metaphysical dependence.

There is a potential objection to my account of metaphysical laws coming from theoretical sides friendlier to metaphysical deflationism. Why go through all of the fuss at all? Of course, one could just be a strict eliminativist about substantive natures and substantive relations of metaphysical dependence, and so take *all* claims about essences and metaphysical dependence to be strictly false. Maybe there is a sense in which this is the right way to go.[195] But I think this may be too extreme. Dialectically, I think it would be more effective, and more interesting, to show that there is an alternative way to think about these metaphysical notions that avoids the metaphysical and epistemological worries motivating eliminativism—so, that if we want to keep such talk, we needn't have qualms about it. Of course, maybe the substantive realist will argue that I'm no longer talking about what they are talking about when they use nature and dependence language. If so, then maybe eliminativism about whatever they take themselves to be talking about is the way to go. But I think nature and dependence talk are philosophically useful and it would be better if we can keep it. When faced with questions about the ontological commitments we needed to make in order to accommodate basic modal claims, many

[195] Cf. Warren 2016.

responded with skepticism and argued for eliminativism. However, basic modal talk is useful, and it turns out that we can accommodate basic modal talk without taking on questionable ontological commitments and keeping a clear epistemology of modality. I think a similar story can be told for talk of counterpossibles, essences, metaphysical dependence, and the like, and is what I have aimed to do in this project.[196] I will return to these ideas in chapter 6.

4.4 Comparative Impossibilities

I now want to briefly address another worry that my own account of counterpossibles can't do the work I set out for counterpossibles in 1.6. Recall from 1.6 that another reason to take metaphysical impossibilities seriously is that we can often sensibly compare impossibilities. First, consider the comparison of possibilities. There are many ways that the world could be, and it seems that we often have considered modal beliefs about comparative possibilities expressed with phrases such as "more possible than," "less possible than", "could easily have been", "less easily could have been", etc.[197] For example, I *could more easily* have been a lawyer than I could have been the first person to land on the moon Europa.

Of course, there are many ways that the world could not be, and we might think that, in addition to possibilities, we have considered modal beliefs about comparative impossibilities. For example, Nolan 1997 notes that worlds where Hobbes squares the circle are far more similar to the actual world than the world where every proposition is

[196] See Nolan (2014) for an interesting historical comparison of intensionality and hyperintensionality.
[197] See. Kment (2006), p. 253-54 and Lewis (1973), p. 52.

true.[198] To put it in comparative terms: the world *could more easily* be such that Hobbes squares the circle than be such that *everything* obtains. Recall that other potential examples of non-vacuous comparative impossibilities are:

(8) Pi could more easily have lacked six consecutive '9's (i.e. the Feynman Point) than it could have been an algebraic number;

(9) Smith could more easily have had different parents than she could have been the number 5; and

(10) Smith could more easily have had different parents than she could have been born in different location than she actually was.

Prima facie, (8) and (9) claims are true while (10) is false.

Consider (9) and (10). On the one hand, it seems that less would need to change about the actual world in order for a human to impossibly have a different biological origin than it would need to change in order for a human to be an abstract number (or for an abstract number to be a human). Thus, (9) seems true. On the other hand, it seems that less, not more, would need to change about the actual world in order for some person to possibly be born in a different city than would need to change in order for her to impossibly have a different biological origin. These considerations provide some reason to think that comparative claims should not all be treated as trivially true or as all false. Or, at the very least, an explanation needs to be given of our intuitions. A substantive realist might argue that we need to posit substantive facts about grounding or fundamentality in order to account for these comparative impossibilities. So, the worry is

[198] Nolan (1997), p. 544.

that since I reject substantive metaphysical properties and relations, I cannot adequately account for comparative impossibilities.

For example, the substantive realist might think that the reason that Smith could more easily have been born in a different location than she could have had different biological parents is that substantive essential properties and grounding relations present varying degrees of resistance to making changes in reality.[199] Essential properties such as a person's biological origin are "stickier" than location properties; in fact, they are impossible to pull away from their objects. On the other hand, compared to a change in kind, biological origin properties are not as hard to pull away. Another explanation appealing to substantive metaphysical explanations might go as follows. We might think that God would have to reach deeper into the structure of reality to change something's kind but not nearly as deep to change its biological origin, much less the physical location of an event. So, Smith could more easily have been born in a different location than she could have had different parents, but she could more easily have had different parents than she could have been an abstract object.

Let me first respond by saying that I do not think that this approach is the best way to go because appealing to substantive notions of essence, grounding, or structure do not really seem to illuminate comparative impossibilities more than they are themselves illuminated. I do think that metaphors appealing to the modal strength of properties or the structure of reality are interesting, and I also think that this way of talking naturally follows from intuitions about comparative impossibilities. But I do not think that this way

[199] Cf. Kment (2006).

of talking is not very illuminating in any non-metaphorical manner. What's more, I am afraid that insofar as these are metaphors for anything other than the function or role comparative impossibilities can play in our cognitive lives and theoretical practices. If we take these metaphors seriously and think that they have non-causal and non-conceptual explanations, it seems that we will be left without a plausible story for how we have cognitive access to these substantive facts.[200]

I think that there are a number of explanations of what we are doing when comparing impossibilities consistent with the normativist picture I have developed. One way to go is that comparative claims are expressions of how comparatively entertainable the claims are. This is similar to Simon Blackburn (1993), who takes certain modal claims to express, but not describe, certain imaginative blocks we come across in our cognitive lives, e.g. as when we try to imagine that 2 = 0. However, it could be that some impossibilities are easier to imagine than others, that is, some impossibilities resist the imagination more than others. For example, I can more easily entertain being born to different biological parents, or even what it would be like to be a talking horse, than entertain what it would be like to be an abstract number.

Based on my arguments above, talk of counterpossibles and metaphysical similarity let's language users express the consequences of making changes to the current semantic normative profile of vocabulary in our "home" language but while remaining in the object language and without actually adopting those changes. I suggest that claims of comparative impossibilities are a way of expressing a willingness to adopt some changes

[200] See also the worries I outline in Chapter 2.3 and 3.4.

over others. For example, perhaps we are more willing to adopt a language governed by rules such that it is true that Hobbes squared the circle than to adopt a language govern by rules where everything is true, because the former language would *better* serve our needs. After all, a language where, for every p, both p and $\sim p$ are true doesn't seem like it would be very useful for expressing much of anything at all. Finally, it might be that a biologist can still find some use for a language in which the name for a particular can pick out something that actually belongs to one species but can belong to different species in certain imagined cases but has little use for a language in which the name can pick out things that are not even living. This might be because the former language does not deviate from her home language as much as the latter language, and so she can still get some empirical work done using the former language. On my account of metaphysical similarity, all of these metalinguistic considerations can be expressed in the object language using comparative claims in which the relevant terms are used and not mentioned.

So, my normativist account of counterpossibles and metaphysical similarity *can* account for comparative impossibilities, and so *can* do the work I outlined in 1.6. Furthermore, this account is theoretically plausible because it focuses on the function and role comparative claims play in our cognitive and theoretical lives as opposed to treating them as attempting to describe features of reality on a par with those features described by the empirical sciences. What's more, because of this, we are able to give a plausible story of how we can know the truth of some comparative claims—we only need to do conceptual analysis in our home language and languages relevantly similar to our home language, perhaps incorporating certain empirical discoveries along the way.

4.5 Summary

I started this chapter with a summary of my account of counterpossibles and metaphysical similarity for modal normativists. I previously argued that my account has certain theoretical and epistemological advantages since it doesn't take on substantive metaphysical properties such as essences and it doesn't take on substantive metaphysical relations such as grounding and fundamentality. However, I pointed out that a substantive realist might want to raise the following worry: my account can't actually do the work I set out for counterpossibles in section 1.6. If my account could not do the work I set out for it at the beginning of this dissertation, then we might have been stuck with two responses. On the one hand, maybe I was wrong and there really isn't any interesting work for counterpossibles to do after all. In that case, we might have to reconsider the theoretical reasons we have for talking about counterpossibles and metaphysical similarity, which might mean we should reconsider Williamson's arguments against non-vacuous counterpossibles. If they aren't that useful, why care about whether they are all true? Or, even if we set aside the issue of vacuity, if counterpossibles are just silly sentences, why even bother with giving an account of when they are assertible? On the other hand, maybe I was right about the work we need counterpossibles to do, but my account can't handle the workload. In that case, we might have to reconsider the plausibility of the substantive approach to counterpossibles and metaphysical similarity since it presumably can do the work. Neither response is an attractive option given what I have set out to do in this project. Fortunately, in this chapter, I was able to argue that my account of counterpossibles and metaphysical similarity can do the work and do it in a clear and plausible manner.

I first looked at the purported differences in content and force between indicative and counterfactual conditionals primarily about metaphysical matters. I looked at a number of examples of metaphysical dependence commonly discussed in the literature on grounding, and I argued that the modal normativist can account for the different roles the mood of a conditional can play in giving evidential versus metaphysical explanations. Then I looked at the claim that counterpossibles reflect the asymmetry of claims of metaphysical dependence, which substantive realists might argue is best explained by substantive relations of grounding or fundamentality. However, I argued that my account of counterpossibles and metaphysical laws gets the job done without having to posit substantive metaphysical properties or relations. Finally, I looked at the argument that the best account of comparative impossibilities is that reality is structured in ways that make certain metaphysical changes more difficult than others; so, we so need to posit substantive metaphysical relations after all; thus, my normativist account can't do the work I said a theory of counterpossibles ought to be able to do. Once again, I argued that this is not a worry for my view, since my theory of counterpossibles can do the work, and can do the work in a clear manner that emphasizes the function and role of comparative impossibilities in our theoretical and cognitive lives.

CHAPTER 5: NEW WORK FOR A THEORY OF COUNTERPOSSIBLES

So far, I have argued that talk and considerations of counterpossibles and metaphysical similarity has an important role to play in our theoretical and cognitive lives. In that case, we need to give a clear plausible account of counterpossibles and metaphysical similarity that enables counterpossibles to fulfill this role. I then explained and responded to the objection that my account cannot do the work that I set out for counterpossibles, and that maybe we should reconsider using resources from more substantive metaphysical theories, e.g. theories of grounding or fundamentality. However, recall from the end of section 4.3 that there is also an objection to consider from a theoretical perspective friendlier to the overall deflationary approach I am taking in my own account of counterpossibles and metaphysical similarity. Maybe we shouldn't really care about debates over essences, over the nature of things, over metaphysical dependence, etc. Many of these debates are intractable, and maybe the best explanation is that people are simply playing word games when debating about the essential properties or natures of things, e.g. when debating about the real nature of Socrates, or when debating about the real nature of free will. If that's the case, then maybe talk of counterpossibles and metaphysical similarity isn't theoretically interesting and not worth all the theoretical fuss I've been making in this dissertation.

In this chapter I will respond to this worry. Again, I think that a project showing that there is an alternative way to think about counterpossibles and related ideas, such as essences, dependence relations, metaphysical similarity, etc., that avoids the metaphysical and epistemological worries *motivating* an eliminativist attitude is more dialectically interesting than adopting eliminativism about those ideas. The main reason is that talk of

counterpossibles and related ideas doesn't seem to be mere nonsense and plays an important role in our cognitive lives, e.g. by serving as object language illustrations of how the rules governing various of our terms can come apart, e.g. the rules governing 'Socrates' versus the rules governing 'the set {Socrates}'. So far, I have provided and argued for a plausible alternative way to think about counterpossibles and related issues. What's more, I argue in this chapter that there is interesting new work for a theory of counterpossibles that comes from recent literature on conceptual engineering, including from literature by those conceptual engineers that are more inclined to a deflationary metaphysics. Given this new work for a theory of counterpossibles, talk and consideration of counterpossibles and metaphysical similarity is *not* nonsense and *is* worth all of the theoretical fuss. So, there is serious work for the account of counterpossibles I have offered here. Furthermore, I argue that my theory can do the work and does it well.

In the next section, I will discuss recent literature on conceptual engineering, conceptual ethics, and metalinguistic negotiation. I will discuss certain limitations of more traditional approaches to conceptual analysis that motivate a new way of thinking about the role conceptual analysis plays in philosophical theorizing. For example, many proponents of conceptual ethics have emphasized the way it can be used to make sense of seemingly intractable disagreements found in many metaphysical disputes.[201] It may turn out that many metaphysical disputes reflect disagreements at the conceptual level rather than disagreements over substantive matters about the nature of reality. However, if that is the case, we need not interpret disputants in metaphysical debates as unknowingly

[201] Cf. Thomasson (2017) and Plunkett (2015).

talking past one another or merely involved in verbal disputes. After all, what concepts we use has implications for our theoretical projects and even moral and practical consequences for how we live. What's more, this means that conceptual engineering can not only make sense of our conceptual lives, but it also has the potential to provide a plausible explanation of many metaphysical disputes as well as a plausible epistemological story of how those disputes can be resolved.[202]

In section 5.2, I will discuss the roles basic modal claims play in conceptual ethics. Then, in sections 5.3-5.5, I will discuss the important work counterpossibles can do in conceptual analysis and conceptual ethics. First, counterpossible conditionals and claims comparing the metaphysical similarity of impossible worlds with the actual world allow us to advocate for or against certain conceptual dependencies between many metaphysical concepts while remaining in the object language. Second, counterpossibles allow us to work out or negotiate the consequences of certain conceptual changes or practices without actually adopting those changes or practices. Finally, I will show how counterpossibles can even do work for those conceptual engineers who take a more conventionalist approach to metaphysical issues. Overall, instead of weighing the costs and benefits of adopting various alternative conventional practices in a metalanguage, counterpossibles allow us to think about and reason about the consequences of alternative conventional practices while remaining in our own object language. In sections 5.6, I will explain how my account of normativist counterpossibles can do the work provided by conceptual engineering.

---

[202] See Thomasson (2017).

5.1 Conceptual Engineering and Metalinguistic Negotiation

Though recent literature on *conceptual engineering* often emphasizes the way it is distinct from conceptual analysis, I think that is helpful to frame a collection of related conceptual projects together under the former heading. Consider the following quote from Blackburn (1999):

> I would prefer to introduce myself as doing conceptual engineering. For just as the engineer studies the structure of material things, so the philosopher studies the structure of thought. Understanding the structure involves seeing how parts function and how they interconnect. It means knowing what would happen for better or worse if changes were made. This is what we aim at when we investigate the structures that shape our view of the world. Our concepts or ideas form the mental housing in which we live. We may end up proud of the structures we have built. Or we may believe that they need dismantling and starting afresh. But first we have to know what they are.[203] (1)

There seem to be at least three distinct but related conceptual projects that can be pulled from this quote. First there is a descriptive project of *conceptual analysis*, which involves explicating and understanding what our *actual* concepts are, i.e., how they are actually used and how they interrelate and fit together. In addition, on some pictures of philosophical analysis, conceptual analysis plays an important role in helping to locate potential satisfiers of those concepts in the world.[204] Second, there is the project of *conceptual development*, which can be seen as a more experimental project of creating new concepts, as well as revising our current concepts.[205] At this point, the question of the usefulness or appropriateness of the new or revised concept will still be open. But, finally, there is a normative project of *conceptual ethics*, which involves asking

---

[203] Also quoted in Tanswell (2017). Tanswell disagrees that we need to or even can understand our concepts fully before embarking on reconstructing our concepts, p. 12.
[204] One such prominent picture is the Canberra Plan, which is thoroughly discussed in Braddon-Mitchell, D. and Nola, R. (2009).
[205] For example, Haslanger (2012).

normative questions about whether we *should* even be using certain concepts at all or asking about whether we *should* adopt a related but different concept, given our goals and values.[206] Given these three projects, there is a host of important descriptive and prescriptive conceptual questions we can ask:

(a) What pieces of syntax are actually associated with which concepts?

(b) What pieces of syntax *should* be associated with which concepts?

(c) What, if anything, is there in the world that satisfies our concepts?

(d) What concepts *best* fit the way the world is (or *really* is)?

(e) What concepts do we actually employ in our theoretical and practical lives?

(f) What concepts *should* we be employing in our theoretical and practical lives, given certain goals and values?

(g) What sorts of theoretical or practical consequences would follow, were we to change our concepts or adopt different concepts?

For the most part, I will focus on the work counterpossibles can do in answering questions (f) and (g), though the discussion will sometimes relate to the other questions. However, in order to set up why anyone might appeal to conceptual ethics to deflate certain metaphysical disputes or adopt a pragmatic approach to conceptual engineering, I want to discuss the limitations we encounter by taking conceptual analysis to solely focus on questions (c) and (d).

[206] For examples see Haslanger (2012), Burgess and Plunkett (2013), Plunkett (2015), and Thomasson (2017).

Much of the recent literature on conceptual engineering has focused on the project of conceptual ethics.[207] Some proponents of conceptual ethics have emphasized the way it can be used to make sense of the intractable disagreement found in many metaphysical disputes. It may turn out that many metaphysical disputes reflect different views about conceptual matters more than reflecting different views about the substantive metaphysical facts of the world. Furthermore, they have emphasized how thinking about many metaphysical disputes in terms of conceptual ethics can provide a plausible epistemological story of how those disputes can be resolved. To better understand the motivation for undertaking conceptual ethics and adopting a deflationary attitude towards these metaphysical disputes, we need to understand why there are limitations to using conceptual analysis to resolve these disputes.

Consider one interesting picture of conceptual analysis called the 'Canberra Plan'.[208] Nolan (2009) suggests that, according to the Canberra Plan, when trying to give a philosophical analysis of a subject matter such as the self, free will, the mind, causation, justice, virtue, etc., we proceed in two steps. In the first step, we attempt to formulate a theory of the subject matter—perhaps by collecting the platitudes surrounding the subject matter—and use that theory to develop a *functional definition* of the subject terms pertinent to the theory. In the second step, we investigate the world, or look to our current best theories of the world and *locate* what it is that that satisfies the functional definition provided by the first step.[209] For example, we might collect all the obvious truths about water such as: it is a clear, potable liquid that falls from the sky as rain, and

---

[207] Ibid.
[208] See Jackson (1998) and Braddon-Mitchell, D. and Nola, R. (2009).
[209] See Nolan (2009).

is found in rivers, lakes, and oceans. We then use this collection as a functional definition of 'water'. Then we investigate the world to find whatever it is that satisfies our functional definition, which, in the actual world, is the chemical compound H_2O.[210]

However, there are limitations to using this overall approach to conceptual analysis to resolve certain philosophical debates. One limitation, pointed out by Nolan (2009), comes from trying to use this method to give an analysis of subjects that seem to be metaphysically fundamental. While the Canberra Plan can be used to provide physicalist answers to many philosophical questions, e.g. answers to questions about the nature of the mind or the nature of ethical properties, it still must rely on ontological and modal assumptions about what exists or what is possible according to some preset scientific or philosophical background theory. For example, one can come up with a functional definition of mental terms, e.g. 'belief', 'desire', etc., but nothing prevents one from taking certain non-physical properties as primitive in their background theory of the world and arguing that those are what satisfy the functional definition of our mental terms.[211] So, if one wants to be a strict physicalist about the mental, additional assumptions need to be made to rule out certain ontologically problematic entities, e.g. that physics alone provides an adequate explanation of the causal role specified by the functional definition of mental terms.[212]

These difficulties mostly stem from the second step: locating something in the world, or in our best theories of the world, that satisfies the functional definitional

---

[210] On versions of the Canberra Plan where we take it to be analytic that water just is whatever satisfies the functional definition, we can even take it to be necessary that water just is whatever actually satisfies the functional definition, i.e. that necessarily water is H_2O.
[211] See Braddon-Mitchell (2009) and Nolan (2009).
[212] Ibid.

provided by the first step. However, it seems that many debates in philosophy are about seemingly fundamental questions, e.g. about whether our fundamental ontology consists only in physical properties, or whether there are nonphysical properties at the fundamental level, or whether certain properties or entities are even possible to begin with, e.g. immaterial souls or philosophical zombies. So, the Canberra Plan might not be of much use in analyzing more fundamental concepts in philosophy such as 'meaning', 'causation', 'personal identity', 'justice', 'negation', 'free will', etc.[213]

Some versions of the Canberra Plan tie conceptual analysis to the view that many of the expressions and concepts to be analyzed are descriptive in nature and so will be satisfied by some feature of the world.[214] But, there is a limitation to descriptive versions of Canberra Plan that comes from theoretical pressure to think that, in many important philosophical debates, e.g. those regarding free will, personal identity, mereological sums, essential properties, etc., we cannot reasonably hope to locate some unique thing in the world, or in our best theories of the world, that satisfies the functional definition provided by the first step. This is because completing the descriptive project will require epistemically suspect methods that go beyond pure conceptual analysis and straightforward empirical methods of inquiry.[215] For example, with certain empirical concepts, e.g. 'gene' in biology, we can use reliable empirical methods to investigate the world and find what it is that best satisfies the relevant functional role developed in the first step of our analysis. However, the problem with many metaphysical debates—e.g.

---

[213] For the most part, I will not distinguish between terms and concepts, and I will use '*p*' to refer to both the term and the concept. However, in the few places the distinction matters to the point I am making, I will use small caps for the concept, e.g. I will use '*p*' to refer to the term the small caps P to refer to the concept.
[214] See Jackson (1998).
[215] See Thomasson (2017).

debates about whether ordinary objects such as tables exist, whether free will is compatible with determinism, whether negation is classical or intuitionist—is that there aren't clear empirical considerations that will help us resolve the debate since, in many cases, the competing philosophical theories are all empirically adequate and are such that, according to both sides of the debate, no merely empirical matter could decide between them.[216]

Another difficulty with descriptive versions of conceptual analysis is that it presupposes *semantic functional monism* about our terms in that it presupposes that the function of most, if not all, meaningful expressions in our language is descriptive in that every term functions to 'pick out' some 'thing' in the world. However, Sellars (1958), Price (2011), and Thomasson (2015) all give convincing reasons to accept a version of *semantic functional pluralism* where various expressions, e.g. prescriptive expressions, modal expressions, ordinary empirical expressions, etc., all have more or less different roles to play in our linguistic (and other, e.g. social,) practices, and no one role is better than another independently of what we are trying to use those expressions to do.

Given the difficulties of agreeing on or finding the right features of the world that satisfy many of our concepts, some propose that we give up the project of giving a descriptive analysis of those concepts and focus instead on the prescriptive elements of conceptual engineering, i.e. on conceptual ethics.[217] Even if descriptive analysis is a hopeless project for many concepts, we still need to decide whether we *should* keep the concepts we currently employ, e.g. we need to decide if we should keep or eliminate

[216] See Thomasson (2015), pp. 14-17.
[217] See Eckland (2014), Simion (2017), and Thomasson (2017).

certain concepts such as ‘race’, ‘free will’, ‘truth’, etc. Moreover, if we decide to keep some version of the concept, we still need to decide what the *best* concept is that will help us meet our theoretical and practical needs, e.g. if we decide to keep race concepts, then which race concepts should we be using?[218] However, even if we place our focus on conceptual ethics, this still leaves a lot of work for conceptual analysis and conceptual development. It is still important to engage in conceptual analysis to determine what our *actual* (perhaps ordinary) concepts are. This way, given a concept, we have a better idea of exactly what role that concept plays, so that we can decide whether we should keep or change the concept. It is still important to engage in conceptual development so that we can construct alternative concepts and understand the consequences of our new or changed constructions and how new or changed concepts fit into our overall conceptual practices.

One final point before moving on. In recent literature, it has been argued that much of conceptual ethics takes place, not at the level of the metalanguage but, at the level of the object language in the form of *metalinguistic negotiations*. Plunkett (2015) notes that, even for ordinary expressions, we sometimes use, rather than mention, the expression to communicate aspects of the meaning of the expression. For example, when communicating how to apply the predicate ‘tall’ in a context, instead of saying, “Apply the term ‘tall’ just in case so-and-so is at least approximately this height,” where the predicate is mentioned in quotes, we might instead use the predicate via an example: “Around here, James is tall,” indicating that to properly apply the predicate ‘tall’ the subject’s height must be at least roughly the height of James, or higher. So, a

[218] See Haslanger (2012).

metalinguistic negotiation takes place when agents implicitly press for views about terms we *ought* to use, or how we *ought* to use them, by *using* terms as opposed to mentioning them.

Thomasson, Plunkett, and Sundell, have extended this idea to philosophical debates more generally.[219] Plunkett argues that there are likely some disagreements in philosophy where disputants take themselves to be engaged in a debate about object level metaphysical issues about the nature or grounds of some aspect of reality, e.g. personal identity, free will, mereological sums, etc., when instead the disagreement is best read as a disagreement over what concepts we should be using. Thomasson (2017) argues that proponents of a deflationary interpretation of metaphysical disputes can appeal to metalinguistic negotiation to explain why the disputants are not just talking past one another or uttering trivial truths and obvious falsehoods—instead they are using the relevant terms to push for using one concept over another or maybe even pushing to get rid of the concept altogether.

Let *metalinguistic engineering* be conceptual analysis, conceptual development, or conceptual ethics that takes place in the object language using rather than mentioning the expressions being analyzed. As I will next show, counterpossibles can do a lot of important work in metalinguistic engineering. To briefly summarize, counterpossibles allow us to: advocate for or against certain dependencies between our concepts while staying in the object language; and work out or negotiate the consequences of certain

[219] See Thomasson (2017), Plunkett and Sundell (2013), and Plunkett (2015).

conceptual changes or practices without actually adopting those changes or practices or even explicitly talking about those changes or practices.

Let me highlight some reasons why engineering concepts in the object language is helpful. The first is that we may not even be able to explicitly define our concepts and conceptual relations in a metalanguage to begin with.[220] In that case, we will not be able to explicitly weigh reasons for or against adopting certain concepts or connections between concepts. So, we may have no choice but to negotiate from the object language. Second, as we will see in 5.5, explicitly reasoning about alternative conceptual practices can create unneeded confusion. Finally, engineering in the object language can help keep the focus on the theoretical and practical consequences of concept use and concept change as opposed devolving into word games.

5.2 Basic Modal Expressions in Conceptual Ethics

Before discussing counterpossibles, I will briefly look at the role basic modal expressions, e.g. 'necessary', 'possible', 'impossible', etc., can play in conceptual development and ethics. Suppose somebody claims that 'necessarily, *p*' or '*p* is impossible'. How to literally read these claims is itself a rich and interesting philosophical topic. One option is to say that these claims are describing modal features about the world, e.g. there something about the metaphysical nature or structure of reality that blocks contradictions from being true or prevents a person from having different biological parents than they actually have.[221] Another option is to be a concrete realist about worlds such that the above basic modal claims are just different ways of allowing

[220] See Thomasson (2007a, 2015) and Chalmers and Jackson (2001).
[221] Kripke (1980), Kment (2014).

us to say things like every world that exists is such that contradictions are not true, or there do not exist worlds such that a person who exists in this world has different biological parents.[222]

However, we need not take basic modal claims to be describing anything at all. For example, recall that Sider (2011) and Cameron (2010) argue that the metaphysically possible/impossible distinction is not a natural distinction, i.e. that it does not carve reality. So, on these views, basic modal claims are simply making classifications. Other approaches skeptical of the literal readings take basic modal claims to be expressive or prescriptive rather than descriptive. According, to some of these views, basic modal claims are still truth apt, but they are neither describing modal features of the world nor are they merely classifying or categorizing truths about the world. For example, according to some expressivist views, modal claims such as '*p* is impossible' express certain imaginative limitations.[223] According to modal normativism, modal claims are not descriptive claims in need of modal truthmakers but instead are object-language expressions of constitutive rules governing nonmodal terms.[224]

Setting aside the issue of what the literal content of basic modal claims may or may not be, or the issue of what speakers take themselves to be communicating with such claims, basic modal claims can play various roles in conceptual ethics and metalinguistic negotiations. One role is simply to further emphasize one's normative views about the words and concepts under dispute. For example, when giving an example of a

[222] Lewis (1986). There are also modal rationalist views, e.g. Chalmers (2002), that take modality to be tied up with conceivability or with apriori conceptual features.
[223] Blackburn (1993).
[224] Thomasson (2010, 2012). See also Brandom (2008).

metalinguistic dispute over 'free will', Plunkett (2015) considers how a incompatibilist might respond to a compatibilist:

> It is part of the essence of free will that only agents that are capable of fully causing their own actions have free will. This means that their actions cannot be fully determined by events over which they have no control, such as the events of the past. (855)

The incompatibilist isn't merely saying that free will is *actually* precluded by determinism but that it is *impossible* for an agent's action to be determined and for that agent to fully cause her actions. Moreover, the incompatibilist claims that this is part of the *essence* of free will. Interpreting the incompatibilist as involved in a metalinguistic negotiation, the modal expressions 'impossible' and 'essence' can be read as linguistic devices used to emphasize the incompatibilist's views about how we *should* structure our concept 'free will', e.g. that it *should* be central to the definition of 'free will' that there is agent causation, and that our actions are not fully determined by physical and psychological causation.

Another role basic modal expressions can play is to advocate for views about how our modal space should be structured more generally. Again, one way to read the basic modal claim '*p* is impossible' is literally, e.g., as a descriptive claim about the modal structure of reality. As discussed in the previous paragraph, an alternative way is to read '*p* is impossible' as metalinguistically emphasizing one's normative views about the nonmodal words and concepts under dispute, i.e. '*p*'. However, on another metalinguistic reading, '*p* is impossible' can be interpreted as negotiating for the exclusion of *p* worlds for our logical space, i.e. that we should not include *p* worlds in our logical space, and so we should not be allowed to rely on *p* claims when making certain arguments. For example, when someone claims that *philosophical zombies*—minimal physical duplicates

of actual people without conscious experiences—are conceivable but impossible, they need not only be seen as metalinguistically advocating for how we should use certain mental words and concepts, but also as advocating for views about how we should think about the relationship between conceivability and possibility.

Yet, given the metalinguistic interpretation of basic modal expressions, I think there is some pressure on how we can literally read basic modal claims. For example, on a realist account, when one says, 'p is impossible', they are literally making a descriptive claim about the *real* nature of our modal space or about how our modal space is *really* structured. For example, according to concrete modal realism, either there is a *p* world or there is not, no matter what you think should be the case. So, it isn't clear to me what sense it would make to accept a version of concrete modal realism but then argue that one is uttering 'p is impossible' as a way of advocating for what the modal space underlying our theoretical and philosophical activities should be like. On the other hand, classificationist, expressivist, or normativist views will have the theoretical flexibility to make sense of both the literal and metalinguistic readings of basic modal claims.[225] I will discuss this and a related concern further in section 5.6.

5.3 Counterpossibles and Negotiating Conceptual Dependencies

The first way that talk of counterpossibles and metaphysical similarity between impossible worlds can help in conceptual ethics and conceptual development is by using them to normatively advocate for or against views that there are certain dependencies

[225] How does this interact with axioms of S4? If you reject $\Box p \rightarrow \Box\Box p$, then you could be interpreted as acknowledging that perhaps $\sim p$ worlds are currently precluded from our modal space but that we should re-center on a world where $\sim p$ is possible. But I think that you can keep S4 and simply say that you are advocating for an entirely different modal space, and entirely different space of possible worlds, all together.

between our concepts. For example, imagine a nominalist engaged in debate with a Platonist about the existence of numbers making the following claim:

(6) If numbers were not to exist, then the physical world would be exactly as it is, i.e. all of the regularities between the microphysical particles and their manifestations would be exactly the same.[226]

Supposing that numbers necessarily exist, (6) is a counterpossible. We need not interpret the nominalist and Platonist in this example as only involved in an object level dispute about the existence of numbers or their relationship with the physical world. We can also interpret them as engaged in a *normative* dispute about how we *should* think about the conceptual or inferential relationships between claims about number and claims about the physical world. Namely, the nominalist is using (6) to metalinguistically negotiate for a language where terms used in the physical sciences are conceptually independent of mathematical terms. For example, the nominalist is advocating for the frame-level content that governs the use of our physical terms to be independent of the frame-level context that governs the use of mathematical expressions. In this example, it is the *entire counterpossible* that is under dispute.

This metalinguistic use of counterpossibles is more forcefully illustrated by considering an example where the relations among our concepts have significant practical and ethical consequences. Imagine two ethicists engaged in a dispute about divine command theory (DCT). While trying to argue against DCT, one of the ethicists claims:

(19) If God were to command people to murder, then murder would still be wrong.

[226] For example, see Liggins (2014).

Suppose that both parties agree that since God's nature presumably precludes God from commanding people to do evil things, it is impossible for God to command people to murder, yet they disagree about (19)—the defender of DCT thinks that it is false while the critic thinks that it is true. Let me put the point in terms of a thesis about worlds:

> *Theological Independence*: Impossible worlds where God commands people to murder and murder is still wrong are more relevantly similar to the actual world than impossible worlds where God commands people to murder but murder is morally permissible.

Again, the defender of DCT thinks that *theological independence* is false while the critic thinks that it is true. Moreover, they might take themselves to be involved in an object level dispute about the nature of morality.

On a substantive realist account of the disagreement, the disputants are involved in a literal disagreement over a descriptive claim that reflects substantive facts about what metaphysically grounds moral wrongness. On this interpretation, whether *theological independence* is true, and thus (19) is true, depends on whether moral wrongness is actually grounded in God's commands. So, the literal disagreement is about what is fundamental or about the grounding structure of reality.

On my view of counterfactuals, (19) is an object level expression of a requirement to infer 'murder is wrong' given 'God commands people to murder' and a range of relevant auxiliary assumptions, an inferential requirement that constitutes, and so is backed by, the actual frame-level contents of 'God', 'wrong', and 'murder'. *Theological independence* is an object level claim that expresses the same thing. So, at the level of the

terms, we can interpret the disagreement over either (19) or *theological independence* as a disagreement over how certain of how the frame-level content of moral and theological concepts are actually related, e.g., whether moral concepts are conceptually downstream from certain theological concepts. But, again, this disagreement need not be considered or expressed in a metalanguage but can be expressed as a disagreement over object level claims such as 'the wrongness of murder is not grounded in God's commands'.

Some might argue that there isn't really any disagreement between the critic and defender of DCT and interpret the apparent dispute as purely terminological. On this reading, the defender of DCT rejects (19) because of how she prefers to use the relevant terms, i.e. 'God' and 'wrong', and the critic of DCT accepts (19) because of how she prefers to use the relevant terms. So, literally speaking, the disputants are speaking past one another. However, there are two reasons to think that this is the wrong interpretation. First, we can easily imagine that the disputants take themselves to be using the terms 'God' and 'wrongness' in the same way. Second, while they may be mistaken about the exact nature of their disagreement, the disputants might legitimately take themselves to have a real disagreement of some kind. So, the disputants are involved in disagreement of some kind. Furthermore, we need not interpret the disputants as disagreeing over substantive grounding relations or the actual conceptual connections between our theological and ethical terms, but instead disagreeing about what sort of concepts we *should* be using.

According a metalinguistic negotiation interpretation of the disagreement over (19), which I will assume comes down to a disagreement over *theological independence*, the critic is uttering (19) to normatively advocating for ethical concepts such that

circumstances where God commands murder but murder is still morally wrong are more relevantly similar to the actual circumstances than circumstances where God commands murder, but murder is morally permissible. That is, the critic is uttering (19) to normatively advocate for ethical concepts that are independent of theological concepts. On my preferred view, the critic is uttering (19) to implicitly push ethical terms that are governed by rules of use that are independent of theological concepts.

There are three reasons that support the metalinguistic negotiation interpretation of the dispute over (19) and *theological independence*.[227] First, there are likely no other object level discoveries or arguments about God, wrongness, or murder that will help resolve the dispute one way or the other. Second, if the disputants both come to believe that the *actual* frame-level content of 'God', 'murder', and 'wrong' *actually* does not support a counterpossibly robust inference from 'God commands people to murder' to 'murder is wrong', they will continue to disagree. In particular, the critic of DCT will still defend *theological independence*. Finally, the disagreement would likely persist even if the disputants were to realize that they are using one or more of the terms in different ways, e.g. the term 'wrong'.

So, the critic of DCT can be interpreted as *using* (19) to metalinguistically press for the normative view that, perhaps given certain goals or values, we *ought to* use moral concepts that do not depend on theological concepts (even if it is discovered that this how our concepts in fact already work). What concepts we use matters because whether (19) is accepted not only matters to philosophical theorizing about the nature of moral

[227] These are the hallmarks of a metalinguistic dispute outlined in Plunkett (2015) and Thomasson (2017).

properties but also has very serious practical implications to society more generally. The conceptual interrelations between theological and ethical concepts have real consequences for what reasons and motivations are acceptable for proposing or enforcing certain social rules or laws. It matters in deciding whether a person's religious beliefs are sufficient reason for them to claim not to recognize or to understand the legal and moral rights of people who do not share their views, e.g., about gay marriage.

One might object that whether we treat counterpossibles as vacuous or nonvacuous has no bearing on the role they play in conceptual ethics just described. Recall that, on the orthodox view of counterpossibles, the literal semantic content of a counterpossible utterance is trivial truth, while on the unorthodox view the content is non-vacuous insofar as the counterpossible may be true or false. This distinction is illustrated by observing that in the same DCT context

(19*)If God were to command people to murder, then two and two would sum to zero

is still trivially true, and hence vacuous, for the orthodox, while likely false for the unorthodox (since the dispute is not about the connection between God's commands and mathematics and so the impossibility in which God commands murder and two and two would sum to zero are irrelevant in the context). Still, nothing prevents a proponent of the orthodox view from agreeing with the metalinguistic reading that what is being communicated with (19) is not the literal content of the counterpossible but instead that we should treat moral concepts as being independent of certain theological concepts. So,

the semantic debate about the vacuity of counterpossibles comes apart from the role counterpossibles play in conceptual ethics.[228]

In chapter 1, I argued that the main arguments against non-vacuous counterpossibles are unconvincing and that an account of counterpossibles that meets all three of the *desiderata* I set out in in the introduction would be better than one that does not. However, I argued that even if counterpossibles are all trivially true, we still need and should provide a general account of what we are doing when we make counterpossible claims and an account of when it is appropriate to do so. For example, a defender of DCT who uttered (19*) in response to the critic might be interpreted as attempting to end the discussion outright because they take the connection between the nature of God and morality to be so fundamental that to question it is beyond absurdity. Instead, it would be better for the defender of DCT to give a less *ad hoc* response and give reasons for why (19) is not appropriate. For example, maybe there are reasons to think that the closest worlds where God commands people to murder are still the best of all the nearest impossible worlds and so certain cases of murder are impossibly morally acceptable in those worlds. But this requires some general account of counterpossibles underwriting the explanation.

So, one role counterpossibles can play in conceptual engineering is to normatively advocate for or against certain conceptual dependencies between our terms while using those terms in the object language. In this case, it is the *entire* counterpossible that is under dispute. However, another way counterpossibles can help in conceptual

[228] Even Williamson seems to agree that counterpossibles can be used to make pragmatic points. See Williamson (2017).

engineering is to work out or negotiate the consequences of certain conceptual changes or practices without actually adopting those changes or practices. I will spend the next two sections illustrating this use.

5.4 Counterpossibles in Metalinguistic Negotiations

In this section, I will argue that counterfactuals and counterpossibles allow us to work out or negotiate the consequences of certain conceptual changes or practices without actually adopting those changes or practices. Very often instead of weighing out the consequences of making certain changes to our linguistic practices, e.g. changes in the rules that govern the use of our terms, in a metalanguage where the terms are *mentioned*, we can *use* the terms in counterfactuals and counterpossibles to consider those consequences.

Recall that in a metalinguistic negotiation the utterers of metalinguistic claims are not primarily communicating the literal content of the claims but are instead normatively advocating for or against certain linguistic changes all while remaining in the object language. Moreover, the disputes are not merely terminological, since how a dispute is settled will often have real practical consequences, e.g. consequences for certain social practices relevant to the term or concept. Consider, for example, two people in a heated debate about whether the racehorse Secretariat is an athlete.[229] According to one interpretation of the debate, they are literally debating whether Secretariat is an athlete, perhaps by appealing to empirical facts about Secretariat or facts about what they take to be the nature of athleticism. According to another interpretation, they are involved in

[229] See Plunkett and Sundell (2013) and Plunkett (2015).

conceptual analysis at the level of the object language and are in disagreement about out whether our actual concept ‘athlete’ includes some, or precludes all, nonhumans. According to an interpretation of the debate as a metalinguistic negotiation, the disputants are each advocating for the way they think the term ‘athlete’ ought to be used, but they are doing so by *using* the expression in the object language rather than mentioning it.

On the view of metalinguistic negotiation put forward in Plunkett (2015), even if the parties to the dispute associate the term ‘athlete’ with different concepts, e.g. ATHLETE* vs. ATHLETE, so that each is speaking a trivial truth in her own language, the disputants need not be interpreted as talking past one another, i.e. as involved in a verbal dispute. Instead, we can interpret them as each pressing for her respective view about which concept the term ‘athlete’ *ought* to be associated with but by *using* the expression in the object language rather than mentioning it and using the term in line with the concept they think the term ought to be associated with.

On the view of metalinguistic negotiation put forward in Thomasson (2017), even if there already exist conceptual practices that determine the appropriate use of the term ‘athlete’ such that one party is uttering trivial truths by her use of ‘athlete’ with respect to those practices and the other trivial falsehoods, they need not be interpreted as making uninformative and pointless claims. Instead, we can interpret them as each pressing for their respective views about how the term ‘athlete’ *ought* to be used but by *using* the expression in the object language rather than mentioning it. In other words, they are implicitly pressing to change or reinforce the actual rules that govern the term ‘athlete’ by using the term in accordance with the rules they think should be adopted.

On either view of metalinguistic negotiation, the disputants are neither trying to communicate facts about the world nor are they trying to communicate facts about how the term 'athlete' is actually used. Instead, what the disputants are *doing* is pushing for a change in linguistic practices based on how they think we *ought* to use the expression 'athlete'. Of course, they might not just be using terms under dispute to push for changes in our linguistic practices but can also be seen as the terms under dispute to push for changes in other social practices, e.g. they are use terms the terms to implicitly advocate for or against treating Secretariat as worthy of the kind of acclaim and respect we give to human athletes. After all, how the dispute is resolved will not only have consequences for how we speak, e.g. maybe it will turn out that we can sometimes apply the term 'athlete' to horses, but also for other social practices, e.g. maybe it will turn out that we can include Secretariat in whatever athletic hall of fames that have been previously reserved for human athletes. Thus, when we interpret the dispute as a metalinguistic negotiation, the disputants are neither merely talking past one another nor berating one another with trivial truths and obvious falsehoods.

Focusing on Thomasson's view of metalinguistic negotiation, one might wonder what happens when the metalinguistic claims made by a person uttering trivial falsehoods with respect to the actual rules that govern the use of the term are embedded in conditional statements.[230] In the case of a counterfactual, an embedded falsehood will mean that the disputants are now potentially uttering a counterpossible. This will

[230] More generally, one might worry that since conditional statements are often taken to be force stripping, the metalinguistic claims will lose their pragmatic force leaving only the literal content of the antecedent and consequent to make sense of the conditional claim being made. However, this is not quite the right way to look at it. Thomasson argues that instead of thinking that disputants are trying to convey some additional informational content with her claim, it is better to think of them as trying to reinforce or change the way the terms are used. See Thomasson (2017), pp. 23-24.

certainly be the case if we try to interpret disputes over claims that are purportedly necessary in metaphysics and logic as metalinguistic negotiations. For example, if we interpret classical and intuitionist logicians as involved in a metalinguistic dispute over the meaning of negation, e.g. whether '¬¬p → p' should be a theorem of a language, or two metaphysicians as involved in a metalinguistic dispute about whether a person survives a teletransporter. In both cases, all parties are involved in a debate about claims that are typically taken to be metaphysically necessary. So, some counterfactuals used in metalinguistic negotiations will involve metaphysical impossibilities, e.g. 'if Jones were to survive (not survive) the teletransporter, then her family would not (would) be entitled to their inheritance'. In the previous section I argued that we can use certain counterpossibles to metalinguistically advocate for or against certain dependencies among our concepts. However, I will argue that counterfactuals and counterpossibles can also be used to work out or negotiate the consequences of certain conceptual changes or practices without actually adopting those changes or practices.

Thomasson (2017) briefly reviews Carnap's distinction between internal existence questions, which are existence questions asked within a linguistic framework, and external existence questions, which are existence questions attempted from outside a linguistic framework. Carnap argued that many external existence questions taken literally, i.e. taken as factual questions, are not well-defined. Nonetheless, external existence questions can sometimes be charitably interpreted as practical questions about whether we *should* adopt the linguistic framework in question. Of course, the distinction

between internal and external *existence* questions can be generalized to include other philosophically important questions such as modal questions.[231]

Thomasson argues that metalinguistic negotiations are one way of understanding Carnap's interpretation of external questions as pragmatic rather than factual questions.[232] On Thomasson's preferred interpretation of the distinction between internal and external questions, internal questions are those questions asked *using* the relevant terms of a linguistic framework, and these questions can be answered by straightforward conceptual analysis or empirical means. As for external questions, those that are asked by *illicitly* using terms independently of a linguistic framework are not well-formed. However, Thomasson (2015) argues that we can charitably interpret some external questions as pragmatic questions that *mention* the relevant terms in order raise a pragmatic question about whether the terms with the associated rules of use should be adopted.[233] However, Thomasson (2017) argues that the idea of metalinguistic negotiations provides yet another way to interpret philosophical disputes seemingly over some external question. On this interpretation, the disputants are indeed using the term under dispute, but we need not interpret them as illicitly trying to use the relevant terms to make factual claims independent of a linguistic framework. Instead, we can interpret the disputants as implicitly pressing to change or reinforce the actual rules that govern the use of the terms in question by using the terms in accordance with the rules they think should be adopted.

I think that very often instead of weighing out the consequences of making certain changes to our linguistic practices, e.g. changes in the rules that govern the use of our

---

[231] I will elaborate on the distinction between internal and external modal questions in section 5.5.
[232] Thomasson (2017), p. 9.
[233] See Thomasson (2015), chapter 1, especially pp. 35-43.

terms, in a metalanguage where the terms are mentioned, we can *use* the terms in counterfactuals and counterpossibles to consider those consequences. For example, instead of weighing out the consequences of changing the use 'athlete' in a metalanguage, e.g. "What would be the case if we started applying 'athlete' to Secretariat?", we can instead *use* the term in object language counterfactuals or counterpossibles, e.g. "What would be the case if Secretariat were an athlete?" What's more, the plausibility of those conditionals will reflect the consequences of adopting the considered linguistic changes and so will have bearing on plausibility of the normative view we are advocating. For example, in advocating for including Secretariat in the extension of 'athlete', one might take the conditional claim, 'if Secretariat were an athlete, then Usain Bolt would no longer be one of the greats', as a reason not to include horses in the extension of 'athlete'.

It is important to see that when we make conditional claims in metalinguistic negotiations, the distinction between indicative conditionals and counterfactuals matters to the role that conditional claim is playing. I will next show that very often indicative conditionals are what tend to matter in cases of *conceptual analysis*, where we are trying to clarify how our concepts are actually used. Alternatively, it is counterfactuals and counterpossibles that matter in cases of *conceptual development*, where we are considering the consequence of alternative conceptual practices, and *conceptual ethics*, where we are advocating for certain conceptual changes.

To begin illustrating these points, consider the following toy example involving the rules of chess.[234] In the game of chess, *castling* is a move where a player can move the king two squares towards the rook and then move the rook to the square over which the king has crossed. According to the standard rules of chess set by the governing organization *FIDE*, you are not allowed to castle after you have already moved the king or rook during the game. Now consider the following conditionals:

(20) If you are able to castle even though you've already moved the king, then you will win the game;

(21) If you were able to castle even though you've already moved the king, then you would win the game.

Imagine two novices playing a game of chess while two expert instructors watch. Imagine one of the novice players uttering the indicative (20) in a situation where they are still unsure of the actual rules of chess. Imagine that Smith, one of the expert chess players viewing the game, responds by saying, "You still might win the game, but you are not allowed to castle since you've already moved the king and that would be against the rules." This response emphasizes what the current, actual chess rules are; and, in this context, responds to the implicit uncertainty regarding the actual rules that is be communicated by (20).

On the other hand, one can imagine Jones, the other expert chess player viewing the game, responding to Smith by uttering a counterfactual like (21) (i.e., but in the third person and about the relevant player). Since Jones is an expert and fully aware of the

[234] I am intentionally using an example about rules because my own theory of counterpossibles relies on a view of modal claims that treats them as object language expressions of linguistic/conceptual rules.

actual official rules of chess, Smith's original response to the player would not be an appropriate response to Jones, nor would it be appropriate for Smith to respond to Jones by saying, "No—there is no telling what would happen because the player must not and cannot castle." We can see why that response is inappropriate by looking at what Jones might be doing when she utters (21). Consider the following three options.

First, Jones is considering, perhaps not for any particular reason, what would follow from chess games relevantly similar to the current game state if FIDE were simply to adopt a different set of rules. Another option is that she is uttering (21) to emphasize how the current rules regarding castling are an objective (via social agreement) barrier to the player winning the game, which also helps explain why the player cannot win (at least not just yet). Or, for yet another possibility, consider that Jones and Smith have a long-standing debate about the FIDE approved castling rules. Jones might utter (21) to make a point about why FIDE *should* relax the castling rules given other goals and goods that come with playing chess: perhaps because it seems that the player who would benefit from the currently illicit move should win the game for other skill related reasons and that the current rule is a severe or unneeded barrier to winning.

On any of these options, Smith and Jones are each communicating and doing something different by uttering (20) and (21) respectively. The indicative conditional is more germane to figuring out what the official rules actually are. While the counterfactual can be used either for a hypothetical consideration of alternative rules (as might be done in a project of general revision), offering an explanation of the game state, or for advocating change in the official rules (as might be done in a more normative project of revision).

I also think that the modal, 'able', that appears in the antecedent is unnecessary. By using the modal 'able', which appears in both antecedents, we are communicating explicitly in about what is allowed by the rules. Can Jones drop the modal and still make the same metalinguistic points? Consider:

(20*) If you castle even though you've already moved the king, then you will win the game;

(21*) If you were to castle even though you've already moved the king, then you would win the game.

Again, it seems that Smith's original response would be suitable if the player had uttered (20*) instead of (20): "You still might win the game, but you are not allowed to castle since you've already moved the king and that would be against the rules." However, Jones's utterance of (21*) is more complicated. First of all, counterfactuals are extremely sensitive to context. So, we must keep in mind certain features of the context of the utterance, namely, that Jones is uttering (21*) *during* the game. There is clearly a context in which someone could falsely say, "If the player had castled even though she had already moved the king, then she would have won the game." In such a context, the utterer might be discussing a previous game, in which case the current rules of chess would be held fixed. But with (21*), Jones, a chess expert, is trying to communicate something about the rules of chess while the game is in play.

Let me introduce some slightly technical terminology and sketch a picture of how we might go about evaluating (21*), which will also further explain how we might go about using counterfactuals used metalinguistic negotiations in metaphysical disputes. Call the set of possible worlds that are consistent with the standard rules of chess set by

the governing organization FIDE the *C-possible* worlds (short for 'Chess-possible'). The *C-necessary* truths are those truths that are true in all C-possible worlds. Thus, the standard rules of chess are C-necessary. In fact, we might even say that the standard rules of chess are constitutive of what it is to be C-necessary. It follows that worlds with nonstandard rules of chess are *C-impossible* worlds. Let the *expert chess players* of these worlds consist in players who know all the standard rules of chess and never forget those rules (say, not even in the heat of battle).

One way to evaluate (21*) is to find a relevantly similar C-possible world where an expert chess player is intentionally playing by her own rules and castles even though she has already moved the king. In such a C-possible world, we might be tempted to say that technically the player does not win the game because she broke the very chess rules we are holding fixed in all C-possible worlds, and you can't win by cheating.[235] In this case, (21*) is false.

Another way to evaluate (21*), however, is as a *C-counterpossible*. Find the C-impossible world most relevantly similar to the world of utterance, which is the C-impossible world with the relevant nonstandard rules where the expert chess player castles. In this case, we hold all other relevant facts about the game fixed and see if the chess player would in fact win the game. If so, (21*) is true.

Returning the conversation between Smith and Jones, to evaluate (21*) in terms of C-possible worlds would miss the potential normative points being metalinguistically

[235] One might point out that on this interpretation we can still say that the player in the antecedent world is still doing something pragmatic by intentionally subverting the rules, e.g. see Thomasson's (2017) discussion of Rosa Parks. While that may be true, I am concerned with what is being done with the entire conditional claim and not just the antecedent.

communicated by Jones.[236] One potential normative point being raised by Jones is that relevantly similar C-impossible worlds with relevantly similar games states where the player castles and wins *should* be considered more similar to the actual world. For example, this could be a way of communicating that certain other rules of chess that the player has strategically relied on during the game should not depend on rules regarding castling. Of course, this is a point that Smith need not disagree with. But suppose the player has demonstrated sufficient skill and mastery of strategy and, given the current game state, is already clearly the winner but for the castling rule. Then Jones might also be using (21*) to consider the consequences of various rule changes (as in the case in conceptual development), or to negotiate the consequences of various rules changes (as in the case of conceptual ethics), without those rule changes actually being adopted. Either way, it is best to interpret (21*) as a non-trivial C-counterpossible, which charitably accommodates the point Jones is trying to advocate.

One might try to object that the players in the C-impossible world are not even playing chess anymore. But this is false and misses the pragmatic point of the utterances. While it might be true to say that the game is changed in such a world, I don't think it means that the C-impossible game is not chess—the game being played in the C-impossible world is sufficiently and relevantly similar enough to the game being played in C-possible worlds to count as chess as opposed to, say, checkers.

[236] I am not necessarily supposing that two chess players, especially chess player not interested in analytic philosophy much less modal metaphysics, would actually go about using this language. I use it to illustrate my overarching point that counterpossibles, whether or not recognized as such by utterers, play a role in metalinguistic negotiations.

All of this discussion supports two points. The first point is that conditional claims, in particular counterfactuals, can play an important role in metalinguistic conceptual development and ethics. So, to return to the Secretariat case, by embedding a claim like 'Secretariat is an athlete' in a counterfactual, disputants can use the counterfactual to negotiate changing our conceptual, linguistic, or social practices with respect to athletes by considering what would follow from the practices being proposed without adopting those practices. The second point is that what mood the conditional is in matters not just to the semantics of the conditional but also to the pragmatic role being utilized, i.e. whether they are making clear the current consequences of the rules or are pressing for different rules by using counterfactuals to explore the consequences of adopting those different rules.

I will now consider a potential objection to this view using the Secretariat case as an example.[237] First, imagine two linguistically well-informed philosophers of athletics and hippology involved in a metalinguistic dispute about whether Secretariat is an athlete. Now suppose that on the standard actual analysis of the concept 'athlete' agreed upon by all of the relevant experts it is conceptually, or metaphysically, true that 'athletes are human', which can be expressed in the object language with the modal claim 'athletes are *necessarily* human'. As in the chess example, we can consider the following two conditionals:

[237] Let me again emphasize that this is a toy example. Pretheoretically we might not really think whether Secretariat is considered an athlete matters because it seems that the property of being an athlete doesn't really have a strict modal profile or is metaphysically interesting in any way. On the other hand, maybe there are theoreticians who take questions about the nature of athleticism seriously. If that's the case, then maybe the example seems less like a toy. Either way, I am only using the example to make a more general point about the conceptual ethics. If you like, you can change the example to one about personhood, mereological sums, material constitution, negation or maybe your favorite metaphysical dispute.

(22) If Secretariat actually is an athlete, then he is one of the greatest athletes of all time;

(23) If Secretariat were an athlete, then he would be one of the greatest athletes of all time.

One might legitimately respond to a person who utters (22)—but who, unlike the two philosophers, is not *in-the-linguistic-know*—by saying, "That's trivial—the supposition fails because we know that Secretariat *cannot* be an athlete." This response *presupposes* the standard and currently agreed upon analysis of the concept 'athlete'. However, this would not be an appropriate response to a linguistically and philosophically well-informed party who is uttering (23) as a means of *implicitly* considering what would follow if we did adopt the linguistic change being pressed for by Secretariat's advocate, namely that Secretariat be included among the athletes. Such a response would miss the point of the hypothetical supposition, which is agreed by all parties to be impossible given the standard and currently agreed upon analysis of the concept 'athlete'. Instead, in the spirit of accommodation, we should evaluate (23) by finding a relevantly similar *impossible world* where Secretariat is an athlete and see whether, given Secretariat's recorded achievements in racing (which the context of discussion holds fixed), he is also one of the greatest athletes. If in such a world Secretariat is one of the greatest athletes, (23) comes out true.

A potential objection to the story I just gave, and my overall point, is that there is a *better* interpretation of (23) on which it is clearly false. Essentially, what I am arguing is that we should not interpret (23) as a literal claim about what would be the case if impossibly the nature of athleticism were such that Secretariat is an athlete, but they are

instead engaged in metalinguistic engineering whereby they are working out the consequences of certain changes in standard linguistic practices without actually adopting those practices and by using rather than mentioning the relevant terms. One might then worry that it would be better to interpret the antecedent of (23) as describing some *possible* world where the community of language users systematically applies the term 'athlete' both to certain humans and certain horses. But, the objection continues, when considering such a world we should still hold our actual linguistic practices fixed, so even in that world Secretariat would still not be an athlete. Thus, (23) is a false counterfactual and not a non-vacuous counterpossible.

This response is similar to arguments that the following counterfactual is false:

(24) If 'tail' were to mean 'wing', then horses would have wings.

If we interpret the antecedent as merely describing a community that uses English in a non-standard way, it will not follow that horses suddenly have wings.[238] I agree, but this is a misrepresentation of what is going on the in the Secretariat example. Note again that we are supposing that the person who is advocating for Secretariat to be considered an athlete is a linguistically well-informed philosopher of athletics or hippology. As such, she presumably understands use-mention distinctions and the differences between the indicative and subjunctive moods of conditionals. Given this assumption, there are at least three reasons why an interpretation of (23) analogous to the interpretation of (24) is an uncharitable interpretation.

[238] See Kripke (1980), Yablo (2002).

First, it is more dialectically effective for the advocate to *use*, as opposed to merely mentioning, the expressions in (23). If the advocate mentions the terms instead of uses them, she risks making her point *strictly* a linguistic point by potentially cutting off other important social and cultural associations with athleticism.[239] In other words, she runs the risk of being interpreted as playing mere word games, which she is not doing since there are practical and social consequences at stake.

Second, there is an important difference between (23) and (24). In (23) the term 'athlete' is *used* in both the antecedent *and* consequent. On the other hand, in (24) the term 'wing' is first mentioned and then used. The reason why we naturally read (24) as false is that in order to read it as true we would need to shift the context of utterance, which fixes the meaning of 'wing' in the consequent, to a context where 'wing' means tail, and this is an extremely awkward shift.[240] However, by using the term 'athlete' in both the antecedent and consequent, we easily treat the 'athlete' as being used in the same way and in accordance with the proposed alteration in the rules that govern the use of 'athlete'.[241] Secratariat's advocate is intentionally using the terms in a counterpossible because, by assumption, she already understands that on the standard analysis of 'athlete' that is agreed upon by all of the relevant experts that 'athletes are human' is actually true, and so understands that relative to the assumed background of ordinary modal beliefs athletes are *necessarily* human.

Another reason why we should not interpret (23) along the same lines as (24) is that it will force us to interpret one of the speakers as making inconsistent claims when

---

[239] See Plunkett (2015).
[240] I will discuss this more in the next section.
[241] Thanks to Amie Thomasson for raising this point.

we have no good reason to do so. Suppose that we interpret (23) and counterfactuals like it along the same lines as (24). Now consider a follow-up claim made by Secretariat's advocate:

(25) If Secretariat were an athlete, then Usain Bolt would still be one of the greatest athletes.

If we interpret (25) along the same lines as (24) and consider a possible world where the linguistic community uses English in non-standard way, but we hold our actual linguistic practices fixed, then of course Usain Bolt will still be one of the greatest athletes. So, on the *possible* world interpretation, (25) is true while (23) is false. But under this interpretation we are forced to interpret Secretariat's advocate as making inconsistent claims when we have no good reason to do so. Instead, it would be better to charitably interpret her as attempting to make consistent claims, and a reading of (23) and (25) as counterpossibles allows us to do just that.

Furthermore, the strict linguistic interpretation does not accommodate the point likely trying to be made when Secretariat's advocate utters (25), which is that Secretariat's inclusion among athletes has no bearing, no effect, on the status of Usain Bolt, who will certainly remain an exemplar of human athleticism. In other words, she is metalinguistically communicating, using the object language, that there will be no unwanted downstream conceptual or social consequences if we make minor adjustments to our linguistic practices or conceptual scheme and allow Secretariat to be an athlete. Instead, when we consider (25) as a counterpossible, we consider relevantly similar impossible circumstances (relative to the assumed ordinary modal beliefs regarding athletes) where Secretariat is an athlete and we see if Usain Bolt's status as a great athlete

holds.[242] If there is no reason to seriously think that his status wouldn't hold, (25) will come out true. Therefore, assuming that (23) is true, reading (23) and (25) as counterpossibles allows us to interpret Secretariat's advocate as making consistent claims.

Now, it could be that the other party to the dispute *argues* that (25) is false because for various reasons Secretariat being an athlete would somehow diminish the significance of Bolt's accomplishments. But, even if that is the case, the other party to the dispute would be offering a compelling explanation for why Secretariat's advocate mistakenly thinks that (25) is true, which is not what is going on if we interpret (25) along the lines of (24). So, it is best not to interpret (23) and (25) using possible worlds of non-standard English users, but as conditionals where we use and not mention the terms, which means that we are dealing with counterpossibles (relative to the assumed ordinary modal beliefs regarding athletes).

5.5 Counterpossibles and Internal/External Modal Claims

In this section I will argue that counterpossibles can even do work for those conceptual engineers who take a more conventionalist or pluralist approach to metaphysical issues.[243] In the last section, I focused primarily on the work

---

[242] Again, for me this is just a way of evaluating meaning constituting inferences under a range of alternative auxiliary hypotheses but while remaining in the object language.

[243] For examples of different ways to be a conventionalist, see Einheuser (2006). By 'pluralist approach to X' I mean an approach that holds that there are many equally metaphysically correct accounts of X. While there may be many different ways of conceptualizing X or ways of X talking, which account of X we end up endorsing, or which account of X is appropriate to endorse, is largely driven by our theoretical goals, practical interests, or the role we take X to play in our cognitive lives. So, for example, consider a debate about the nature of free will such that one party argues that free will is compatible with determinism and the other party argues that free will is not compatible with determinism. One might hold a pluralist attitude about this debate insofar as one thinks that it doesn't make sense to ask which theory is more metaphysically "correct" than the other without first asking what role talk about free will plays in our cognitive and social lives, e.g. what role talk of free will plays in our practice of attributing moral

counterpossibles can do in answering normative questions about which concepts we should use. However, recall that there is also the project of *conceptual development*, which can be seen as a more experimental project of creating new concepts, as well as revising our current concepts.[244] At this point, the question of the usefulness or appropriateness of the new or revised concept will still be open. However, counterpossibles play a role in conceptual development by allowing us to think about and reason about alternative conceptual practices while remaining in the object language.

Just as there might be internal and external existence questions, there might be internal and external modal questions.[245] I take internal modal claims be modal claims made against the backdrop of our actual linguistic and conceptual practices perhaps along with empirical facts. Internal modal claims capture "ordinary" modal beliefs, e.g. 'bachelors are necessarily unmarried', 'water is necessarily H2O', 'nothing can be red all over and green all over', 'agent causation is possible', etc.[246] *Internal modal questions*, e.g. whether agent causation possible or whether water is necessarily H2O, can be settled either by analyzing our current linguistic and conceptual practices, e.g. seeing whether our ordinary concept of free will is incompatible with our views about determinism, or by empirically investigating the world, e.g. discovering the actual chemical composition of the watery stuff in our environment.[247] Like Thomasson (2015), I take internal modal questions and claims to be claims made using the relevant terms.

---

responsibility to others and whether we should continue the practice of attributing moral responsibility to others.

[244] For example, Haslanger (2012).

[245] This idea and what follows is built from work found in Einheuser (2006) and Thomasson (2015, 2017).

[246] These are just example of potential beliefs, I'm not claiming that any of them are actually ordinary modal beliefs.

[247] Cf. fn. 5.

External modal questions can be seen as questions regarding alternative modal frameworks, e.g. a modal framework where it impossible for an agent to possess free will while her actions are fully determined versus a framework where it is possible. I will next show how counterpossibles work in helping us settle external modal questions, when these external modal questions are interpreted in a deflationary manner.

An example of an external modal question might go along the lines of "Is agent causation *really* impossible?" Heavyweight realists about certain modal questions hope to answer this question by appealing to substantive facts about either modal properties, the nature of things, grounding relations, relations of fundamentality, or the metaphysical laws that structure reality.[248] For example, they might inquire into the *real definition* of free will by looking into the metaphysical grounds of free will or by looking at the fundamental structure of reality to determine whether agent causation is *really* impossible.[249] However, following the concerns already discussed 5.1, and elsewhere in this dissertation, there are good reasons to doubt that we can really locate substantive facts that might answer the heavyweight realist's reading of external modal questions. For example, there is reason to be suspicious of the epistemic reliability of methods used to answer these questions, e.g. purported faculties of rational insight, which are neither empirical nor conceptual. Instead, it might be that the intractability of debates over many first-order modal claims, e.g. 'is agent causation possible', results because such claims are attempting to evaluate modal claims independently of a linguistic or conceptual framework.

[248] Cf. Kment (2014).
[249] See Rosen (2015).

Alternatively, instead of interpreting questions about whether *p* is *really* possible (or *really* impossible) as factual questions about the modal nature of reality, one might think that external modal questions are best interpreted as practical questions regarding alternative linguistic and conceptual frameworks that each determine their own spaces of metaphysical possibility. So, it might be that the questions we really want to ask, and answer, are pragmatic questions such as "Should we adopt such-and-such rules that govern the use of '*p*'?" The frame-level rules govern the use of 'free will' or 'it is *not* the case that p', will determine whether worlds with agent causation, or worlds where ¬¬*p* is true but *p* lacks a truth value, are possible or impossible. So, on this interpretation, external questions, we are concerned with which spaces of possibility are best to adopt given other theoretical or practical considerations.

Consider the following explanation of a debate about negation in logic. Suppose that the theorems of classical logic capture our ordinary logical practices and serve as one of our ordinary default starting points when deciding what is possible and what is impossible (additional default starting points might include the physical laws, facts about essences, standard mathematics, etc.). From an *internal* perspective, e.g. the perspective of most mathematical practices, '*p* necessarily follows from ∼∼*p*' is a trivial modal truth. However, *external* to our ordinary position, i.e. from "outside" our ordinary logical practices, whether '*p* necessarily follows from ∼∼*p*' is true cannot be settled by looking at what classical logic tells us since, external to our ordinary linguistic practices, there are alternative logical practices in which it does not.

So, on a picture of logical pluralism that I have in mind, the question of whether theorems of classical logic are *really* necessary is not settled by trying to figure out the

logical nature or structure of the world (or even the logical nature or structure our cognitive lives). Instead, external to our ordinary logical practices, we settle the debate by deciding what logical systems we *ought* to adopt given our theoretical goals and needs, perhaps even only for a given context or for a given purpose. This doesn't mean that the choice between logical systems is completely arbitrary since there may be other theoretical considerations at play, e.g. standards for what counts as a good proof or standards for systemizing a collection of empirical data.

However, in the previous section I began to set up an argument for the claim that counterpossibles allow us to consider or negotiate certain consequences of other adopting alternative linguistic practices, e.g. adopting alternative logics, *internally*, i.e. within a given default position. This can be accomplished by *using* terms in counterpossible conditionals while relying on considerations of conceptual similarity and context. Since conceptual development and conceptual ethics often needs to take place in the object language and against the backdrop of our ordinary modal beliefs, counterpossibles are extremely useful for reaching outside of our actual linguistic practices to weigh reasons for or against adopting certain changes to our actual linguistic concepts while using rather than mentioning the terms.

To illustrate how this works and demonstrate its advantages, let me contrast the use of counterpossibles in considering external questions with Einheuser's (2006) work on counterconventionals. Let *mereological nihilism* be the view that only mereological primitives exist. Let *ontologese* be a language (or set of conventional practices) in which mereological nihilism is true. In ontologese, the sentences 'there are tables', 'there are mountains', 'there is a mereological sum of the mereological primitives a, b, and c', etc.

are all false. Now suppose that within our *ordinary linguistic practices* many claims such that many mereological sums do exist are true, i.e. our ordinary language is such that it is true that there are tables and mountains. Also suppose that internal to our ordinary language, it is necessarily false that, when the right constituents and relations obtain, only the mereological primitives exist. That is, against the backdrop of our ordinary conceptual practices, given the existence of certain mereological primitives along with other relevant facts (e.g. about how they are related or the intentional activities of agents), there *necessarily* exists a mereological sum of those primitives. For example, if particles arranged tablewise exist, then necessarily a table exists.[250]

Now consider the following conditional:

(26) If we had adopted ontologese, then there would have been no mountains in Africa.[251]

There is a clear sense in which this is false: our linguistic practices have no bearing on the physical geological formations in Africa. Yet, Einheuser (2006) argues that there is also a sense in which conditionals like (26) are true: we can consider (26) as a *counterconventional* where we hold non-conventional aspects of the world fixed while changing our conceptual practices. On this way of reading (26), we hold fixed certain microphysical facts and pretend that we are in the ontology room speaking only ontologese, in which case there are no mountains in Africa. What exactly is the point of considering counterconventionals? Einheuser doesn't have much to say about this, but

[250] Cf. Thomasson (2017).
[251] Einheuser (2006) uses the conditional: "Had our conventions been suitably different, then there would have been no mountains in Africa."

one reason she offers is that it gives us a clearer picture of certain conventionalists view, say conventionalism about mathematics, and helps conventionalists address certain objections to their own views.

However, Einheuser's counterconventionals explicitly mention a set of conventional practices, and this creates some difficulties in how to interpret them. Einheuser gives two different ways to read a conditional like (26). Again, on a *straight counterfactual* reading, (26) comes out false since non-linguistic facts about physical bodies in Africa do not depend on what linguistic conventions we adopt. On the counterconventional reading, (26) comes out true and captures the conventionalist intuition. Einheuser argues that the counterconventional readings are less natural than straight counterfactual readings. In normal discourse, the referents of the terms in the consequent of a subjunctive conditional are fixed by the context in which the conditional is uttered. So, in normal discourse, the referents of 'mountains' and 'Africa' in the consequent of (26) are fixed by our actual linguistic practices. Therefore, in order to make the consequent come out true, the counterconventional reading requires us to shift the context of utterance to the antecedent world, i.e. to shift from our actual conceptual practices or linguistic conventions to alternative conceptual practices or alternative linguistic conventions. In many contexts, such a shift is awkward and so creates confusion when evaluating counterconventional conditionals or when trying to understand conventionalist views.[252]

[252] Ibid., pp. 473-475.

I think that staying in the object language and using counterpossibles allows us to easily bypass all of the confusion created by counterconventional conditionals. Consider an object language analogue of (26):

(27) If mereological nihilism were true, then there would be no mountains in Africa.

Supposing that mereological nihilism is necessarily false, this is a counterpossible. However, it seems non-trivially true, since in the *right context* it is false to say that

(27*) If mereological nihilism were true, there would be mountains and no mountains in Africa.[253]

So, (27) allows us to do two things. First, by using counterpossibles in the object language, we can avoid the ambiguous reading that comes with Einheuser's counterconventional conditionals. Second, using counterpossibles in the object language allows us to implicitly evaluate various proposed changes to our actual linguistic practices without actually adopting those practices and potentially outright contradicting ourselves. Conversely, for those sympathetic to any kind of conventionalist or pluralist approach to metaphysics, this (and the work discussed in previous sections) explains what counterpossibles are doing and why we should care about them.

5.6 Why Normativism Does the Best Job

Recall from section 5.2 that one way that counterpossibles can help in conceptual ethics and conceptual development is using them to normatively advocate for or against

[253] The right context perhaps because there is context where it comes out trivially true, e.g. a context where we are feeling ontologically intolerant and our beliefs about ordinary objects are firmly held fixed and we carefully and convincingly reason by deduction from the assumption of mereological nihilism to an outright contradiction.

certain dependencies between our concepts. Recall the debate between two ethicists over DCT where one of the ethicists utters:

(19) If God were to command people to murder, then murder would still be wrong.

According to metalinguistic negotiation, we can interpret the disputants not as involved in an object level debate about the nature of moral properties and whether such properties depend on God's will, but as communicating their views about how certain concepts *should* be related. In particular, in uttering (19), one of the ethicists is metalinguistically negotiating for moral concepts that are independent of God's will, i.e. that we *should* adopt moral concepts such that worlds where God commands murder but murder is still morally wrong are more relevantly similar to the actual world than worlds where God commands murder but murder is morally permissible or obligatory.

On a substantive realist view, when deciding whether (19) is literally true we need to look at the divine and moral structure of impossible worlds and compare their similarity to the divine and moral structure of the actual world (assuming that God exists in all of the worlds being considered). But taking the counterpossibles to be made true by substantive facts about the nature of things, grounding, or fundamentality creates tension with taking counterpossibles to play the role they also need to play in conceptual ethics. Let me explain.

Conceptual engineering is driven by the idea that there is a certain amount of indeterminacy in our concept choice or even in the concepts themselves. For example, it might be that our concept of 'person' in our theories of personal identity will need to change in light of new technology or that our working concept of polyhedrons turns out

to be vague and needs to be further sharpened for fruitful mathematical inquiry.[254] Consider the question of whether persons can survive the use of a teletransportation machine (TM). Consider now the concept PERSON_A, in which someone that uses a TM survives its use, and the concept PERSON_B, in which nobody who uses a TM survives. According to the picture of conceptual ethics developed in Burgess and Plunkett (2013), in addition to the metaphysical good of truth, there are other goals and values that we can consider when debating about which concept of person we should use.[255]

Now, consider a substantive interpretation of personal identity debates according to which there is a metaphysical fact to the matter of whether I survive the use of a TM and whether the person at the other end of the TM is the same person as the person that used it. Moreover, this metaphysical fact is not one that can be discovered through conceptual analysis or straightforward empirical methods. But if this is the case, then from the point of view of the substantive realist, there is pressure on the idea that conceptual ethics is very useful and perhaps even reason to think that could be *harmful*. For if the device results in death as a matter of substantive metaphysical fact but, for whatever reason, we have conceptually negotiated that we use the concept PERSON_A, then by the lights of the substantive realist story I have told, we are making a significant mistake and living with a potentially harmful ideology of personal identity.

Overall, it seems that the substantive realist may think that either the two concepts we are negotiating are both concepts that imperfectly fit the nature of the world or that one of them does perfectly fit the nature of the world and it is the job of philosophers to

[254] Burgess and Plunkett (2013) and Tanswell (2017).
[255] Burgess and Plunkett (2013).

figure out which. If the two concepts imperfectly fit the nature of the world, then we are dealing with two conceptual attenuations of the real concept PERSON. If that's the case, then there is some tension in normatively advocating for an imperfect concept since, especially for a realist, it often seems that metaphysical facts determine important moral facts. For example, if I ought to A implies that I can A, then if it is metaphysically impossible for me to A, it is not the case that I ought to A. On the other hand, supposing that we are mistaken and one of the imperfect concepts really does capture the notion of personhood, then we again encounter tension in trying to negotiate for the imperfect concept.

A deflationist, on the other hand, may have an easier time with the issue. The metaphysical deflationist doesn't think that there is a privileged metaphysical description of the world that really gets at the real metaphysical nature of the world or "carves at the joints" of the world. If facts about personal identity are ultimately backed by our linguistic and conceptual practices, then while our linguistic and conceptual practices do in fact have very real practical outcomes, the deflationist who is in the business of conceptual ethics does not have the same kind of potential conflict as the substantive realist in the same business.

The same concern applies if you are a substantive realist about moral facts and modal facts. In that case, you think that our modal space is such that worlds where God commands murder but murder is still morally wrong *really are* or *really are not* more relevantly similar to the actual world than worlds where God commands murder but murder is morally permissible or obligatory. In that case, normatively advocating for conceptual independency between our moral and theological concepts comes with the

same risks just outlined. Specifically, the more substantive leanings you have about counterpossibles, the less work they can do for you in the project of conceptual ethics.

However, on my preferred understanding of counterpossibles, what is at stake in deciding whether (19) is acceptable is a matter of determining conceptual similarity, my account does not rely on substantive metaphysical relations of grounding or fundamentality. Furthermore, my account is not committed to a Fregean third realm of concepts. So, there is less tension in advocating that we *should* treat worlds where God commands murder but murder is still morally wrong as more relevantly similar to the actual world than worlds where God commands murder but murder is morally permissible because my account can avoid positing any substantive metaphysical or conceptual facts. Likewise, for counterpossibles (23) and (25).

Suppose that the actual application conditions of 'athlete' preclude application to non-humans. Now suppose that we look at impossible worlds where Secretariat is an athlete and that the context requires us to also accept that in such a world both Secretariat and Usain Bolt are some of the greatest athletes, but the context also requires us not to accept that in such a world Secretariat is not an athlete or that flying pigs are gold medalists. The reason, however, has nothing to do with what some subject *a priori* knows about actual athletes and it has nothing to do with any sort of robust metaphysical law regarding the nature of *athleticism*. Instead, in nearby worlds where the application conditions of 'athlete' are such as to include Secretariat, the concept will be sufficiently similar to our actual concept to thereby consider Secretariat one of the greatest athletes. Moreover, if we accept (25), then it gives reason to think that minor changes to the conceptual role of 'athlete' will not have any downstream conceptual effects that

preclude Usain Bolt from being an athlete, a practical consequence we likely want to avoid and would make people more resistant to changing their actual linguistic practices.

Conceptual ethics combined with a pluralistic attitude provides a clear and plausible story about the metaphysics and epistemology behind many philosophical debates. It also does this while avoiding trivializing those debates as being mere word games. Given the work that conceptual engineering creates for counterpossibles, especially when conceptual engineering is combined with a pluralistic attitude, it is important to also have a clear and plausible story about the metaphysics and epistemology of counterpossibles. My normativist account does just this. It provides an account of counterpossibles that does not rely on strange concrete impossible worlds nor does it rely on substantive considerations of similarity between worlds based on substantive facts about the nature of things, grounding, or fundamentatlity. According to my account, object level talk of metaphysical similarity expresses features of our concepts and their relations. So, in order to understand and use counterpossibles with distinctly metaphysical content, we only need to understand our concepts or how to use counterpossibles to explore our conceptual space relative to the conceptual universe.

5.7 Summary

I started this chapter by considering the following worry: if we want to take a more deflationary view of many debates in metaphysics, e.g. when considering whether mereological sums exist, or agent causation is constitutive of free will, then maybe we don't really need counterpossibles or to think about metaphysical similarity. One reason a deflationist might think this is that they believe that many claims in metaphysics are not

necessary in a substantive sense, and so their negations are not impossible in a substantive sense. Another reason a deflationist might think this is that questions of essence or metaphysical dependence are not really that important or that disputes about these issues are really just verbal disputes. So, counterpossibles really are not that interesting and not worth all the theoretical fuss I've been making in this dissertation.

I responded to this objection in two ways. The first was to argue that recent work in conceptual engineering provides new work for a theory of counterpossibles. The second way was to argue that even if we take a prescriptive or even conventionalist approach to conceptual engineering counterpossibles have a lot of work to do. However, this does not commit them to any substantive views about counterpossibles or metaphysical similarity between impossible worlds.

CHAPTER 6: TWO EASY WAYS TO THINK ABOUT IMPOSSIBLE WORLDS

6.1 Summary of the Project

In this dissertation I motivated the need for finding an account of counterpossibles that meets at least three *desiderata*:

1. An account of counterpossibles should be consistent with our semantic intuitions, i.e. provide for the non-vacuity counterpossibles—in particular, allow that there are some false counterpossibles;
2. An account of counterpossibles needs to enable counterpossibles to do the theoretical work we want them to;
3. An account of counterpossibles should avoid problematic ontological commitments and provide a clear story about how we come know certain counterpossible claims as well we related claims about metaphysical similarity and the metaphysical laws.

I did this by criticizing two views that suggest finding such an account is unnecessary: epistemic two-dimensionalism and recent objections raised in Williamson (2016, 2017). In responding to these objections to non-vacuous counterpossibles, I outlined interesting theoretical work counterpossibles can do in making sense of philosophical debates and the role they play in metaphysical explanations.

I then raised additional worries for two existing approaches to counterpossibles. I argued that epistemic approaches might fail to adequately account for the different dimensions of similarity that are relevant when evaluating counterpossibles. For

example, when considering what would be the case if an impossible steel Penrose triangle were to exist, we do not necessarily care about how that impossible object appears to, or is named by, different subjects, but we are interested in that impossible object. However, I argued that if we respond to the worries raised for epistemic views by adopting a criterion of metaphysical similarity that posits the existence of to substantive properties of essence or substantive relations of dependence or fundamentality—properties and relations that cannot be discovered through conceptual analysis or straightforward empirical methods—we will inherit theoretical and epistemological worries that might obscure what we are up to when we talk about counterpossibles and metaphysical similarity.

Given the remaining need for an account of counterpossibles, I offered a novel account underwritten by modal normativism. I developed a normativist reading of counterpossibles such that 'If *p* were the case, then q *would* be the case' *expresses* a *requirement* to accept that *q* given *p* and a range of *relevant auxiliary assumptions* determined by the context. However, I argued that it is perfectly acceptable to use talk of worlds to consider various ranges of auxiliary assumptions and talk of similarity to constrain what auxiliary assumptions are relevant. A more picturesque way of explaining my view is as follows. Considering both possible and impossible worlds, if we are going to *hypothetically* accept that the world is some way other than it actually is, then counterfactuals and counterpossibles convey a requirement about what kinds of worlds we ought to accept: if you're going to hypothetically accept a *p* world, it *ought* to also be a *q* world. However, another way of making the same point is to say that *p q* worlds are more relevantly similar to our world than every *p* non-*q* world, but where claims about

worlds and similarity are not primarily descriptive but, instead, are expressive claims that illustrate the actual constitutive rules that govern the use of '*p*' and '*q*' along with additional relevant conceptual or empirical information being held fixed by the context. So, my account respects other extended Lewis-Stalnaker accounts of counterfactuals and counterpossibles by keeping talk of similarity and worlds *simpliciter*, which allows for talk of both possible and impossible worlds.

Novel and interesting features of my account are:

1. In a very qualified sense, I develop a *metalinguistic theory of counterpossibles*, where the main qualification is that my account is a *prescriptive* account of counterpossibles and *not* a descriptive.
2. In the end, I develop an account of counterpossibles whereby competently evaluating counterpossibles comes down to having:
    i. a tacit understanding of, or ability to follow, the rules and permissions that govern the terms of our "home" language, perhaps along with other relevant empirical information; and
    ii. a tacit understanding of, or ability to follow, the rules governing a language that is different, yet relevantly similar to our "home" language

The main gear in my account of counterpossibles is my novel account of metaphysical laws. On my account, *the metaphysical laws just are hyperintensional object-language expressions of certain constitutive rules that govern the use of ordinary vocabulary, i.e. the conceptual laws*.

Importantly, my account meets all three *desiderata*. My account allows for non-vacuous counterpossibles. I argued that my account enables us to do the philosophical work we need counterpossibles for. And, in the end, my view avoids problematic ontological commitments and provides a clear story about how we come know certain counterpossibles since we only need conceptual analysis, perhaps along with some straightforward empirical work, to understand counterpossible claims, claims of metaphysical similarity, and claims of metaphysical laws.

In the previous two chapters, I looked at two general worries to my account. One worry was that it can't do the work I outline to motivate giving an account of counterpossibles that meets all three desiderata. In particular: my account of counterpossibles can't adequately explain why indicative conditionals and counterfactuals sometimes comes apart; my account of counterpossibles can't account for the role counterpossibles play in explaining certain asymmetries in metaphysical explanations; and my account of counterpossibles can't explain comparative impossibilities. I responded by explaining in detail how my account can do all of that work.

The second worry was that the work I outlined in chapter 1 doesn't need to be done to begin with, so we don't need to give an account counterpossibles to do non-existent work. I responded by arguing that recent work in conceptual engineering provides new work for a theory of counterpossibles. I also argued that even metaphysical conventionalists and pluralists have reason to take talk of counterpossibles seriously.

Thus, I accomplished my central aim: *I gave a novel theory of counterpossibles that can account for semantic intuitions, that can do interesting philosophical work, and that avoids the worries that come with certain epistemic readings and substantive metaphysical* readings.

6.2 Do We Need Impossible Worlds?

Throughout this dissertation I have made liberal use of talk about possible and impossible worlds, yet I have yet to say much about what I take impossible worlds to be. Of course, the metaphysics of possible worlds, let alone impossible worlds, is a rich and complicated issue, which could be a dissertation project all on its own. So, I cannot hope to adequately address this issue here. However, talking about metaphysically impossible worlds is likely to raise two eyebrows and not just one. After all, we aren't just talking about possible worlds in which Dorothy isn't in Kansas anymore, and we aren't even just talking about worlds in which Dorothy encounters talking inanimate objects and flying monkeys—we're now talking about worlds in which Dorothy encounters deviant logics, steel Penrose triangles, squared circles, talking numbers, etc. So, one may be worried that our ontology will get out of hand if we admit that there are impossible worlds, even if we are already willing admit that there are possible worlds. So, in the hopes of assuaging some of these worries, I will close this dissertation by sketching two easy ways to talk about impossible worlds.

Before getting to that, however, one might wonder if I really need to say anything about impossible worlds at all since I take talk and consideration of counterpossibles and metaphysical similarity to be expressive and not descriptive. On my account, we do not

need impossible worlds to serve as truthmakers for counterpossibles. Moreover, while we use talk of metaphysical similarity between impossible worlds in the object language, maybe we can ultimately ascend into the meta-language and eliminate all object level talk of worlds and counterparts. However, I do not think that this is a necessary move to make for two reasons.

First of all, Lewis (1986) notes that if it's possible that *p*, then there are many *ways* in which it is possible that *p*. If it is possible that there is a talking donkey, then there are many ways in which it is possible, e.g. there is a possibility in which the talking donkey is grey, a possibility in which it is blue, a possibility in which it is green, etc. Of course, Lewis is talking about possible worlds, but the same point can be made for impossible worlds: if it is impossible that *p*, then there are many ways in which it is impossible that *p*. For example, suppose that it is impossible for me to have a different biological origin than the one I in fact have. However, there seem to be many distinct impossibilities in which I have a different biological origin than the one I in fact have, e.g. an impossibility where I am the eldest biological offspring of Elizabeth II, an impossibility where I am the eldest biological offspring of David Lewis, an impossibility where I am the eldest biological offspring of the number pi, etc. In other words, it seems that we can quantify over and count impossibilities, so it seems we need to provide a story of *what* we are quantifying over and *what* we are counting. However, the second reason I think that it is unnecessary to eliminate talk of impossible worlds is that it is easier to provide an ontology for impossible worlds than one initially thinks, so there isn't really be anything to worry about to begin with.

The first easy way to think about impossible worlds is to be a fictionalist about them. This is likely an attractive view for those already committed to fictionalism about possible worlds. The remaining two ways utilizes pleonastic entities and builds on the work of Schiffer (2003), Thomasson (2007, 2015), and Steinberg (2013). Roughly, impossible worlds can be thought of either as sets of pleonastic facts or pleonastic entities in their own right. The existence of pleonastic entities is not ontologically or epistemically problematic because talk of the existence of such entities falls out of a conservative extension of a language that does not initially mention the entities. To provide a contrast which highlights some of the ontological and epistemological challenges with impossible worlds, I will briefly discuss the option of being an extended Lewisian realist about impossible worlds.

6.3 Problems with (Extended) Lewisian Realism

When giving an account of impossible worlds, one option is to be a Lewisean realist about possible worlds and claim that impossible worlds are no different from possible worlds—they are spatiotemporally and causally isolated concrete mereological sums of objects.[256] Call this view *extended Lewisian realism*. Further, simply for the sake of brevity, let's call the relevant worlds 'concrete worlds'. In addition to appeals to theoretical utility, Yagisawa (1988) justifies a version of extended Lewisian realism by noting that if quantifying over *possible* ways the world could be commits us to the existence of possible worlds, then quantifying over *impossible* ways the world could not be commits us to the existence of impossible worlds. As we will see, I think there is a sense in which we can agree, but I disagree about what the real nature of worlds is

[256] See Yagisawa (1988). However, Yagisawa no longer holds the same view, c.f. Yagisawa (2010).

(specifically, I don't think they have a substantive nature at all since, on my preferred way of understanding it, talk of impossible worlds falls out of the way we talk, i.e., they are pleonastic entities, though nothing in my overall view of counterpossibles commits me to taking this easy way).

One difficulty for extended Lewisean realism is discussed in Nolan (1997). Suppose that extended Lewisian realism is true—so that possible and impossible worlds are concrete worlds. Now suppose that God does not actually exist. Then there is a world in which God does not exist, so it is impossible for God to necessarily exist. This means that Anselm's necessarily existing God is impossible. However, according to extended Lewisian realism, there is an impossible world in the total domain of worlds where Anselm's necessarily existing God *literally* exists, i.e. a world where God exists at every possible world. But the actual world is a possible world. Hence, God both literally exists and does not literally exist in the actual world, which is a contradiction. Since a similar argument could be made on the assumption that God actually does exist, the problem is with extended Lewisian realism.

Another problem with extended Lewisian realism are the ontological and epistemological concerns that the view inherits from Lewisian realism about worlds. First, one worry is that it is hard to see how facts about spatiotemporally isolated concrete mereological sums of objects, possible or otherwise, *make a difference* to modal facts that obtain in the actual world.[257] Assume that concrete possible worlds do in fact exist.

[257] The following argument is a version of a relevance argument found in Salmon (1998) and Blackburn (1993). However, this objection focuses on the explanatory connection between modal facts and facts about impossible (as opposed to possible) worlds. A similar argument can be found against Platonism about mathematical facts in Liggins (2015).

According to the Lewisian realist all of the facts about what is or is not possible just are facts about what worlds do and do not exist. But something seems wrong with this picture. Suppose that if *p* is possible then it is necessarily possible and that if *p* is necessary, then it is necessarily necessary (i.e. assume modal logic S5). Then facts about what is and is not possible are necessary. Then, since we are assuming Lewisian realism, the possible worlds necessarily exist. But now consider the following *counterpossible*:

(28) If all of the concrete possible worlds were to suddenly go out of existence (or they had never existed at all), then we would still know all of the modal facts that we know and, since knowledge is factive, all of the modal facts would still be true.

For example, upon the destruction of all the non-actual worlds, it seems that I would still know that it is metaphysically possible for me to be a bachelor, so it would still be true that it is metaphysically possible for me to be a bachelor—else I don't know that it is metaphysically impossible for me to be a bachelor, which is absurd. So, even if we grant that concrete Lewisian possible worlds exist, their existence seems to make no difference to our actual modal knowledge, and so no difference to the actual modal facts. Furthermore, since extended Lewisian realism is a realism in Lewis's sense, the same explanatory problem arises. For example, if all of the concrete impossible worlds were to go out of existence, then we would still know what is impossible, and since knowledge is factive, all of the facts about what is impossible would be true.[258]

There are other related epistemological worries for both Lewisian and extended Lewisian realism. Warren (2017) argues that a reasonable epistemology for a subject

[258] For most, the corresponding conditional is not even a counterpossible!

should entail that our beliefs about that subject are both *sensitive* and *safe*; otherwise, we lack knowledge or our knowledge is just a matter of pure luck. Roughly: a belief that *p* is *sensitive* just in case if *p* were false, we would not believe that *p*; and a belief that *p* is *safe* just in case if we were to believe that *p*, then *p* would be true. Applying Warren's idea, a reasonable epistemology for extended Lewisian realism should entail that our beliefs about what worlds exist are both sensitive and safe. However, there are plausible reasons to think that such beliefs are neither sensitive nor safe.

Assume that concrete Lewisian worlds exist. Note that concrete Lewisian worlds are causally disconnected from us. Recall that *philosophical zombies* are supposed to be minimal physical duplicates of actual people without conscious experiences. Suppose that philosophical zombies are in fact possible because there are concrete Lewisian worlds at which philosophical zombies exist. Let's start with sensitivity. Many people believe that zombies are possible, which, given our assumptions, on Lewisian realist picture means these people truly believe that there exists a concrete world with philosophical zombies. But, similar to what I just argued above, it seems true to say that if it were the case that there are no concrete worlds with philosophical zombies, then actual people with zombie beliefs would still believe that there are zombies. However, according to concrete Lewisian realism, the zombie believers would then be in widespread error. Therefore, the beliefs of zombie believers are not sensitive to what worlds exist.

Now let's look at safety. Again, we are assuming that concrete Lewisian realism is true and that there are concrete Lewisian worlds with philosophical zombies. To say that philosophical zombies are impossible is to say that there is no possible world at which philosophical zombies exist. However, it seems true to say that if there were major

breakthroughs in cognitive science that provided strong empirical reasons in favor of believing that philosophical zombies are impossible, if David Chalmers wrote a book forcefully arguing that philosophical zombies are impossible, and if there was a miraculous philosophical consensus that philosophical zombies are impossible, then, according to concrete Lewisian realism, there would still be worlds with philosophical zombies. So, our beliefs about what is possible and impossible, i.e. our beliefs about which concrete Lewisian worlds exist, are not safe.

Therefore, following Warren's suggested connection between a reasonable epistemology and sensitivity and safety, it seems to be a miracle that we have any modal knowledge at all. But presumably we know many modal things and we know them reliably. Therefore, we have reason to be suspicious of concrete Lewisian realism. The same arguments apply to extended Lewisian realism.

6.4 Two Easy Ways to Think About Impossible Worlds

Fortunately, we do not need to be extended Lewisian realists about possible or impossible worlds. There are theories of worlds that provide a reasonable ontology and a reasonable epistemological story of our talk and consideration of impossible worlds.

The first option is to be a fictionalist about impossible worlds.[259] On this option, impossible worlds exist, but only within a fiction where we pretend there are impossible worlds. For example, on a straightforward, literal reading, the claim that there is an impossible world with steel Penrose triangles is strictly false. But, according to the fictionalist, there may be theoretical reasons to adopt a fiction according to which it is

[259] See Nolan (1997) and Kim and Maslen (2006) suggestion fictionalism about impossible worlds.

true that there are impossible worlds. Of course, I have argued in this project that having a language in which we can quantify over impossible worlds, count impossible worlds, and talk about the similarity between impossible worlds is very useful.

For example, talk of impossible worlds can help us organize many of our considered modal beliefs. Recall that according to the considered modal beliefs of many, it is impossible for me to be the eldest biological offspring of Elizabeth II, and it is impossible for me to be the biological offspring of David Lewis. However, it is sometimes useful for us to talk about these as two distinct ways the world could not be, that is, as two distinct impossibilities or as taking place in two different impossible worlds. Or according to the considered modal beliefs of many, the world could more easily have been such that Hobbes squared the circle than it could have been a world in which every proposition is true. In order to make sense of this modal judgment, it is useful to consider how much more metaphysically similar an impossible world in which Hobbes squared the circle is to the actual world than a world in which every proposition is true, but this requires talking about impossible worlds. Thus, according to a version of fictionalism, if we want to make sense of certain of our modal beliefs, we are theoretically justified in *pretending* that there are impossible worlds. So, under the pretense of an impossible worlds fiction, it is true to say, e.g., that there is an impossible world with a steel Penrose triangle, or that there is an impossible world in which Hobbes squares the circle.

The second easy way to think about impossible worlds involves the idea of pleonastic entities. The existence of pleonastic entities is not ontologically or epistemically problematic because talk of the existence of such entities falls out of a

conservative extension of a language that does not mention the entities. For example, Thomasson (2015) defends a deflationary meta-ontological view according to which we can start with an uncontroversial truth, e.g. 'the wagon is red', that does not involve a singular term, e.g. 'the property of redness', through a trivial transformation from the uncontroversial truth, derive a true sentence, e.g. 'the wagon has the property of redness', that now has a new singular term referring to properties. We can do likewise for fact talk: there is a trivial transformation from '*p*' to 'It is a fact that *p*'. For example, suppose that it is an uncontroversial truth that 'There is not an elephant in the room'. Then via a trivial transformation, we are able to introduce talk of negative facts 'It is a fact that there is not an elephant in the room.' The catch is that we cannot appeal to these properties and facts as truthmakers for our claims. But since modal normativism adopts a deflationary view of truth, we don't need to treat claims as requiring truthmakers, where truthmakers are intended as metaphysical posits required to explain what makes our claims true. Moreover, talk of properties, facts, or worlds might have other theoretical uses that justify their introduction. Of course, I have argued throughout this dissertation that talk of impossible worlds has many theoretical uses despite not needing them to serve as truthmakers for counterpossibles and claims of metaphysical similarity between impossible worlds.

According to modal normativism, modal facts and properties exist, but we derive the ability to talk about them through hypostatizations of true modal claims. Such modal facts and properties are pleonastic in the sense just described. Therefore, there is no need to reduce or ground modal facts in any non-modal facts—we get talk of modal facts easily from transformations of sentences such as 'necessarily, all bachelors are male' to

'it's a *fact* that necessarily all bachelors are male.'[260] Thomasson also notes that one could make same move regarding possible worlds. For example, we get talk of a possible world in which there is a blue donkey (as opposed to a red donkey) from transformations of sentences such as 'it is possible that there is a blue donkey' to 'there is a possible world in which there is a blue donkey.' Steinberg (2013) elaborates on these ideas further. He takes the claim that *if it is possible that p, then there is a possible world at which p* to be a conceptual truth that licenses the acceptance of a possible world in which *p* given that *p* is possible.

All of this can be used to give a pleonastic account of impossible worlds. Nolan(1997) offers something like the following principle:

> *Nolan's Comprehension Principle*: for every proposition that cannot be true, there is an impossible world where that proposition is true.[261]

From this principle and pleonasticism, we get the following argument for the existence of impossible worlds:

(1) It is not possible that *p*.

(2) If it is not possible that *p*, then there is an impossible world at which *p* (*Nolan's Comprehension Principle*).

(3) Therefore, there is an impossible world at which *p*.

Therefore, we are able to derive talk of impossible worlds from less controversial claims about what is impossible.

---

[260] Thomasson 2007
[261] Nolan (1997), p. 542.

There are two clarifications that need to be made. The first clarification is establishing that Nolan's Comprehension Principle (NCP) is a conceptual truth. The second clarification is accounting for how we individuate impossible worlds. While, I cannot fully either clarification here, I will at least make some suggestions on how to start doing so.

First, is (NCP) a conceptual truth? Both Thomasson (2007b, 2013) and Steinberg (2013) seem to hold that *if it is possible that p, then there is a possible world in which p* is a conceptual truth. In section 3.2 above, I suggested that someone who says, "*p* is possible but there is no possibility at which *p*" is guilty of a massive conceptual confusion. Does the same apply for (NCP)? It isn't hard to imagine someone saying, "*p* is impossible and there is no world at which *p*". But to do so would be to take the notion of world simpliciter as dependent on the notion of possible world. But this is a bad idea for two reasons. First, I have just spent this dissertation arguing that it is useful to talk of distinct impossibilities and talk of similarity between them, especially when evaluating counterpossibles, which also have function in discourse (though a non-descriptive function). So, we need to talk about impossible worlds which we cannot do if talk of worlds simpliciter depends on what is possible. Second, even if in the end we should abandon fictionalism about worlds in favor or some other account, it is hard to deny that talk of worlds has a close affinity to talk of stories. And, certainly, it seems absurd to say that the notion of story depends on the notion of possible stories—impossible fictions and stories are everywhere! So, instead of taking the notion of possible world to be basic, we

should take the notion of world simpliciter as basic.[262] In that case, "*p* is impossible and there is no world at which *p*" is false.

So, is someone that claims that "*p* is impossible and there is no impossible world at which *p*" guilty of a massive conceptual confusion? I grant that this is claim is not *as* conceptual confused as the claim "*p* is possible but there is no possibility at which *p*". However, while not massively confused, I think that it is confused in the following sense. If someone refuses to accept that if *p* is impossible, then there is an impossible world at which *p* but at the same time agree with most of what I have argued for in this dissertation, then anyone else who agrees with what I have argued for in this dissertation has a right to rebuke them for refusing to accept (NCP).

The second clarification is more complicated than the first. Many might be willing to agree with the premises and the conclusion of the argument for the existence of an impossible world using (NCP). However, disagreement might occur over the *number* of impossible worlds we get from the argument. For example, conservatives about what worlds there are, in particular about what impossible worlds there are, will think that there is only one impossible world—the trivial world where everything is true. However, I argued above that it is sometimes useful to talk about *distinct* impossibilities and compare them. So, I reject the claim that the only impossible world that exists is the trivial world. But then: how do we individuate the impossible worlds?

On Steinberg's (2013) account we get possible worlds from possibilities. This works in part because we are able to individuate possibilities. For example, when two six-

[262] For similar arguments, see Kment (2014), pp. 62-63.

sided dice are rolled there are two distinct ways, and hence two distinct possibilities, in which the outcome will yield 11: one where die A = 5 and die B = 6 and the other where die A = 6 and die B = 5. Steinberg suggests the following way to individuate possibilities. Let 'x ⊩ p' stand for 'according to x, q.' Then:

> the possibility that p_1 = the possibility that p_2 if and only if, for all q, the possibility that $p_1 \Vdash q$ just in case the possibility that $p_2 \Vdash q$.[263]

Furthermore, for Steinberg: for all q, the possibility that $p \Vdash q$ just in case $\Box(p \rightarrow q)$.

Like Steinberg's account of possibilities, perhaps we can individuate impossible worlds as follows:

> I1. the impossibility that p_1 = the impossibility that p_2 if and only if, for all q, the impossibility that $p_1 \Vdash q$ just in case the impossibility that $p_2 \Vdash q$.

So, for example, it is impossible for me to be the eldest biological offspring of Elizabeth II, and it is impossible for me to be the eldest biological offspring of David Lewis. But given (NCP) and pleonastic we can pleonastically derive two distinct impossibilities:

> i1: the impossibility in which I am the eldest biological offspring of Elizabeth II; and
> i2: the impossibility in which I am the biological offspring of David Lewis.

I want to claim that according to i1, I am the eldest biological offspring of Elizabeth II but not David Lewis. Likewise, for i2. Thus, i1 is not the same impossibility as i2. Now the complication comes in deciding how to explain the 'according to x, q' operator, i.e. how to explain when *the impossibility that p* $\Vdash q$ but *the impossibility that p* $\nVdash \sim q$?

---

[263] Steinberg (2013), p. 777.

We cannot rely on *any* consequence relation between sentences to explicate the 'according to x, q' operator because that will put into a dilemma. On the one hand, if we use the consequence relation of classical logic, we will be stuck with the explosion world any time we have an impossibility, since according to classical consequence relations anything follows from an impossibility. Thus, a pleonastic approach that uses a classical explication of 'according to' will fail to yield distinct impossibilities.

On the other hand, if we suppose that the consequence relation backing our 'according to' operator is classical when dealing with possible worlds (assuming classical logic is necessary) but non-classical when dealing with impossible worlds, then we end up with a disjunctive account of 'according to' and, more importantly, we end up leaving out impossibilities. Suppose that we let some non-classical consequence relation *K* determine when *q* holds according to *any* impossibility. Then there will likely be some non-trivial world that violates *K* that will either end up not really being a world or end up being conflated with the trivial world.[264] Thus, a pleonastic approach that uses a non-classical explication of 'according to' will either fail to yield all of the impossibilities or it will conflate distinct impossibilities.

One solution might be to say that: *for all q, the impossibility that p* $\Vdash$ *q just in case* $\Box(p \rightarrow q)$. The problem, of course, is that *p* is impossible, so the conditional will be vacuously true. But maybe we can take the box operator to be capable of applying across all worlds, possible or otherwise, but exactly what worlds it applies to will vary according to an accessibility relation determined by context. On this view, what worlds are

[264] See Nolan (1997). He raises this concern as a criticism of the general proposal that we need to change our notion of logical consequence in order to reason about impossibilities.

accessible according to the box operator will depend on similarity with respect to the information that is held fixed by the context. In this case, maybe we could even adopt a strict conditional analysis of counterfactuals and counterpossibles. Another way to respond to these worries is simply to take the 'according to x, p' operator as a primitive operator that is not necessarily deductively closed. Then we let context pick out the impossibilities that are relevant to our considerations.

Given these challenges, I want to briefly outline a slight variant on the pleonastic approach that might help. One might be tempted to take worlds to be sets of syntactically individuated sentences that deductively closed for possible worlds and not deductively closed for impossible worlds.[265] In this case, we might have a set of sentences that contains the sentence 'The rocket hit Hesperus' but not the sentence 'The rocket hit Phosphorus'. However, this will likely not work because it still might not be clear how to would individuate the impossible worlds—it is reasonable to think that one and the very same impossibility can be expressed in Sanskrit, English, etc. However, another option for possible and impossible worlds is to take them as sets of *pleonastic propositions*. On Thomasson's (2015) account of easy ontology, we can trivially infer the existence of propositions along the following lines:

> Undisputed claim: Snow is not blue.
> Conceptual truth: That *p* is true iff *p*.
> Derived claim: That snow is blue is false.
> Ontological claim: There is a proposition that snow is blue.[266]

[265] For example, see Brogaard and Salerno (2013), p. 655.
[266] Thomasson (2015), p. 135.

So, impossible worlds are just those non-deductively closed sets that contain pleonastic propositions such as the proposition that Hesperus is not Phosphorus, the proposition that I am the eldest biological offspring of Elizabeth II, or the proposition that I am the eldest biological offspring of David Lewis, etc.

On either the fictionalist or pleonastic approach, we end up with a reasonable epistemology. On the fictionalist account, our beliefs about what possible or impossible worlds are determined by what fictions we adopt, so those beliefs are sensitive and safe. Assume fictionalism about worlds. If philosophical zombies are impossible, then, according to the impossible worlds fiction, there are worlds with philosophical zombies, and my belief that this impossible world exists is safe. On the one hand, if I were to believe that there are impossible worlds with philosophical zombies, then according to the fiction of impossible worlds it would be true that there is an impossible world with philosophical zombies. On the other hand, if I were to believe that there is no such impossible world, then either would be no such world because I have either abandoned the impossible worlds fiction or have adopted a different fiction. My belief that impossible zombie worlds exist is sensitive. If it were not the case that there are impossible zombie worlds, then I wouldn't believe that there are such worlds because either I haven't adopted the worlds fiction or I have adopted a fiction in which there are no zombie worlds.

Now consider the pleonastic approach. I proposed that talk of metaphysically impossible worlds conceptually falls out of talk of what is metaphysically impossible. Given a plausible ontological and epistemological story of metaphysical possibility, we have a plausible ontological and epistemological story of metaphysical impossibility, i.e.

what is not metaphysically possible. Given what I have argued so far, we know what it metaphysically impossible primarily through conceptual analysis; since the existence of impossible worlds is trivially derived from truths about what is metaphysically impossible along with (NCP), I know that impossible worlds exist primarily through conceptual analysis. My beliefs in what impossible worlds exist are both sensitive and safe. If it is true that philosophical zombies are impossible, then, by a trivial transformation along with (NCP), there is an impossible world with philosophical zombies. On the one hand, if I were to I believe that there are impossible zombie worlds, then it would be true that there are impossible zombie worlds. On the other hand, if I were to believe that there are no impossible zombie worlds, then there would be no impossible zombie worlds perhaps because zombie worlds are possible rather than impossible, which might be the case because the actual linguistic framework is such that the application conditions for philosophical zombies are satisfiable (though not necessarily satisfied). My belief is sensitive because if there were no impossible zombie worlds, then I wouldn't believe that there are zombie worlds perhaps because in that scenario the actual linguistic framework is such that the application conditions for philosophical zombies are satisfiable. Thus, my beliefs in impossible worlds are sensitive and safe.

In conclusion, we do not need to be extended Lewisian realists about possible or impossible worlds. There are theories of worlds that provide a reasonable ontology and a reasonable epistemological story of our talk and consideration of impossible worlds.

Works Cited

Bernstein, S. (2016). "Omission Impossible." *Philosophical Studies*, 173, pp. 2575–2589.

Berto, F., French, R., Priest, G., Ripley, D. (2017). "Williamson on Counterpossibles." *Journal of Philosophical Logic*, pp.1-21.

Blackburn, S. (1993). "Morals and Modals." In *Essays in Quasi-Realism*. Oxford University Press.

Blackburn, S. (1999). *Think*. Oxford University Press.

Braddon-Mitchell, D. and Nola, R. (2009). *Conceptual Analysis and Philosophical Naturalism*. MIT Press.

Braddon-Mitchell, D. (2009). "Naturalistic Analysis and the A Priori." In *Conceptual Analysis and Philosophical Naturalism*, pp. 23-44. MIT Press.

Brandom, R. (2000). *Articulating Reasons: An Introduction to Inferentialism*. Harvard University Press.

Brandom, R. (2007). "Inferentialism and Some of Its Challenges." *Philosophy and Phenomenological Research*, Vol.74(3), pp.651-676.

Brandom, R. (2008). *Between Saying and Doing: Towards an Analytic Pragmatism*. Oxford University Press.

Brandom, R. (2014). "Modal Expressivism and Modal Realism: Together Again." Unpublished: Available on http://www.pitt.edu/~brandom/currentwork.html.

Brogaard, B., & Salerno, J. (2013). "Remarks on counterpossibles." *Synthese*, 190, 639-660.

Burgess A. and David Plunkett. (2013). 'Conceptual Ethics I'. *Philosophy Compass* 8, no. 12 (2013): 1091–101.

Burgess, A., and David Plunkett. (2013). 'Conceptual Ethics II'. *Philosophy Compass* 8, no. 12 (2013): 1102–10.

Cameron, R. (2010). "On the Source of Necessity." In B .Hale and A. Hoffman, eds. *Modality: Metaphysics, Logic and Epistemology*, pp. 137-152. Oxford University Press.

Chalmers, J. & F. Jackson (2001). "Conceptual Analysis and Reductive Explanation." *Philosophical Review*, Vol.110(3), p.315.

Chalmers, D. (2002). "Does Conceivability Entail Possibility?" T. Szabó Gendler and J. Hawthorne, eds. *Conceivability and Possibility*, pp. 145-200. Oxford University Press.

Chrisman, M. (2008). "Expressivism, Inferentialism, and Saving the Debate." *Philosophy and Phenomenological Research*, Vol.77(2), pp.334-358.

Einheuser, I. (2006). "Counterconventional Conditionals." *Philosophical Studies*, Vol. 127, No. 3, pp. 459-482.

Eklund, M. (2014). "Replacing Truth." Burgess, A. and Sherman, B., eds. *Metasemantics: New Essays on the Foundations of Meaning*, pp. 293–310. Oxford University Press.

Fine, K. (1994a). "Essence and Modality." *Philosophical Perspectives*, 8, 1–16.

Fine, K. (1994b). "Senses of Essence." Walter Sinnott-Armstrong, Diana Raffman & Nicholas Asher, eds. *Modality, Morality and Belief: Essays in Honor of Ruth Barcan Marcus*, pp. 53-73. Cambridge University Press.

Fine, K. (2012). "Guide to Ground." Fabrice Correia & Benjamin Schneider, eds. *Metaphysical Grounding: Understanding the Structure of Reality*, pp. 37-80. Cambridge University Press.

Haslanger, S. (2012). *Resisting Reality: Social Construction and Social Change*. Oxford University Press.

Horwich, P. (1990). *Truth*. Oxford University Press.

Jago, M. (2014). *The Impossible: An Essay on Hyperintensionality*. Oxford University Press.

Jenkins, C. S., & Nolan, D. (2012). "Disposition Impossible." *Noûs*, 46(4), 732–753.

Kim, S. & Maslen C. (2006). "Counterfactuals as Short Stories." *Philosophical Studies*, May, Vol.129(1), pp. 81-117.

Kment, B. (2006). "Counterfactuals and the Analysis of Necessity." *Philosophical Perspectives*, 20, pp. 237-302.

Kment, B. (2014). *Modality and Explanatory Reasoning*. Oxford University Press.

Krakauer, B. (2012). *Counterpossibles*. Dissertation.

Kripke, S. (1980). *Naming and Necessity*. Oxford University Press.

Lange, M. (2009). *Laws & Lawmakers*. Oxford University Press.

Lewis, D.K. (1973). *Counterfactuals*. Blackwell.

Lewis, D.K. (1983) "Scorekeeping in a Language Game." In *Philosophical Papers: Volume I*. Oxford University Press.

Lewis, D.K. (1986). *On the Plurality of Worlds*. Oxford University Press.

Lewis, D.K. (1986b). "Counterfactual Dependence and Time's Arrow." In *Philosophical Papers: Vol. II*. Oxford University Press.

Liggins, D. (2014). "Grounding and the Indispensability Argument." In D. Molinini, F. Pataut, & A. Sereni (Eds.), *Synthese*, special issue Indispensability and Explanation. doi: 10.1007/s11229-014-0478-2.

Lycan, W. G. (2005). *Real Conditionals*. Oxford University Press.

Menzies, P. and Price, H. "Is Semantics in the Plan?" Braddon-Mitchell, D. and Robert N., eds. *Conceptual Analysis and Philosophical Naturalism*, pp. 183-200. MIT Press.

Nolan, D. (1997). "Impossible Worlds: A Modest Approach." *Notre Dame Journal of Formal Logic*, 38, pp. 535-73.

Nolan, D. (2009). "Platitudes and Metaphysics." Braddon-Mitchell, D. and Robert N., eds. *Conceptual Analysis and Philosophical Naturalism*, pp. 267-300. MIT Press.

Nolan, D. (2011). "The Extent of Metaphysical Necessity." *Philosophical Perspectives*, V. 25, Issue 1, pp. 313–339.

Nolan, D. (2014). "Hyperintensional Metaphysics." *Philosophical Studies*, Vol.171(1), pp.149-160.

Nolan, D. (forthcoming). "Causal Counterfactuals and Impossible Worlds." Forthcoming in Beebee, H., Hitchcock, C. and Price, H., eds. *Making a Difference*. Oxford University Press

Plunkett, D. and Tim Sundell (2013). 'Disagreement and the Semantics of Normative and Evaluative Terms'. *Philosopher's Imprint* 13 (23): 1-37.

Plunkett, D. (2015). "Which Concepts Should We Use?: Metalinguistic Negotiations and The Methodology of Philosophy", *Inquiry*. Vol. 58, Nos. 7-8, 828-874.

Price, H. (2011). *Naturalism Without Mirrors*. Oxford University Press.

Rayo, A. (2015). "Essence Without Fundamentality*", Theoria, 30,* pp. 349-363.

Ripley, D. (2012). "Structures and Circumstances: Two Ways to Fine-Grain Propositions." *Synthese*, Vol.189(1), pp.97-118.

Rosen, G. (2006). "The Limits of Contingency." McBride, F., eds. *Identity and Modality*, pp. 13–39. Oxford University Press.

Rosen, G (2015). "Real Definition." *Analytic Philosophy*, Vol. 56 No. 3, pp. 189–209.

Ryle, G. (2000). *The Concept of Mind*. University of Chicago Press.

Salmon, N., 1988. 'Review of *On the Plurality of Worlds*, *Philosophical Review*, 97(2): 237–244.

Schaffer, J. (2010). "Monism: The Priority of the Whole." Philosophical Review 119.1: 31–76.

Schaffer, J. (2017) "Laws for Metaphysical Explanation." *Philosophical Issues*, October 2017, Vol.27(1), pp.302-321.

Schiffer, S. (2003). *The Things We Mean*. Clarendon Press.

Schroeter, L. (2017). “Two-Dimensional Semantics.” Stanford Encyclopedia of Philosophy: https://plato.stanford.edu/archives/sum2017/entries/two-dimensional-semantics/.

Sellars, Wilfrid. (1958). “Counterfactuals, Dispositions and the Causal Modalities”. Herbert Feigl, Michael Scriven, and Grover Maxwell, eds. *Minnesota Studies in Philosophy of Science Volume 2: Concepts, Theories and the Mind-Body Problem* pp. 225-308. University of Minnesota Press.

Sider, T. (2011). *Writing the Book of the World*. Oxford University Press.

Sidelle, A. (1989). *Necessity, Essence, and Individuation: A Defense of Conventionalism*. Cornell University Press.

Simion, M. (2017). “The ‘Should’ in Conceptual Engineering.” *Inquiry*, DOI: 10.1080/0020174X.2017.1392894.

Stalnaker, R. (1968). “A Theory of Conditionals.” *American Philosophical Quarterly*, 98-112.

Steinberg, A. (2013). “Pleonastic Possible Worlds.” *Philosophical Studies*, 164, pp. 767-789.

Steward, S. (2015). “Ya Shouldn’ta Couldn’ta Wouldn’ta.” *Synthese*, Vol.192(6), pp.1909-1921.

Tanswell, F. S. (2017). “Conceptual Engineering for Mathematical Concepts.” *Inquiry*, DOI: 10.1080/0020174X.2017.1385526.

Tennant, Neil. (1990). *Natural Logic*. Edinburgh University Press.

Thomasson, A. (2007a). *Ordinary Objects*. Oxford University Press.

Thomasson, A. (2007b). "Modal Normativism and the Methods of Metaphysics." *Philosophical Topics*, 35/1&2, 2007 (released 2010), pp. 135-160.

Thomasson, A. (2009). "Non-Descriptivism about Modality: A Brief History and Revival." *Baltic International Yearbook of Cognition, Logic and Communication* Volume 4: 200 Years of Analytical Philosophy, pp. 1-26.

Thomasson, A. (2013). “2012 Nancy D. Simco Lecture: Norms and Necessity”. *Southern Journal of Philosophy*, 51 (2), pp. 143-160.

Thomasson, A. (2015). *Ontology Made Easy*. Oxford University Press.

Thomasson, A. (2017). “Metaphysical Disputes and Metalinguistic Negotiations.” *Analytic Philosophy*, Vol. 58 No. 1, pp. 1–28.

Thomasson, A. (forthcoming). *Norms and Necessity*. Book manuscript.

Vander Laan, D. (2007). “Counterpossibles and Similarity.” F. Jackson & G. Priest, eds., *Lewisian Themes: The Philosophy of David K. Lewis*, pp. 258–276. Oxford University Press.

Vetter, B. (2016). “Counterpossibles (Not Only) for Dispositionalists.” *Philosophical Studies*, 173, 2681–2700

Warren, J. (2014). “Conventionalism, Consistency, and Consistency Sentences.” *Synthese*, 192(5), 1351–1371.

Warren, J. (2015a). “The Possibility of Truth by Convention.” *Philosophical Quarterly*, Vol. 65 Issue 258, pp. 84-93.

Warren, J. (2015b). “Talking with Tonkers.” *Philosophers Imprint*, 2015 Sep, Vol.15(24), pp.1-24.

Warren, J. (2016). “Sider and the Epistemology of Structure.” *Philosophical Studies*. Vol.173(9), pp. 2417-2435.

Warren, J. (2017). “Epistemology versus non-causal realism.” *Synthese*, 194, 1643–1662.

Williamson, T. (2007). *The Philosophy of Philosophy*. Blackwell.

Williamson, T. (2016). “Counterpossibles.” *Topoi*, pp. 1-12.

Williamson, T. (2017). *Counterpossibles*. Forthcoming: http://www.philosophy.ox.ac.uk/__data/assets/pdf_file/0010/37756/Counterpossibles.pdf.

Wilson, A. (2016). “Grounding Entails Counterpossible Non-triviality.” *Philosophy and Phenomenological Research*: doi: 10.1111/phpr.12305.

Yablo, S. (1996). “How in the World?” *Philosophical Topics*, Vol.24(1), pp. 255-286.

Yablo, S. (2002). “Coulda, Woulda, Shoulda.” T. Szabó Gendler and J. Hawthorne, eds., *Conceivability and Possibility*, pp. 441-492. Oxford University Press.

Yagisawa, T. (1988). “Beyond Possible Worlds”, *Philosophical Studies*, 53: 175–204.

Yagisawa, T. (2010). *Worlds and Individuals, Possible and Otherwise*. Oxford University Press.