

The Problem of Counterpossibles

MSc Thesis (*Afstudeerscriptie*)

written by

Jorge Viterbo Ferreira

(born June 24th, 1994 in Porto, Portugal)

under the supervision of **Prof. Dr. Franz Berto** and **Dr. Luca Incurvati**,
and submitted to the Board of Examiners in partial fulfillment of the
requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam*.

Date of the public defense: **Members of the Thesis Committee:**
June 29, 2018

Dr. Maria Aloni
Prof. Dr. Franz Berto
Dr. Luca Incurvati
Dr. Martin Lipman
Dr. Katrin Schulz (*chair*)
Dr. Shane Steinert-Threlkeld



INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

The standard semantics for counterfactuals, due to Lewis (1973) and Stalnaker (1968), defends that a counterfactual $p \Rightarrow q$ is true if, and only if, q is true at all the closest possible worlds where p is true. This semantics predicts that all counterfactuals with metaphysically impossible antecedents are vacuously true. This goes against our linguistic intuition - some counterfactuals with metaphysically impossible antecedents seem false. This is a problem for the standard semantics - the *problem of counterpossibles*.

Those that think we should accept that all counterpossibles are vacuously true are called *vacuists* and their opponents are called *non-vacuists*. Some non-vaculist semantics have been proposed, most notably by Nolan (1997) and Berto, French, Priest and Ripley (2017). These semantics keep the general structure of the standard semantics intact but add *impossible worlds* - points of evaluation where impossible propositions can be true. The vacuists, most notably Williamson (2017), claim that this cure is worse than the disease.

In this thesis I will try to do three things. 1- Clarify the terms of the debate between vacuists and non-vacuists; 2- Point out where I think the existing literature is mistaken; 3- Develop a moderate form of non-vacuism (also defended by Kment 2014) which will help highlight what the minimal theoretical commitments of non-vacuism are.

The structure of the thesis is as follows. Chapter 1 introduces the problem of counterpossibles in detail, sketches the general vacuist strategy and discusses the methodological assumptions implicit in the debate. The vacuist accuses the non-vaculist of being committed to a number of theoretical costs that offset the benefits of fully respecting our linguistic intuition. The rest of the chapters each deal with one of these costs. Chapter 2 deals with the use of impossible worlds in semantic theorizing. Chapter 3 discusses the accusation that the non-vaculist must have an objectionably weak counterfactual logic. Chapter 4 is concerned with the effects of accepting non-vacuism for our epistemology and understanding of metaphysical modality. Chapter 5 deals with the question of whether non-vacuism is committed to the failure of the substitutivity of identicals in the scope of counterfactuals and whether this is a problem. Chapter 6 concludes.

Acknowledgements

I have been fortunate to be in contact with a lot of amazing people during my two years in Amsterdam. I hope you indulge me in some lengthy acknowledgements.

My greatest debt of gratitude is to my supervisors, Franz Berto and Luca Incurvati. Franz is the best supervisor one can hope for. He helped me find my interests from the very first semester I was in Amsterdam. Without his patience and constant guidance, I would not have developed whatever philosophical personality I have now. But, more importantly, he was a great mentor. He provided words of encouragement at times where I desperately needed them. And the fact that he had some confidence in my philosophical abilities often allowed me to have some as well. Luca was an inspiring supervisor and professor. He was always a model of the philosophical and technical precision that I lack. I have no doubt that my work is much better for my contact with him.

Another person who made my academic experience in Amsterdam much better was Martin Lipman. Apart from having taught the best course I took during my time here, he went above and beyond his obligations to make me understand the realities of a life in professional philosophy better than I ever could on my own. If in the future I manage to find my own small place there, I owe it to him. For further academic guidance I must also thank the participants in the Amsterdam Metaphysics Seminar and the Logic of Conceivability Project. I want to single out Peter Hawke and Tom Schoonen in particular, who taught me so much.

Thanks to Felicia, Veronica and Silvia for making Amsterdam a livable place. I owe my Father, Mother and Grandmother more than words can express for allowing me the opportunity to study in the Master of Logic. I can only hope to be deserving of their confidence.

Thanks to my Mother, in particular, for having taught me the basic skills necessary for being a philosopher from early on in my life. If you'll allow me the only sentences my grandmother will understand in this thesis: Olha avó, agora estou a falar contigo! Muito obrigado por tudo. Amo-te mais que tudo e tenho muitas saudades tuas!

Finally, as with everything I do, I would have done none of it if not for Catarina.

Contents

1	Introduction and Methodological Considerations	3
1.1	What is a Counterfactual?	3
1.2	Lewis-Stalnaker Analysis	5
1.3	The Problem of Counterpossibles	9
1.3.1	False Counterpossibles	9
1.3.2	Metaphysical Modality	10
1.3.3	Vacuism and Non-Vacuism	12
1.4	Vacuism	12
1.4.1	Connection with Compositional Intensional Semantics . .	12
1.4.2	Hybrid Theories	14
1.4.3	Lewis' Motivation for Vacuism	14
1.4.4	Pragmatics	16
1.5	Methodological Considerations	19
1.5.1	Semantic Deontologism	20
1.5.2	Analogy with Universal Quantification	20
1.5.3	Sophisticated Semantic Deontologism	22
1.6	The Structure of the Thesis	23
2	Impossible Worlds	25
2.1	The Ontology of Impossible Worlds	25
2.1.1	Using Impossible Worlds	25
2.1.2	Adaptation of Ontological Stories	28
2.1.3	An Ontology: Linguistic Ersatzism	30
2.2	Using Impossible Worlds in Semantics	33
2.2.1	Two Objections: Compositionality and Synonymy	33
2.2.2	A Pragmatic View of Impossible Worlds	35
2.3	A Restriction for the Semantics of Counterfactuals	37
3	Counterfactual Logic	40
3.1	Counterfactual Logic	40
3.1.1	What is a Counterfactual Logic?	40
3.1.2	Standard Counterfactual Logic	41
3.1.3	A Non-Vaculist Semantics	44
3.2	Weak Logic	45

3.2.1	The Weak Logic Argument	45
3.2.2	SIC-Based Strategy	47
3.2.3	Against SIC	49
3.3	A Possible Solution	51
3.3.1	Using Only Logically Possible Worlds	51
3.3.2	Counterlogical Vacuism	53
3.3.3	Defensive Moves	54
3.3.4	Upshot	57
4	Counterfactuals and Metaphysical Modality	60
4.1	Williamson's Logical Equivalence	60
4.1.1	Counterfactual Modal Epistemology	61
4.1.2	A Conception of Metaphysical Modality	63
4.2	Non-Vacuism and the Logical Equivalence	65
4.2.1	Berto et al.'s Solution	65
4.3	A Non-Vacuist Counterfactual Modal Epistemology	66
4.3.1	Skeptical Scenarios	68
4.3.2	Constitutive Facts	70
4.4	The Conception Problem	72
5	Counterfactuals and Worldly Opaqueness	74
5.1	Preliminaries	74
5.2	Worldly Opaqueness	76
5.3	Defensive Moves	78
5.3.1	Peer-based Intelligibility	79
5.3.2	Impossible Worlds Model	81
5.4	Counterfactuals are Transparent	85
6	Conclusion	87

Chapter 1

Introduction and Methodological Considerations

In this chapter, I start by introducing the problem that will be the focus of this thesis - *the problem of counterpossibles*. It is a problem for a particular semantics for counterfactuals, the one defended by Lewis (1973) and Stalnaker (1968). I will explain what kind of sentences counterfactuals are in section 1. The Lewis-Stalnaker semantics is introduced in section 2. Section 3 presents the problem of counterpossibles - the semantics makes all counterfactuals with impossible antecedents vacuously true. Section 4 sketched the moves a defender of the Lewis-Stalnaker semantics can make in response to the problem. Section 5 discusses the methodology used in the debate surrounding the problem of counterpossibles. Section 6 outlines the structure of the rest of the thesis.

1.1 What is a Counterfactual?

In this thesis, we will be occupied with a debate regarding the semantics of sentences like

(1) If I had taken her Knight, she would have a crushing attack
against my King

(1) can be uttered in the context of a chess game, when analysing a given position where I did not take my opponent's Knight but could have. The sentence is exploring what would be the case if things had gone differently and I had taken the Knight. In particular, (1) is asserting that my opponent would have a crushing attack against my King in that situation. The general form of such sentences can be represented as follows:

(2) If it had been the case that A , then it would have been the case that B

Where ' A ' and ' B ' are variables ranging over declarative sentences.¹ In philosophical parlance such sentences have been called *counterfactual* or *subjunctive* conditionals. The label “counterfactual” tries to define the class of conditionals we are referring to *semantically*: the antecedent is false (or assumed to be false, or known to be false). The label “subjunctive” tries to define the class based on the *form* of the conditionals involved: that the subjunctive mood is used. Moreover, “subjunctive” marks an opposition with another philosophically significant class of conditionals, the indicative conditionals. To see the difference, consider the following pair:

- (3) If Oswald hadn't shot Kennedy, someone else would have
(4) If Oswald did not shoot Kennedy, someone else did

(3) is a subjunctive conditional while (4) is an indicative conditional. It is clear that they can differ in truth value: in the actual world, (3) seems false and (4) seems true. For (3) to be true, there would have to be a second shooter on site at the time of Kennedy's murder, something for which we have no empirical evidence. But (4) takes it for granted that Kennedy was, in fact, murdered - we know this much about the actual world. If there was a murder, there was someone doing the murdering - and that is what (4) asserts. Providing a semantics for either type of conditionals is a hard thing to do. But in this thesis we will not be occupied with indicative conditionals.

As it turns out, both labels, “counterfactual” and “subjunctive”, have problems. Both of them fail to include some conditionals we want to be talking about. The antecedent of a counterfactual does not always need to be false:

- (5) If Jones had taken arsenic, he would have shown just exactly those symptoms which he does in fact show²

(5) can be uttered by a doctor trying to justify his diagnosis of a patient, wanting to provide evidence for the *truth* of the antecedent of (5).

The “subjunctive” label also faces problems. Iatridou (2000) argued that the subjunctive mood is not necessary for a conditional to be a counterfactual. In fact, some languages who do have the subjunctive mood do not use it when they want to produce sentences such as (1) and (2).

Keeping all this in mind, in this thesis I will use the label “counterfactuals”, as is entrenched in the philosophical literature. If you want, a counterfactual is a type of conditional that *typically* has a false antecedent, though not always. We will not need a completely precise characterisation to proceed with our discussion.³

¹The slightly unnatural constructions involving “been the case that” are there to ensure grammaticality when the variables are instantiated with actual English sentences.

²This example was first presented by Anderson (1951).

³Maybe using Iatridou's notion of “fake past” provides a better platform for a correct formal characterisation of the conditionals I want to talk about. Getting into these issues is outside the scope of this thesis, but see Schulz (2014).

Let me take a moment to fix some terminology. Recall the schematic sentence (2):

(2) If it had been the case that A , then it would have been the case that B

Given a particular counterfactual, I will call the declarative sentence that instantiates A the *antecedent-clause* and the one that instantiates B the *consequent-clause*. This allows me to keep the more natural *antecedent* and *consequent* for the *propositions* expressed by these sentences. I will be assuming that declarative sentences express propositions and that these are the primary bearers of truth, sentences being true derivatively, in virtue of the truth of the proposition they express. I will use the symbol $\lceil \Rightarrow \rceil$ to represent the counterfactual conditional. I could represent (2) by $\lceil \alpha \Rightarrow \beta \rceil$, where α is the proposition expressed by A and β is the proposition expressed by B .⁴

1.2 Lewis-Stalnaker Analysis

We now turn our attention to the *semantics* of counterfactuals: under what conditions is such a sentence true or false?

As we saw, a distinctive feature of counterfactuals is that they typically have false antecedents. The most familiar kind of conditional, at least for those who were victims of some philosophical education, is the truth-functional *material* conditional, $\lceil \rightarrow \rceil$. It seems intuitively clear that the counterfactual conditional is not the material conditional. A material conditional is always true when its antecedent is false. But not all counterfactuals are true. Therefore, we need a different analysis.

A natural thought is to use what is called a *strict* conditional. A strict conditional is just a necessitated material conditional. We can write it as $\Box(p \rightarrow q)$, where the box is to be read as metaphysical necessity.⁵ It is also clear that counterfactuals cannot simply be identified with strict conditionals. Consider the following:

(6) If I had struck the match, it would have lit

In a great many contexts of utterance, hopefully, (6) will be true. But the corresponding strict conditional is not true. It is metaphysically possible for me to have struck the match without it having lit. This means that it is possible for the material conditional corresponding to (6) to have been false and this means that the strict conditional is false as well.

But there is still hope for the strict conditional analysis. Counterfactuals are context-dependent. Just consider the following:

(7) If Caesar had been in command during the Korean war, he would have used atomic bombs

⁴I will sometimes use \Rightarrow to abbreviate the natural language counterfactual construction as well, when no confusion threatens.

⁵If you do not know what that is, do not worry - it will be explained below.

(8) If Caesar had been in command during the Korean war, he would have used catapults

Both these counterfactuals can be true, albeit in different contexts. (7) can be uttered in a context where the relevant counterfactual situations are those in which Caesar's temperament stays fixed while the weaponry at his disposal is that of a US commander in chief. And (8) can be uttered in a context where the weaponry at Caesar's disposal during his lifetime is held fixed.

Context-dependency opens the door for the possibility of context providing some parameter which can be used in a semantics of counterfactuals. Someone wishing to use the strict conditional to provide a semantics for the counterfactual conditional can exploit this as follows. There is an *accessibility function* f , from contexts to sets of worlds, that selects the accessible worlds at a given context. What makes (6) true is not that the corresponding material conditional is true at every possible world. It is sufficient that the material conditional be true at every *accessible* possible world. The result is the following semantic analysis of counterfactuals:

Strict Conditional Analysis: $p \Rightarrow q$ is true at a context c if and only if $p \rightarrow q$ is true at all accessible worlds selected by $f(c)$

What worlds are accessible in a given context is a hard question to answer: it corresponds to what Goodman (1955) called the *problem of co-tenability*. But the semantics of counterfactuals can go on without a perfectly precise characterisation of how this might go.

The standard semantic analysis of counterfactuals nowadays starts from a criticism of the strict conditional analysis. This analysis was developed by Stalnaker (1968) and Lewis (1973), so I will call it the *Lewis-Stalnaker analysis*.⁶

The strict conditional analysis predicts that certain patterns of inference which seem intuitively invalid are, in fact, valid. Consider the following pair:

(9) If I had struck the match, it would have lit

(10) If I had struck the match and it had been wet, it would have lit

Assume that (9) and (10) are uttered in the same context. This means that they are both strict conditionals over a certain set of worlds $f(c)$. In this case, the inference from (9) to (10) is valid. If a conditional $p \rightarrow q$ is true at every member of a set of worlds, then the conditional $(p \wedge r) \rightarrow q$ is also true at every member of that set of worlds. If it were not, there must have been a world where $(p \wedge r)$ is true while q is false. But this would also be a world where p was true and q was false, thus contradicting the assumption that $p \rightarrow q$ is true at every world.

The problem is simple. It seems that (9) can be true and (10) false. The ability of the match we are considering to light is conditional on the fact that it is not wet. Therefore, the strict conditional analysis makes a wrong prediction.

⁶Stalnaker's semantics was different from Lewis', as I will point out in due course. But their general strategy was essentially the same for present purposes.

There seems to be a way for the strict conditional theorist to respond. Maybe there was a context shift between (9) and (10). This would mean that the accessible worlds in each counterfactual evaluation were not the same and the inference turns out invalid. Lewis thinks that this is not the way to go:

It is defeatist...consign[ing] to the wastebasket of contextually resolved vagueness something much more amenable to systematic analysis than most of the rest of the mess in that wastebasket (Lewis 1973: 13)

Every argument form can be invalidated by positing a context shift. But the context shift must be motivated and Lewis thinks that, in the case of (9) and (10), it cannot be.⁷ What, then, is the “systematic analysis” that Lewis proposes instead?

Lewis thinks that the context provides a *closeness ordering* instead. There is a function f from contexts to total orderings of the space of possible worlds relative to a base world. And this can be used in the semantic analysis of counterfactuals. A counterfactual will be true if the consequent is true at all the *closest worlds* where the antecedent is true. The result is the following semantic analysis:

Lewis-Stalnaker Analysis: $p \Rightarrow q$ is true at a context c and base world w_b if and only if q is true at all the possible worlds w such that (i) p is true at w ; (ii) If p is true at w' , then $w \preceq_{w_b} w'$, where \preceq_{w_b} is the value of $f(c)$.⁸⁹

We read $w \preceq_{w_b} w'$ as “the world w' is at least as close as the world w to the world w_b . Here is how the inference from (9) to (10) comes out invalid in this semantics. (9) and (10) are uttered in the same context, there is no need for “waste basket-consigning”. But what the context provides is a closeness ordering. In the closeness ordering where (9) is true but (10) is false the following happens. The closest worlds where I strike the match do not include any world where the match is also wet. This makes (9) come out true. But when we move farther along the closeness ordering, we find worlds where the antecedent of (10) is true. And, in all of the closest of these, the match does not light when I strike it. This makes (10) false.

⁷For semantic treatments in the tradition of dynamic semantics that disagree with Lewis on this, see von Stechow (2001) and Gillies (2007).

⁸This formulation assumes what Lewis calls the Limit Assumption. The Limit assumption tells us that when there is at least one world where the antecedent is true, there must be at least one *closest* world where the antecedent is true. An infinite chain of ever closer possible worlds where the antecedent is true cannot happen. This will not make a difference in what follows.

⁹Stalnaker’s semantics is different from Lewis’. Stalnaker assumes that the contextual parameter is a *selection function*, that selects the unique closest antecedent world at a given context. You can reproduce Stalnaker’s semantics in Lewis’ framework by making the Uniqueness Assumption, that when there is at least one world where the antecedent is true there is *exactly one* closest world where the antecedent is true. The differences between the two semantics, such as the validity of conditional excluded middle $\models (p \Rightarrow q) \vee (p \Rightarrow \neg q)$, will not matter in this thesis.

This is one way into the Lewis-Stalnaker analysis. Some more points of detail are in need of clarification.

We will call the world relative to which the closeness ordering is built a “base world” and not the “actual world”. This is due to the fact that embedded counterfactuals, of the form $p \Rightarrow (q \Rightarrow r)$ can shift the base world when p is actually false. The embedded counterfactual $(q \Rightarrow r)$ depends on the truth value of r at the closest q -worlds *relative to* the closest p -worlds to the world of utterance. In most cases this will not matter much, and we might refer loosely to the base world as “actual world” or “world of evaluation”.

Lewis has an intuitive interpretation of the closeness ordering in mind. He wants to interpret the closeness ordering using the notion of *similarity*, a notion we must have some grip on. In other words, he thinks the following is true:

Closeness-to-Similarity Principle: A world w is at least as close as a world w' to the base world if and only if w is at least as similar (in the relevant aspects) as w' to the base world

Similarity is understood to be a highly context dependent matter: what aspects of similarity we value might change from context to context. Which thing is more similar to a red circle, a red square or a blue circle? The answer is “It depends”.

What aspects, in particular instance, influence the similarity metric is a hard question to answer. As with the problem of co-tenability, the lack of precision here does not prevent further discussion.¹⁰

To see the Lewis-Stalnaker analysis in practice, recall the example that opened this chapter:

- (1) If I had taken her Knight, she would have a crushing attack against my King

How would the evaluation of this counterfactual go? We start from the actual world where I did not take my opponent’s Knight. We go back to the position where I might have done so and consider the worlds where I did take the Knight in that position. In the closest of these worlds the rules of chess are the same, the skill level of my opponent stays fixed and the accepted meaning of “crushing attack” remains unchanged. Worlds where this is the case are closer to the actual world than world where these things change for no reason. We now ask of each one of these closest worlds: does my opponent have a crushing attack against my King? If she does have a crushing attack at *all* of them, then (1) is true. If she fails to have a crushing attack at one of them, if there is a defensive variation that my actual chess skills could find, then (1) is false.

Most people find this story intuitively compelling. The Lewis-Stalnaker analysis can be considered one of the success stories of the use of possible worlds in

¹⁰For an exchange that advanced the understanding of the notion of similarity immensely, see Fine’s (1975) review of Lewis (1973) and Lewis (1979) reply.

semantics.¹¹ But it is not without its problems. This thesis will be occupied with one of them, which I will present in the next section.¹²

1.3 The Problem of Counterpossibles

1.3.1 False Counterpossibles

Consider the following counterfactual:

(11) If Bill Gates was my father, I would be poor

(11) seems false. Bill Gates is a very rich man. We have no reason to think that he would not give enough money to an hypothetical extra son to, at least, stay above the poverty line. One problem with the Lewis-Stalnaker analysis is that, given certain metaphysical assumptions, it predicts that (11) is true.

The Lewis-Stalnaker analysis uses possible worlds. Propositions are true at (or true according to) possible worlds. Because the worlds are possible, no impossible proposition can be true at them. But the antecedent-clause of (11) expresses an impossible proposition.

The modality in question is *metaphysical modality*. “Metaphysical modality” is something of a technical term in philosophy, given its most influential characterization in Kripke (1980). There he argued that each object has its origins necessarily. The table where I am writing could not have been made from a different block of wood. And I could not have developed from a sperm cell and egg different from the ones I actually developed from. This implies that Bill Gates could not have been my father. The proposition that Bill Gates is my father is metaphysically impossible. And it seems clear that the antecedent-clause of (11) expresses this proposition.

Whether a counterfactual is true or not, in the Lewis-Stalnaker analysis, depends on an universal quantification at the level of the metalanguage. The counterfactual is true if the consequent is true at *all the closest possible worlds* where the antecedent is true. If the antecedent is impossible, it is true at no possible worlds. *A fortiori*, this means that there are no closest possible worlds where the antecedent is true. Since universal quantifications are true when the domain they quantify over is empty, a counterfactual with an impossible antecedent is always true, irrespective of its consequent.

Call a counterfactual with a metaphysically impossible antecedent a *counterpossible*. Since (11) is in such a condition, assuming the necessity of origins, (11) is true. This goes against our intuitions. A semantic analysis that does not respect ordinary speakers’ intuitions is problematic. This is the *problem of counterpossibles*.

Some philosophers have been moved by the fact that counterpossibles are true no matter what their consequent is to say that the Lewis-Stalnaker analysis

¹¹See Menzel (2016), for example.

¹²For other problems with the Lewis-Stalnaker analysis, see, for example, Fine (2012). All further discussion in this thesis will be circumscribed to the general approach of using worlds and similarity for a semantic analysis of counterfactuals.

makes all counterpossibles *vacuously* true. I do not think this label is especially informative - I do not see what the adverb “vacuously” adds to truth *simpliciter*. What is vacuous is an universal quantification in the metalanguage not the truth of the counterpossibles. This being said, this terminology has given the name to the two main positions regarding the problem of counterpossibles. Those that think that all counterpossibles are true are called *vacuists*. Those that think that some counterpossibles are false are called *non-vacuists*. I will put my reservations to the side and follow the literature in this respect. This thesis will deal with certain questions that arise in the debate between vacuists and non-vacuists.

1.3.2 Metaphysical Modality

The reasoning we did with respect to (11) is, of course, perfectly general. To get an idea of the dimension of the problem, it will be helpful to have a better idea of what metaphysical modality is.

The first thing to note is that metaphysical modality is supposed to be a *real* or *objective* kind of modality. Though it is not easy to say what this amounts to, the intended opposition is with *representational* kinds of modality.¹³ Again roughly, a modality is representational if it depends on how representational agents represent things. Consider *epistemic* modality. A proposition p is epistemically possible for an agent A if and only if p is compatible with what A knows. Generally speaking, whether a proposition is epistemically possible depends on what representation of reality A has. A real modality is not supposed to be anything like this.

There are various kinds of real modalities. An illuminating contrast is between metaphysical modality and logical modality and between metaphysical modality and nomological modality. Something is logically possible if it is not ruled out by logic alone.¹⁴ Something is nomologically possible if it is logically consistent with the (actual) laws of nature.

It is normally assumed that metaphysical necessity is broader than nomological necessity. It is nomologically necessary that nothing moves faster than the speed of light. But it is metaphysically possible that somethings move faster than the speed of light. All metaphysical necessities are nomological necessities. But not all nomological necessities are metaphysical necessities.¹⁵

In contrast, metaphysical necessity is less broad than logical necessity. It is metaphysically impossible that Bill Gates is my father, but there is no way to

¹³For an attempt at a fuller characterization, see Williamson (2016).

¹⁴Whatever that logic turns out to be. In this thesis, I will assume that there is a correct logic and that the correct logic is classical logic. This much is generally assumed in the debate I want to participate in. This, of course, begs some questions both against non-classical logicians and logical pluralists, who think there is more than one correct logic, such as Beall and Restall (2005).

¹⁵Some philosophers would object to this, saying that metaphysical possibility coincides with nomological possibility. The most salient example is that of scientific essentialists, see, for instance, Shoemaker (1998). In general, I will not engage with such extensionally revisionary conceptions of metaphysical modality.

determine this based on logic alone. On the other hand, every logical necessity is metaphysically necessary.

It would be nice to give a more substantial explanation of what metaphysical modality is. The problem is that philosophers have not made up their minds.¹⁶ Some think that something is metaphysically necessary if it follows from the essences of things.¹⁷ Some think that what is metaphysically necessary depends on our conventions.¹⁸ Some others still define metaphysical modality in terms of counterfactuals.¹⁹ We will return to the problem of finding a conception of metaphysical modality in chapter 4.

For all this variation in terms of what metaphysical modality is, there is a relative consensus regarding its extension, at least in terms of which broad categories make for metaphysical necessity. Sider (2011) puts forward the following list of kinds of metaphysical necessities:²⁰

1. Metaphysical Truths (those discovered by the philosophical sub-discipline of metaphysics; maybe some other philosophical truths should be included)
2. Mathematical Truths (such as the proposition that $2+2=4$)
3. Logical Truths (such as the proposition that this ball is red or not red)
4. Analytic Truths (Roughly, those propositions true in virtue of the meaning of the sentences expressing them)

A proposition is a metaphysical necessity if it appears on this list or if it is a logical consequence of that which appears in this list. Something is metaphysically impossible if its negation is metaphysically necessary. This will be the rough extensional characterization of metaphysical modality we will assume in this thesis.

A counterpossible will be any counterfactual with a metaphysically impossible antecedent as per above. Sometimes it can be helpful to have more discriminating labels such as:

- (11) If Bill Gates was my father, I would be poor [Counter*metaphysicals*]
- (12) If Fermat's Last Theorem were false, Andrew Wiles would be a farse [Counter*mathematical*]
- (13) If this ball were round and not round, we could play football with it [Counter*logical*]
- (14) If some bachelors were married men, it would not be the case that all clergymen are bachelors [Counter*analytics*]

¹⁶For an essay problematizing this situation, see Rosen (2006).

¹⁷Fine (1994).

¹⁸Sidelle (1989) and Sider (2011) are two different implementations of this strategy.

¹⁹Notably, Williamson (2007).

²⁰Sider also wants to *define* metaphysical modality in terms of this list, but let us not worry about that for now.

The epistemological status of metaphysics is worse than the epistemological status of mathematics, logic or the science of synonymy. Due to this, the example I opened with, a countermetaphysical, is more controversial than the examples from the other categories. This does not matter for the discussion of counterpossibles. If metaphysics does deal with factual questions, then one of each two contradicting propositions in metaphysics is false and, hence, impossible. And counterfactuals with antecedent-clauses which express such a proposition will be counterpossibles. Our imperfect knowledge of metaphysics does nothing to falsify this.

1.3.3 Vacuism and Non-Vacuism

Generally speaking, any semantics that predicts that all counterpossibles are true should be called vacuist. And, conversely, any semantics which denies this should be called non-vacuist. The scope of this thesis, however, will be narrower than this.

The representative vacuist position will be the Lewis-Stalnaker analysis. This analysis uses a worlds and similarity framework, where the worlds are to be understood as metaphysically possible worlds.

The non-vacuist positions we will be discussing will *also* be developed in a worlds and similarity framework. This type of non-vacuism tries to solve the problem of vacuism without sacrificing the general approach of the Lewis-Stalnaker semantics. The way it does this involves the use of worlds which do make (or represent) certain metaphysical impossibilities as being true. If a metaphysical impossibility is true at some worlds, a counterpossible is evaluated for truth or falsity in exactly the same way as a counterfactual with a possible antecedent.

The most influential papers defending this form of non-vacuism are Nolan (1997), Brogaard and Salerno (2013) and Berto, French, Priest and Ripley (2017). The best developed defense of vacuism is due to Timothy Williamson (2007; 2017). The discussion between Berto, French, Priest and Ripley (2017), henceforth Berto et al. (2017), and Williamson (2017) will orient most of this thesis.

I will start, in the next section, by presenting the main lines of the vacuist position, as defended by Lewis (1973) and Williamson (2017).

1.4 Vacuism

1.4.1 Connection with Compositional Intensional Semantics

As I said above, our attention will focus on worlds and similarity approaches. A first thought you might have is that introducing impossible worlds seems rather extreme as an answer to the problem of counterpossibles. Maybe something else can be done using only possible worlds and a bit of ingenuity. It seems

possible that there is some fix of the Lewis-Stalnaker analysis that can solve the problem. Maybe the problem of counterpossibles is something parochial to the Lewis-Stalnaker analysis. Williamson (2017: 197) tries to show that this is not the case.

Assume that you want to provide a compositional semantics for counterfactuals using only possible worlds. Let A and B be declarative sentences and let their intension be written as $|A|$ and $|B|$. If the semantics is to be compositional, the intension of a counterfactual must be a function of the intensions of its antecedent- and consequent-clauses:

$$(15) |A \Rightarrow B| = f(|A|, |B|)$$

Plausibly, our grasp of the semantics of the counterfactual conditional tells us that the only consequent-worlds that make a difference are also worlds where the antecedent is true. If this is legitimate, (15) amounts to the following:

$$(16) |A \Rightarrow B| = f(|A|, |A| \cap |B|)$$

If A expresses an impossible proposition, then $|A| = \{\}$. Therefore the following is true:

$$(17) \text{ If } |A| = \{\}, \text{ then } |A \Rightarrow B| = f(\{\}, \{\})$$

This is already enough to conclude that all counterpossibles will have the same truth-value, no matter what their consequent is. What that truth-value is is still open. But it seems natural to think that $A \Rightarrow A$ should always be true. If that is so, then all counterpossibles should be true.

If correct, this reasoning shows that vacuism is no accidental feature of the Lewis-Stalnaker analysis. It follows from the general framework where the theory is developed, what Williamson calls “compositional intensional semantics”.

I think this argument is a good one in that it places precise conditions on what a non-vaculist theory should deny. But I do not think it proves that all theories in “compositional intensional semantics” must be vacuist unless a overly restrictive conception of this term is being used.

The objectionable step is the one from “ A expresses an impossible proposition” to $|A| = \{\}$. This assumes that all there is to the intension of a declarative sentence is the set of possible worlds where it is true. This does not seem mandatory. Yablo (2014), for example, tries to show that some distinctions between sentences which are true at the same possible worlds can be captured in a possible worlds-only framework. What distinguishes them is that they are *about* different things, that their *subject matter* is not the same. Imagine that we have such a theory where the intension of a proposition is an ordered pair of the set of possible worlds where it is true and its subject matter, whatever that turns out to be. We could then give an account of counterfactuals where the step above is invalid. The set of possible worlds collapses when the proposition expressed is impossible, but its subject matter does not. This could allow us to distinguish between the truth-values of different counterfactuals. I am not saying that this is a *good* theory of counterfactuals. What I am claiming is that it deserves to be called an *intensional* semantics and it is one that avoids Williamson’s argument.

All that being said, Williamson’s point is well taken. In a *possible* worlds and similarity framework vacuism looms large, depending only on some reasonable sounding assumptions. This is part of the motivation to enlarge the space of worlds being used, as we shall see.

1.4.2 Hybrid Theories

There is one further requirement that Williamson (2017: 211) thinks a successful account of counterpossibles must meet. It seems intuitively plausible that what we have been calling “counterfactuals” is a reasonably natural semantic kind: each counterfactual resembles more other counterfactuals, in matters of relevance for semantics, than members of other classes of sentences. This suggests a *prima facie* advantage for theories of counterfactuals that postulate a single mechanism of semantic evaluation for all counterfactuals. This places what we can call *hybrid theories* at a disadvantage. If a theory treats a subclass of counterfactuals in one way and a disjoint class of counterfactuals in a very different way, this is a cost for that theory, a cost that needs to be explained.²¹

What this means for the case of counterpossibles is that a theory that accounts for our semantic evaluation of counterpossibles in a very different way from the account it gives in the case of counterfactuals with a possible antecedent is to be dispreferred. What would explain our use in natural language of such a disjunctive class of sentences? What justifies the lack of economy of cognitive mechanisms that a hybrid theory needs? Why is the metaphysical impossibility of the antecedent sufficient to trigger whatever difference of semantic evaluation is in cause?

I think that all these considerations are somewhat defeasible. Nevertheless, it is a good point to keep in mind. A semantics of counterfactuals that treats counterpossibles and counterfactuals very differently is to be mistrusted.

1.4.3 Lewis’ Motivation for Vacuism

At this point you might reasonably ask yourself if the vacuist has any intuitive story to motivate vacuism. I know only of one such positive account, the one given by Lewis (1973: 24). This passage has been forgotten in more recent literature, but I think it is very plausible and it will play an important role in this thesis.

Lewis says the following:

There is at least some intuitive justification to make an “would” counterfactual with an impossible antecedent come out vacuously true. Confronted by an antecedent that is not really an entertainable supposition, one may react by saying, with a shrug: If that were so, everything you like would be true! (Lewis 1973:24)

I think this is a very reasonable thought to have. Do not worry about what “entertainable” means precisely at this point. It does seem that sometimes

²¹We will focus of this ideology of costs and benefits in more detail below.

we are faced with counterfactuals whose antecedents we cannot develop into anything of substance. In those cases, our process of counterfactual evaluation breaks down, as it were. We could interpret this breakdown as a truth-value gap. But we might not want truth-value gaps for other theoretical reasons - maybe we want to keep classical logic intact. If we are to attribute a truth-value to such counterfactuals, it is going to be an arbitrary one. Since we want $A \Rightarrow A$ to always be true, we end up with vacuism with regards to counterfactuals with unentertainable antecedents.

I suspect that one of the reasons this story has been ignored has to do with the fact that it is hard to believe that all metaphysical impossibilities are unentertainable. In what sense can I not entertain the scenario where Bill Gates is my father? It seems that I can, and that I can know a lot of things that would be true under that counterfactual supposition. If this is so, then Lewis' story cannot be used as a complete solution to the problem of counterpossibles.

Moreover, contemporary readers will tend to read "entertainable" as "conceivable". There is a huge body of literature on the link between conceivability and possibility, on the question of whether what is possible is what can be conceived.²² The general consensus is that the link does not exist, at least for non-idealized versions of conceivability.²³ It is easy to read Lewis as making this mistake here, understandable as that is given the time when the passage was written.

But I think this is reading too much into it. Insofar as "conceivability" describes a natural kind of cognitive abilities, it is not clear what that is. It is very hard to see the difference between "imagining", "supposing" and "conceiving", for example.²⁴ I think that Lewis meant something less technical by "entertainable", a certain kind of intuitive feeling of puzzlement we get when we are asked what would be the case if something very weird was the case. The key property of unentertainable propositions is our inability to counterfactually suppose them successfully. At least, that is how I will use the expression in what follows. I will leave it open whether there is some form of conceivability that corresponds to this.

I think that unentertainability considerations can be used to explain why some counterfactuals can be reasonably thought to be vacuously true. If a certain class of propositions seems unentertainable, then vacuism about that class of propositions seems somewhat justified. One problem is that what is unentertainable seems context dependent. Sometimes context gives us enough to develop a counterfactual supposition and sometimes it does not. Consider the following:

(18) If 6 plus 6 were equal to 13, then 0 would be equal to 1

Just uttered out of the blue, I think (18) is a good candidate for unentertainability. How do we go about knowing what would follow from the fact that such

²²See the papers in Hawthorne and Gendler (2002) for an introduction to the topic.

²³See Chalmers (2002) for discussion of idealized versions of conceivability.

²⁴See Balcerak Jackson (2016) for a discussion of the interplay between these three terms.

a basic arithmetical fact changes? But sometimes context helps. If we were in a context where this was the *only* basic arithmetical fact that changes, the counterfactual seems true. Just subtract 12 to each side of $6+6=13$ and you get $0=1$. Without knowing that this is something you can do, the situation is sufficiently strange to count as unentertainable. With some background facts to help maybe it is not.

We should keep this in mind when we want to use the unentertainability strategy. That being said, it can be dialectically very useful for the vacuist to have an intuitive story to back her truth-value predictions, at least in some cases.

1.4.4 Pragmatics

Part of the story of the non-vacuist involves a pragmatic explanation of how accepting non-vacuism does not affect significantly our ordinary linguistic behavior. We use sentences for various communicative purposes in everyday life. Oftentimes, our selection of which declarative sentence to use involves the notion of truth. We include for consideration candidates that we take to be true and we exclude from consideration candidates we take to be false. Vacuism implies that all counterpossibles are true. This means that we cannot use the notion of truth to guide our linguistic use of them, as sketched above. The vacuist must have some alternative story. In what follows we will analyse three topics discussed in Williamson (2017) - why not all known counterpossibles can be asserted, the case of *reductio ad absurdum* and the heuristics which explain why we had the intuition that some counterpossibles were false in the first place.

Unassertable Counterpossibles

Williamson (2017: 204) asks us to consider the following counterfactual, uttered in a discussion of seventeenth century English politics:

(19) If Hobbes had squared the circle, he would have become Lord Chancellor

(19) is a counterpossible, therefore it is true. Furthermore, I can know it to be true if I know that its antecedent is an impossibility and I recognize the logical implication between an impossible antecedent and a true counterfactual. Moreover, (19) seems on topic, it is about seventeenth century English politics. Why, then, can I not utter (19)?

Williamson's reply is that it violates the Gricean conversational norm of quantity (Grice 1989). In the situation described I would also be in the position to assert the following:

(20) Hobbes could not have squared the circle

(20) implies (19) and, in fact, it is my ground for asserting (19). This makes the straightforward assertion of (19) misleading. The interlocutors of the person who asserts (19) can be misled into thinking that that person knows of some

hidden political significance of squaring the circle, maybe gaining the recognition of an important secret society. Since the grounds for (19) are much more mundane, this fails to accord to the general principles of cooperation that give rise to conversational norms.

This process is perfectly general and allows us to explain why the explosion in the number of true counterfactuals does not correspond to an explosion in the number of assertable sentences. This is very important insofar as that prediction would clash head-on with our ordinary linguistic behavior.

Reductio Ad Absurdum

One particularly important use of counterpossibles seems to be in proofs by *reductio ad absurdum*. Such proofs are based on the truth of counterfactuals of the form

(21) If α were true, then β would have been true

Where α is a mathematical proposition to be proven false and β is something considered a manifest absurdity in this mathematical context, maybe an explicit contradiction.²⁵

Since α is a false mathematical proposition, it is metaphysically impossible. This makes (21) a counterpossible. Its truth is independent of the fact that β is absurd. But the fact that β is absurd seems important to explain the behavior of the mathematicians. It is the reason why they choose (21) instead of another counterfactual with the same antecedent.

What else could explain this choice? What is so special about (21)? Williamson (2017:214) claims that there is something distinctive about the context of mathematical practice that allows us to explain this.

Not everything that is true can be uttered in the process of giving a mathematical proof. It seems that we can only utter sentences which are either axioms of the mathematical theory we are working in or sentences for which we have a previous mathematical proof. What is special about (21) is that we have a previous mathematical proof for it. The absurd proposition can be *derived* from the antecedent. This is what makes (21) assertable and all the counterpossibles with the same antecedent unassertable.

What saved the practice of the mathematicians in this case was that assertability of a counterfactual required a further connection between the antecedent and the consequent, that of mathematical proof, that is not present in all the true counterfactuals with the same antecedent. Again, this can be generalized somewhat, though Williamson himself does not do so.

Consider the context of metaphysical argumentation. Arguments that use counterfactuals with the form of (21) but where α is a metaphysical proposition are quite common. A metaphysician assumes that something is true in order

²⁵Williamson (2017: 213) recognizes that maybe all such arguments can be rewritten without the use of counterfactuals but claims that it is natural to assume that they do not *need* to be so rewritten. I will just assume he is right in what follows. Adjudicating the question would presumably be an empirical project in the semantics of actual mathematical practice.

to show that it has absurd consequences and with the ultimate goal of showing that the assumption is false. The context of metaphysical discussion does not seem to require that there is a mathematical proof for everything asserted. But maybe it does require that there must be an *argument* for what is asserted. If this is true, then what distinguishes the uttered counterpossibles from the inappropriate ones that share the same antecedent is that there is an argument for the former and no argument for the latter.

This is one more pragmatic strategy open for the vacuist. In general, it must be used with care. If we postulate artificial pragmatic requirements to any given context, we can dispense with the pragmatic utility of truth altogether. Maybe *all* sentences could be false and we would have the same linguistic behavior, ruled only by the additional requirements for utterance we postulated. Whether the strategy is plausible, then, will depend on a case-by-case analysis.

Heuristics

When a theory tells us that our intuitions must be revised, asking what led us astray is a very natural thing to do. Providing such an explanation can help us learn to live with the revised intuitions. Philosophers often call such stories *error theories*. Williamson (2017: 215) provides such an account in the case of counterpossibles.

Williamson's story is that we use a certain heuristic when evaluating counterfactuals that has two characteristics, (1) it is very successful in most cases, and (2) breaks down when it comes to counterpossibles. We can present the heuristic by noting that sometimes examples of counterpossibles are presented in pairs. Consider the following classic examples by Nolan (1997: 544):

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

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

(22) seems true and (23) seems false. A plausible way to reach this conclusion is as follows. We start by supposing that Hobbes squared the circle. We assume that this involves some kind of geometrical treatise writing in England, where he lived. We then ask ourselves whether the sick children in the far-away South America would have cared about this. Since the children plausibly had other problems occupying their minds, we conclude that they would not. (22) is true. After we conclude this, we judge (23) as false almost automatically. This is due to the fact that the consequent of (22) is the negation of the consequent of (23). What is guiding our intuitions is the following heuristic (Williamson 2017: 219):

(HCC*) If you accept one of $\alpha \Rightarrow \beta$ and $\alpha \Rightarrow \neg\beta$, reject the other²⁶

(HCC*) is, generally speaking, quite useful. It eases the cognitive demands on counterfactual evaluation. When we evaluate a counterfactual we start by

²⁶(HCC*) is the label Williamson uses.

supposing the truth of its antecedent. We then ask whether the consequent is true under that supposition. We develop the supposition trying to find that the consequent is true. What do we do if we find that something inconsistent with the consequent is true? It is reasonable to assume that we should stop our search. Continuing would be a waste of time and cognitive resources. We can conclude instead, using (HCC*) that the counterfactual is false. Unfortunately, (HCC*) is false given vacuism. When a proposition is impossible, all counterfactuals which have it as an antecedent are true. We have a good heuristic, one which we have good reason to adopt, but which leads us the wrong way when counterpossibles are around.

At this point, it can seem that all that Williamson did is quite circular. The explanation can have some role to play in vacuist apologetics but it will not do much to convince a non-vacuist. But Williamson goes on to claim that even non-vacuists should accept that (HCC*) is false. He considers the following pair of counterfactuals:

(24) If $p \wedge \neg p$ were true, then p would be true

(25) If $p \wedge \neg p$ were true, then $\neg p$ would be true

It seems that the non-vacuist should accept both (24) and (25) as true. But the pair is a counterexample to (HCC*). The only option is to reject the following reasonable looking principle of counterfactual logic:

Conjunction Elimination: $\models (\alpha \wedge \beta) \Rightarrow \alpha$

This provides independent justification to reject (HCC*) and, therefore, makes Williamson's case stronger. He found an heuristic which seems to have significant practical value but which must be considered wrong by every party in the debate.

1.5 Methodological Considerations

We have seen the main defensive moves of the vacuist position. Vacuism entails that some of our intuitions regarding the truth-value of counterfactuals are wrong. And we surveyed some techniques that can allow us to live with this state of affairs. We now have to ask: why be a vacuist in the first place? Why should we learn to live with our intuition's being mistaken and not prefer an account that avoids this state of affairs?

Williamson's reply is that being a non-vacuist implies accepting other theoretical costs which offset the ones associated with the violation of our truth-value intuitions. Examining the nature of these costs and possible replies on the non-vacuist's behalf is going to be the main focus of this thesis. But first, we must bring to light a hidden methodological assumption behind Williamson's position.

1.5.1 Semantic Deontology

Why isn't vacuism simply wrong? In a sense, this is one of the central questions of this thesis. Vacuism is an account of the semantics of counterfactuals. Semantics is in the business of explaining how natural language works. Its data is ordinary speakers' semantic intuitions and linguistic behavior. Vacuism does not correctly account for the data. Therefore, it is a false semantic theory.

The reasoning above can be codified in the following methodological maxim:

Semantic Deontology: If a semantic theory has an intuitive counterexample, then it is a false semantic theory.

By "intuitive counterexample" I mean a sentence S such that the theory predicts that S has a certain truth-value but our linguistic intuition tell us that S has a different truth-value.

One way to question the status of a purported counterexample is to reject the intuitions in question. Maybe they are unstable across languages or across individuals. But no such thing has been tried in the case of counterpossibles, as far as I know. Therefore, semantic deontology would imply that vacuism is a false semantic theory.

The question then becomes: is semantic deontology true? Williamson has an argument to the effect that it isn't.

1.5.2 Analogy with Universal Quantification

Williamson (2017) develops an analogy between counterpossibles and vacuous universal quantification. I will interpret this analogy as an argument against semantic deontology and claim that Berto et al. (2017) have misunderstood its point.

The relevant fact about the semantics of universal quantification is this. A sentence like

(26) Every A is an F

is true when there are no A 's. Williamson (2017: 205) claims that this follows from our best theory of quantification. In this case, we can say that (26) is vacuously true. This seems to go against our linguistic intuitions. Consider the following:

(27) Every golden mountain is a valley

Intuitively, (27) is false. Mountains cannot be valleys, independently of their material constitution. But, given that there are actually no golden mountains, (27) is true.

Here is how Berto et al. understand the point of this analogy:

The intended analogy with vacuous quantification is clear: if there are no circumstances in which the antecedent of a counterfactual is

true, then counterfactuals with that antecedent are true. This is because for a counterfactual to be untrue there must be circumstances at which its antecedent is true. But the analogy looks question-begging: what is at issue is whether there are such circumstances, not what would happen in their absence. (Berto et al. 2017:16)

This response seems to assume that Williamson’s point with the analogy is the following. Everyone accepts that the correct semantics for universal quantification makes universally quantified statements vacuously true when the subject term is empty. Both parties agree that the truth conditions for counterfactuals involve a universal quantification in the metalanguage. When that subject term is empty counterfactuals are vacuously true.

If this were Williamson’s argument, Berto et al.’s response would be fair. Since his opponents do not think that the fact that a counterfactual has an impossible antecedent is sufficient to make the relevant universal quantification vacuously true, it would be dialectically inefficient to use this argument. But I just do not see any textual evidence that this is what Williamson meant.

One use that Williamson has for the analogy is in giving us some confidence that we can live with the vacuity of counterpossibles. This point is purely defensive. It is based on the fact that vacuous universal quantifications and counterpossibles share the following characteristics: (a) Every member of the class is true (sometimes against intuition), and (b) membership to the class can be determined without considering the semantic value of all constituents of the expression in question (What F is does not matter for something being a vacuous universal quantification anymore than what the consequent is matters for something counting as a counterpossible). This similarity holds out hope that both classes of expressions can receive a similar pragmatic treatment.

But I think the main point of making the analogy is to argue against semantic deontologism. Williamson is using the analogy for this effect when he says:

In the case of the universal quantifier, proper understanding was finally achieved through systematic, highly general semantic and logical theorizing, rather than by a more data-driven approach. The same may well hold for the counterfactual conditional. At any rate, it is methodologically naive to take the debate over counterpossibles to be settled by some supposed examples of clearly false counterpossibles. (Williamson 2017:224)

The argument goes like this. Actual semantic practice does not adhere to semantic deontologism, as is evidenced by the case of vacuous universal quantification. And actual semantic practice has been very successful. This shows that semantic deontologism is not the right methodology to use.²⁷

From the quote above, we can also identify Williamson’s alternative methodology. We can call it *semantic consequentialism*:

²⁷Notice the inference from a descriptive premise to a prescriptive conclusion here. Though I take it to be reasonable in this instance, we should be aware of this fact about Williamson’s move.

Semantic Consequentialism: 1) Semantic Deontology is false; 2) The correct semantic theory is arrived at by a process of theoretical *cost-benefit analysis* in which respect for people’s truth-value intuitions is only one of many factors (though possibly a very weighty one).

Semantics stands in certain relations with other fields of philosophy. Our semantics of counterfactuals affects which positions we can take in other fields, such as metaphysics or epistemology. If a particular semantics of counterfactuals forces us to reject good positions in these fields, or to accept bad ones, then that is a cost for that semantic theory. And such a cost can offset a certain (limited) lack of empirical adequacy with the data the semantic theory should account for.

This seems to be the methodological picture envisaged by Williamson. We can now ask ourselves whether his argument is convincing.

1.5.3 Sophisticated Semantic Deontology

Williamson’s argument seems sufficient to establish that semantic deontology is false. But, or so I will argue, there is a more plausible methodological principle that also implies that vacuism is false and which is not vulnerable to Williamson’s argument.

A striking feature of Williamson’s example is that our intuitions regarding universal quantification sentences may not be consistent. We seem to hold the following intuitions (see Peters and Westerstahl 2006: 124 for discussion):

- There are sentences of the form “Every A is F ” that are false even if there are no A ’s.
- A sentence of the form “Not every A is F ” is true only if there is an A that is not a F .
- “Not every A is F ” is the negation of “Every A is F ”.

Faced with “Every golden mountain is a valley” we have the intuition that the sentence is false, even though there are no golden mountains. But we also think that “Not every golden mountain is a valley” is true only if there is a golden mountain that is not a valley. Our intuitions tell us that, when there are no golden mountains, both a sentence and its negation are false. Something must go. And the semantic analysis Williamson thinks is successful dispenses with the first intuition.

The case of counterpossibles seems disanalogous. Our intuitions about counterpossibles seem perfectly consistent. That this disanalogy matters can be made clear by considering the following methodological principle:

Sophisticated Semantic Deontology: If a semantic theory has an intuitive counterexample and our intuitions regarding the object of the semantic theory are consistent, then it is a false semantic theory.

Subscribing to Sophisticated Semantic Deontology is enough to make vacuism false and still explain what is going on in the case of vacuous universal quantification. This methodology still takes semantics to be essentially descriptive and is therefore a “data-driven approach” as Williamson puts it. Semantics is in the business of providing precise models for the meaning of natural language sentences. In doing this it must respect ordinary speakers’ intuitions. But the models must be consistent. This means that when the intuitions to be respected are inconsistent, then some of them must go and be explained away. But this is the *full extent* of semantics’ revisionary power with respect to linguistic intuition.

It seems to me that this is a very reasonable position to take in the methodology of semantics. It is as *prima facie* attractive as the first version of semantic deontology was.²⁸ And Williamson presents no argument against this particular methodological principle.

I do not know if Sophisticated Semantic Deontology is true. Arguing for methodological matters is quite hard. Doing so tends to beg questions against some of the positions involved (in this case, against vacuism). And it is unclear what considerations can be used for both sides.

That being said, the discussion between Williamson (2017) and Berto et al. (2017) assumes semantic consequentialism. Insofar as this thesis extends this discussion, I will be assuming it too. I hope that the discussion of this section makes this assumption clearer than it was before. If you are persuaded by Sophisticated Semantic Deontology (or something similar to it) you will find the ensuing debate to be a waste of time. Hopefully, some of the points I make in discussing it can still be of more general interest.

1.6 The Structure of the Thesis

Let us, then, assume semantic consequentialism. Williamson (2017) goes on to claim that being a non-vacuumist has a number of costs that outweigh the benefit of accounting for our truth-value intuitions regarding counterfactuals. The thesis will be structured around a discussion of these costs.

I hope to do three things in what follows:

1. Make the debate between vacuumists and non-vacuumists clearer, both by unearthing both parties’ assumptions and by clarifying the arguments given;
2. Point out where the current literature went wrong;
3. Defend a moderate non-vacuumist position which has not been adequately developed in existing literature.

²⁸Further examples that can seem to motivate sophisticated semantic deontology include the problems with a naive theory of truth and some impossibility results regarding indicative conditionals. For an exposition of the first see Glanzberg (2013) and for a classic paper on the second see Gibbard (1981).

The structure of each chapter follows roughly the structure dictated by these goals. I must confess at the outset that I am more confident of having made progress towards 1. and 2. than towards 3. Nevertheless, I think it is important to understand better the structure of the conceptual space of answers to the problem of counterpossibles.

The moderate non-vacuumist theory I have in mind, which was also defended by Kment (2014), will make some counterfactuals with metaphysically impossible antecedents come out false. But it will not attempt to completely eradicate all forms of vacuumism. Counterlogicals and counterfactuals whose antecedents violate the necessity of identity will still be vacuously true. My motivation for defending this view is two-fold. First, it will help us to understand better what is involved in being a non-vacuumist. The position of Nolan (1997) and Berto et al. (2017) is very extreme - it uses so many impossible worlds that every proposition whatsoever is true at least in one of them. The consideration of a more moderate version will make clear which parts of their theory are indispensable for non-vacuumism and which are not. Secondly, I think a moderate form of non-vacuumism highlights better the connection between vacuumism and metaphysical modality - the real problem with vacuumism, one might say, is that it takes metaphysical modality too seriously in semantic practice.

Here is a rough outline of the rest of the thesis.

The non-vacuumist positions we will be considering use impossible worlds. There are general worries about using these objects in developing semantic theories. Chapter 2 presents these worries and sketches a way to answer them. In the end, I explain which impossible worlds I intend to use in my preferred non-vacuumist position.

Using impossible worlds seems to commit us to a serious weakening of our logic of counterfactuals. Impossible worlds can be logically anarchic and their unruliness infects the logic of counterfactuals. Chapter 3 describes this problem and argues that the currently more popular approach among non-vacuumists to solve it does not work. Finally, I defend the position that only logically possible worlds should be used in the semantics of counterfactuals in order to solve the problem.

This dissociation between counterfactuals and metaphysical modality may not be a significant cost for our semantics of counterfactuals. But maybe it is a significant cost for our understanding of metaphysical modality. In chapter 4 I evaluate the extent of this problem. Williamson uses counterfactuals to provide both an epistemology for, and a conception of, metaphysical modality. I argue that losing the epistemology is not a significant cost. Losing the conception might be. But it is hard to know for sure given the current state of the literature.

Non-vacuumists have tended to defend that counterfactuals do not respect the substitutivity of identicals. Williamson claims that this is a problem when coupled with the thought that counterfactuals are worldly sentences, that they do not depend on the way that representational agents represent things. In chapter 5, I will argue that the answer given by Berto et al. does not work. I will suggest that this motivates a non-vacuumist semantics where the substitutivity of identicals is indeed valid.

Chapter 2

Impossible Worlds

The non-vaculist semantics that we will consider in this thesis use impossible worlds. In this chapter we will consider a cluster of questions that arise when impossible worlds are used in semantic theorising. Section 1 clarifies what it means to “use impossible worlds in a formal semantics” and sketches a picture of the dialectical situation when it comes to the ontology of impossible worlds. Section 2 considers two objections that have been made against the use of impossible worlds in natural language semantics. I try to clarify what the problems are and evaluate their severity. I then sketch a way of thinking about the use of impossible worlds that provides a deflationist path to answer the objections. In section 3 I put forward the restrictions on the space of impossible worlds I will be using to develop my alternative non-vaculist semantics and explain how evaluating its merits blends into the structure of the chapters that follow.

2.1 The Ontology of Impossible Worlds

2.1.1 Using Impossible Worlds

Philosophers often develop *formal semantics* as part of their answer to a given problem. They do this because it can be easier to be precise about what a particular thesis says, and about what its consequences are, in a formal setting. What follows is a rough sketch of how this process might go with respect to a propositional language. It will allow us to see the way in which impossible worlds enter the picture.

The first step in presenting a formal semantics is defining a *formal language*. A set of symbols, the *alphabet* of the language, is singled out and a characterization of which strings of symbols of the alphabet are to be considered well-formed formulas is given. The set of well-formed formulas can be identified with the formal language itself.

The second step is often the development of a *model theory* for the formal language. A model is an ordered n -tuple of set-theoretical structures that can be

used to specify the notion of truth-at-a-model for the formal language in question. Some model theories define truth-at-a-model via an intermediate step. These models contain an element that is a set of objects C , which we will call *circumstances*, and a *valuation function* V that assigns to each member of the set of circumstances the set of formulas true at that circumstance.¹ This information is enough to define the relation of true-at-a-circumstance(-at-a-model). A formula φ is true at a circumstance c at a model M if and only if φ belongs to $V(c)$. Truth-at-a-model can be defined as truth in every circumstance of the model.

At this point we have a formal language and a model theory that specifies what a model for the language is and which tells us how the notion of truth applies at the level of circumstances and at the level of models. Given what we have said until now, we have no way of pairing up the formulas of the formal language in question with the sentences in natural language which we might be interested in as philosophers. Moreover, the objects that are members of the set of circumstances can be any objects whatsoever and the fact that a valuation function pairs a circumstance to a set of formulas has no particularly deep philosophical meaning.

The next step in developing a formal semantics tries to correct this situation.² An interpretation of the models is given. Here is a way this can be done. We can define an *interpretation function* that assigns the set-theoretical constructions of our models to set theoretical constructions whose elements are entities in the real world. This builds in a significant assumption: that the interpretation function is *ontologically serious*, that whatever is assigned by the interpretation function to an element of our model *exists*.³

I will assume that we will interpret the formulas in our language as *proposi-*

¹I am trying to be perfectly general at this point. I am saying nothing about the truth value of the formulas that are *not* assigned to a circumstance by the valuation function for example. Furthermore, at this point, any formula can be assigned to a circumstance by the valuation function, irrespective of its syntactic complexity. It is often the case that the assignment of truth values to formulas up to a certain level of syntactic complexity is enough to extend the assignment to the whole language. None of this is relevant at the present level of abstraction.

²We can characterise this step as a transition from “pure semantics” to “applied semantics”. This terminology was apparently coined by Plantinga (1974). Lewis says the following regarding the pure semantics of modal logic: “For that job we need no possible worlds. We need sets of entities, which, for heuristic guidance, ‘may be regarded as’ possible worlds, but which in truth may be anything you please. We are doing mathematics, not metaphysics” (Lewis 1976: 17). The point of using possible worlds in that context is to fix the “correct” or “intended” interpretation of the semantics of modal logic (Lewis 1976: 19-20).

³The usual argument for ontological seriousness is quinean in nature (see Lewis 1973: 84). We quantify over circumstances in the metalanguage we are using to provide the semantics. We are ontologically committed to whatever we quantify over (the quinean premise). Therefore, we are ontologically committed to the circumstances we use. Evaluating the merits of this argument is outside the scope of this thesis and I will take ontological seriousness as an assumption. This assumption begs the question against certain views on the interpretation of formal semantics. Modal fictionalism, for example, thinks, roughly, that the benefits of the use of possible worlds can be reaped by understanding our sentences ranging over them in the metalanguage as “true according to the possible worlds fiction”. For more on modal fictionalism, see Nolan (2016).

tions, whatever their ontological nature turns out to be.⁴ A translation between the formulas and the sentences of natural language which might be the focus of our philosophical interest can be defined as follows (provided that we know which propositions the sentences in the natural language express in the first place). A formula φ is the translation of the natural language sentence S if and only if S express the proposition that is the interpretation of φ .

The interpretation of the valuation function depends on the interpretation we give to the circumstances. The valuation function is the thing that assigns to a circumstance the formulas we want to interpret as being true at that circumstance. The relation between the kind of thing a circumstance is and the notion of truth of a proposition (our assumed interpretation of what the formulas are) will determine the ontological character of the valuation function. An example: if a circumstance is interpreted as being a book in the library of the University of Amsterdam, maybe $V(b)$, for book b , contains the formulas which express the same proposition as some declarative sentence in that book.

It is when we wish to interpret what circumstances are that impossible worlds come into our story. Impossible worlds are candidates to play the circumstance role in a formal semantics. The minimal content of the notion of impossible world is this: a circumstance c that is such that it can happen that (i) p is a metaphysically impossible proposition; (ii) $V(c)$ contains a formula φ which is interpreted as p . These conditions are not sufficient for being an impossible world: I bet that a lot of books in the library of the University of Amsterdam contain declarative sentences which express impossible propositions as well. Furthermore, it tells us nothing about what an impossible world is, ontologically speaking.

It is not easy to say which conditions are sufficient for something counting as an impossible world and I will not try to.⁵ Roughly speaking, the philosophical literature treats something as an impossible world if it is like a possible world, ontologically speaking, but meets the conditions (i) and (ii) above. I will use this as my working definition.⁶

Possible worlds are circumstances that are taken to be at least (a) maximal: if a formula does not belong to $V(w)$ then it is false at w ; (b) consistent: if

⁴For an overview of the debates surrounding the metaphysics of propositions, see McGrath and Frank (2018). Some of the more coarse-grained conceptions of propositions will turn out not to be very apt for the interpretation of many propositional languages of philosophical interest. This will not be a problem in what follows: a proponent of such a conception of propositions can even interpret the formulas of the formal language directly as the sentences of natural language she is interested in.

⁵This makes the debate between the impossible world theorist and someone like a situations-theorist, in the style of Barwise and Perry (1981), very unclear. What exactly makes an impossible world allowed to be non-maximal different from a situation? Luckily, the debate that will be central to this thesis is between the possible worlds theorist and the impossible worlds theorist, a much clearer contrast.

⁶I do not intend it as anything stronger than that. This idea has been called “Parity Thesis” (Priest 1997: 580-1). For an ontology of impossible worlds which explicitly denies the Parity Thesis, defending a Lewisian realism for possible worlds and ersatzism for impossible worlds, see Berto (2010). For a survey of other possible characterisations of impossible worlds, none of which can cover the full use they are given in practice, see Berto (2013).

a formula is false at w then it does not belong to $V(w)$; (c) Metaphysically possible: if p belongs to $V(w)$ then p is metaphysically possible.

A number of different suggestions about what possible worlds are has been suggested in the literature.⁷ I will call each possible suggestion in this context an *ontological story* about what possible worlds are. At the level of abstraction I am working at, it will be useful to quantify over and make claims about ontological stories.

Given my working characterization of impossible worlds, what they are is dependent on what ontological stories can be told about possible worlds. An ontological story for impossible worlds is a certain adaptation of an ontological story about possible worlds: such is the import of “being like a possible world” in my characterization above. Given this, there better be some ontological stories that permit such an adaptation. In the next section I will present a brief argument for the conclusion that there are.

2.1.2 Adaptation of Ontological Stories

Let us start by clarifying a bit further what an ontological story is supposed to be. An ontological story for possible worlds is a specification of a set of properties that an object must satisfy to count, according to the story, as a possible world.⁸ The same goes for impossible worlds, of course.

Some ontological stories are better than others. We want to use possible worlds as circumstances in a formal semantics. And we have assumed that the interpretation of the formal apparatus is ontologically serious. So, a satisfactory story about possible worlds should be at least such that there are *some* objects that satisfy the set of properties in question.

In the case of possible worlds, a further condition on the adequacy of ontological stories emerges. We have a certain theoretical notion of what is metaphysically possible. Call a scenario (a word that is purposefully vague) that is metaphysically possible a *possibility*. Our ontological story must not conflate between different possibilities. If two possibilities are different they must be represented differently in the space of possible worlds. In other words, there must be *enough* possible worlds. Call an ontological story that has this property *sufficient*.

But there cannot be too many. No possible world can represent something that is metaphysically impossible. A large class of ontological stories guarantees that this requirement is met by brute force. Given whatever notion of representation is adequate to the objects in question, one of the properties of the objects that are to count as possible worlds is that they do not represent an impossibility. Call this requirement, and any similar requirement that uses modal notions explicitly, a *modal* requirement.

⁷For an overview of the full range of options here, see Menzel (2016).

⁸It is possible that we would have to place some restrictions on languages eligible to express this specification. Maybe they should be particularly simple, unambiguous, precise and so on. Let us take some such conditions as understood in what follows.

Philosophers have argued that the fact that this requirement is enforced directly is not an obstacle to the explanatory role possible worlds play in formal semantics.⁹ But this does render an historically important role for possible worlds impossible to fill: that of providing a reductive analysis of metaphysical modality.¹⁰

A reductive analysis of metaphysical modality is provided by an ontological story about possible worlds if and only if it has the following property:

(*Reduction*) I) The ontological story does not contain any modal requirement; II) A proposition is metaphysically necessary if and only if it is true at every possible world according to the ontological story.

The reductive part is captured in the first condition, designed to avoid circularity. The second one guarantees extensional adequacy.

I have now said enough to state my claim about the adaptation of ontological stories about possible worlds to ontological stories about impossible worlds.

Claim: Any sufficient ontological story that does not meet (*Reduction*) due to the failure of II) characterizes at least one impossible world.

It is clear that such an ontological story about possible worlds would not be satisfactory: somethings that count as possible worlds make impossibilities true. But this problem can be solved by adding the modal requirement above to obtain a satisfactory story. The adaptation we want is obtained simply by dropping this requirement. By doing this, we turn an ontological story about possible worlds into an ontological story about impossible worlds.

It is easy to see that *Claim* is true. Condition II) can fail in two ways. There can be some proposition that is not metaphysically necessary but true at every (putative) possible world or some proposition that is metaphysically necessary that is false at some (putative) possible world.

The first way is blocked by sufficiency. If a proposition is not metaphysically necessary there is a scenario where it is false which represents a possibility. Since the proposition is true at every object that meets this story, there is a possibility that is not represented.

Then it must be the case that there is some proposition that is metaphysically necessary that is false at some object. But the negation of a metaphysically necessary proposition is metaphysically impossible. This means that there is a proposition that is metaphysically impossible and which is true at one of the objects characterized by the ontological story. This is an impossible world.

Let us summarize what can be concluded from this argument. Philosophers advance ontological stories about possible worlds. Some of them do not satisfy (*Reduction*) which forces them to include modal requirements by brute force.

⁹See Plantinga (1974) and Stalnaker (1976) for two early non-reductive theories of possible worlds.

¹⁰That was, of course, one of the main goals of Lewis (1986).

If we take one of those ontological stories and we drop the modal requirements we have an ontological story that characterizes at least an object that is an impossible world.

This result is dialectically important. Impossible worlds are normally introduced in philosophical discussion to remedy a perceived defect of a possible worlds theory. The problem of counterpossibles is a case in point. Therefore, one of the most natural dialectical opponents of the impossible worlds theorist is the possible worlds theorist. What the argument above shows is that the possible worlds theorist has two options. The first is to claim to have provided a reductive analysis of metaphysical modality based on her ontology of possible worlds. As far as I know, only the genuine modal realism of Lewis (1986) claimed such a result and even this has been challenged by Divers and Melia (2002). The second is to concede that there is no objection to be levelled against impossible worlds based solely on their ontology. Unless you are a Lewisian modal realist, you cannot think that the ontological status of impossible worlds is worse than the ontological status of possible worlds.

2.1.3 An Ontology: Linguistic Ersatzism

It is tempting at this point, after the argument against possible worlds theorists of the last section has been made, to leave the ontology of impossible worlds to one side. Alas, I think this is a mistake. A lot of philosophers end up thinking of impossible worlds as sets of sentences of the language that is being given a semantics. As we will see, this motivates some objections that are not so convincing when an explicit ontology for impossible worlds is in place.

A very popular ontological story amongst impossible worlds theorists is that based on what Lewis (1986: 3.3) called *linguistic ersatzism*. Jago (2014; 2015) and Kment (2014) are very weighty examples. I too will adopt a version of this story as my official ontology of impossible worlds, drawing on their work.

According to linguistic ersatzism, worlds, possible and impossible alike, are indeed sets of sentences of a language. But that language is not English or any other natural language, but of a different sort. What follows is a sketch of the strategy which highlights some of the points that will be relevant to us later on. I cannot discuss every interesting issue that arises in this context and I defer to the more detailed treatment of Jago and Kment at various points.

There are at least three reasons why English is not a good choice to serve as a basis for linguistic ersatzism. First, we do not have a complete method to say what a given declarative sentence in English represents. Second, English is a vague and ambiguous language. Third, there are more possibilities than sets of sentences of English. Let me expand on the third reason, since it is decisive enough to disqualify English for this particular theoretical role and it makes for a smooth transition to the discussion of the worldmaking language that we will end up using.¹¹

English is a finite language in the following sense: its vocabulary is finite

¹¹This objection can be found in Lewis (1973: 90).

and so is the length of its sentences. There are at most continuum many sets of sentences of such a finite language. But there is an argument that there are at least as many possibilities as the size of the powerset of the continuum. It is possible for there to exist a continuum of space-time points (whether or not this is actually the case). Each of these points can be occupied with matter or not. To each pattern of occupation of space-time points by matter corresponds a distinct possibility. There are as many of these patterns as the size of the powerset of the continuum. Therefore, possible worlds conceived as sets of English sentences must conflate distinct possibilities: this ontological story is not sufficient.

We must move to an infinite language instead. Here I will consider what Lewis (1986) called a *lagadonian* language. The alphabet of this language is constituted by *actually existing entities*. You and I will end up being symbols of this alphabet. The idea of using objects such as planets and animals as symbols in an alphabet of a language might be weird at first, but it is a natural extension of our usual idea of a language - letter tokens are objects too. And this extension is necessary to make the strategy of building worlds out of sets of sentences work.

The ontological commitments such a language requires are the following: (i) All actually existing concrete particulars; (ii) All actually existing properties and relations;¹² (iii) All sets and numbers. Insofar as you have no objection to the ontology codified in (i)-(iii) you think that the alphabet of this language exists. The alphabet is clearly infinite as there are infinite sets. Moreover, we will allow infinite strings of objects to count as sentences.

The atomic sentences of the language are just strings of n -ary predicates followed by n terms. The logical vocabulary will be assumed to be that of first-order logic, with truth functional connectives, quantifiers and an identity predicate.

But which sequences of symbols stand for terms, predicates and the logical vocabulary?

An n -ary predicate will just be the corresponding n -ary property.¹³ The logical vocabulary can be made to correspond to particular pure sets. The difficult case is the one of names.

There will be two kinds of terms: proper names and referential non-attributive descriptions.¹⁴ Each actually existing object will be its own proper name. My desk is a rigid designator that refers to itself at any world where it exists and so are you and so am I. This is the way in which we refer to actually existing objects in the Lagadonian language.

This still leaves out the objects that do not actually exist.¹⁵ To refer to these,

¹²We are thinking here of properties in an ontologically serious sense in which being committed to them represents a genuine ontological commitment.

¹³I am here assuming a particularly rich conception of which properties there are which would include non-natural properties and maybe even uninstantiated properties. I do not have the space to go into this here. See Jago (2014: 5.4) for discussion.

¹⁴I am following Jago (2014: 152) here.

¹⁵If there are such things. Some philosophers, such as Williamson (2013), are *necessitists*.

we will have to revert to some trickery. Donellan (1966) defended the idea that some definite descriptions could be used to refer to an object in the way that proper names do. Imagine you are at a party and see a man holding what seems to be a glass of champagne, both to you and to your interlocutor. The man’s name is John. You say “The man holding the glass of champagne is exquisitely dressed”. Now, unbeknownst to you and your interlocutor, the glass did not contain champagne but some other fizzy beverage. Donellan made two claims about this use of the description. First, what you said was referential and not quantificational. What you said is closer to “John is exquisitely dressed” than to “The one and only man that is such that he is holding a glass of champagne is exquisitely dressed”. Secondly, you referred to John successfully even in the case where he did not have the property your description attributed to him. Hence, the description you used can be said to be non-attributive. Putting these together, such a description can be said to be referential and non-attributive.

Whether the semantics of definite descriptions in natural language allows for this phenomenon or whether it should be treated pragmatically is irrelevant for our purposes. We can just introduce an artificial operator with this semantics. Kaplan (1978) did just that and called it the DTHAT operator. DTHAT(The F) picks out the unique F but does not attribute *Fness* to it. That allows it to designate whatever it picked at every world, just like a rigid designator would.

A non-actual particular is to be named using the DTHAT device. Given a set of properties $\{F_1, \dots, F_n\}$, DTHAT(The F_1, \dots, F_n) can be a name for a particular. For example, suppose I want to name my non-existing sister. I can use the operator DTHAT(Being my sister) as a name for her and use that to represent my sister as being such-and-so at a world. Because we are naming a non-existing individual, the DTHAT operation in cause is empty, it refers to nothing. But that does not detract from its representational *potential*: it is a meaningful expression just like any other and it can be used to build meaningful sentences.

In my official story, actual individuals and non-actual individuals get named in very different ways. Actual individuals are named by themselves. Non-actual individuals are named by the DTHAT construction. This is optional. Jago (2014: Section 5.5) uses the DTHAT construction to name every individual, even the actual ones. I do not have any objection to that approach other than the fact that it complicates the semantics of the lagadonian language somewhat.¹⁶ The reason Jago chooses to do so has to do with an expressive limitation of the lagadonian language I will be using.¹⁷ Given the impossible worlds I will be using later on, I do not need this extra expressive power.¹⁸

The limitation is this. A natural language sentence like “Hesperus is identical

They believe that, necessarily, everything exists necessarily (where “exists” is understood as “being identical with something”). Being a necessitist would greatly streamline the following discussion.

¹⁶As Jago (2014: 159) admits. You can find more discussion of this phenomena there.

¹⁷This consequence is also is pointed out by Kment (2014: 78).

¹⁸If his ontology is satisfactory, I could use it and add a restriction to the worlds I will use in the semantics of counterfactuals, a procedure I will justify below.

to Phosphorus” is translated to the lagadonian language as $\lceil a = a \rceil$, where a is the object Venus and $=$ is whatever object is playing the role of the identity sign. There are impossible worlds where a is not identical with itself, those that contain the sentence $\lceil \neg(a = a) \rceil$. But we have no way to represent that Hesperus is not Phosphorus. The reason is that, by design, each actual object has exactly one proper name in this semantics, namely itself. If you think that impossible worlds should represent every impossibility, then this is a problem.¹⁹

Worlds are to be identified with sets of sentences of this worldmaking language. We say that a proposition is true at a world if and only if that world contains a sentence of the worldmaking language that expresses that proposition.

I hope to have provided a necessarily sketchy explanation of what a world is, given linguistic ersatzism. It is clear that some of these worlds will be impossible: just consider the set of sentences that has two elements, a sentence and its negation. It is a world but one who represents something impossible. To get a proper ontology for possible worlds you would have to impose a modal requirement by brute force: that any world can represent that p only if p is possible and that the set of sentences that are worlds are closed under logical consequence.

I hope it is clear enough that the language we are using to build the impossible worlds is nothing like English. In the next section we will evaluate two objections that seem completely decisive against the use of impossible worlds in semantics if worlds are sets of the natural language under study. It will be clear that the arguments are not nearly as decisive with an actual ontology of impossible worlds at hand. That being said, actually solving the problem will involve arguing for a certain methodological picture of how we can use impossible worlds in practice.

2.2 Using Impossible Worlds in Semantics

2.2.1 Two Objections: Compositionality and Synonymy

Even if we have succeeded in providing a satisfactory ontological story for impossible worlds, this does not mean that using them is justified. Impossible worlds might just be *bad* circumstances. Presumably, a set of circumstances can be good for some purposes and bad for some others. In this section I will consider two objections that wish to conclude that impossible worlds should not be used in a formal semantics which wants to model phenomena in natural language semantics.

The first is the *compositionality objection*.²⁰ A satisfactory semantics for a

¹⁹In the case of possible worlds there was a clear minimum amount of discrimination that was required: no possibilities could be collapsed. Nothing as clear as this exists in the case of impossible worlds unless, maybe, by stipulation. For what is worth, I do not think that representing all impossibilities is an intrinsically worthwhile goal for an ontology of impossible worlds.

²⁰Pressed, for example, by Fine (2018: 2).

fragment of natural language should be compositional. Roughly, a semantics is compositional if the meaning of a complex expression can be derived from the meaning of its constituent expressions and the way they are put together, its syntactic construction. This is not assumed to be just a brute force requirement. Compositionality helps to explain how speakers can understand and produce a potentially infinite range of sentences with different meanings given their limited memory and cognitive resources.²¹ It seems that impossible worlds cannot be used when giving a compositional semantics.

Start by assuming that impossible worlds are sets of English sentences. Now assume you want to give a compositional semantics for the English sentence represented by $\ulcorner a \oplus b \urcorner$, where \oplus is some binary operator on sentences and a and b are sentences. It seems that the following two objects are both worlds (impossible ones, since they are non-maximal) : $w_1 = \{a, b, a \oplus b\}$ and $w_2 = \{a, b\}$. Both worlds are the same when it comes to the truth-value of the constituent expressions a and b . But they differ in the truth-value of $a \oplus b$. This is a failure of compositionality.

But, as we have seen, impossible worlds are not sets of sentences of English. We are not using sets of sentences of a language to provide a semantics for that very same language. We are using sets of a very different language to provide a semantics for natural language. Does the argument work when we keep this in mind?

It does given an extra assumption I will call *syntax mirroring*: the syntactically simple expressions of the natural language are translated to syntactically simple expressions of the world making language and any step of syntactic construction in the natural language is mirrored by a step of syntactic construction in the worldmaking language.

Given that worlds are sets of sentences, what worlds there are depends on the syntax of the worldmaking language. If the language being given a semantics and the worldmaking language are similar enough syntactically, then the compositionality objection works.

The second objection is quite similar. Let us call it the *synonymy* objection.²² Sometimes we want two different expressions to be synonymous in a semantics. Synonymy, we will assume, amounts to the fact that a substitution of an expression by a synonymous expression preserves semantic value. It seems that if we use impossible worlds this cannot be obtained.

Assume again that worlds are sets of English sentences. Let a and b be two distinct expressions, let φ be a sentence with one or more occurrences of a and let $\varphi[a/b]$ be the result of substituting one or more occurrences of a by an occurrence of b . It seems that the following are two different worlds: $w_3 = \{\varphi\}$ and $w_4 = \{\varphi[a/b]\}$. Since each sentence is true at the world where the other is false, we can conclude that a and b are not synonymous. Since a and b can be any expressions whatsoever, it seems that there can be no synonymous expressions at all.

²¹See Dever (2012), for example, for a survey of the philosophical issues surrounding compositionality.

²²It is put forward, for example, in Williamson (2017: 212).

Again, the fact that worlds are not sentences of English can be remedied by the assumption of syntactic mirroring. The fact that different English expressions correspond to different worldmaking sentences is sufficient for the argument to go through.

The best way to resist these arguments in their present form is to reject syntax mirroring. In fact, the details I gave about the worldmaking language above are sufficient to block the synonymy objection. “Hesperus” and “Phosphorus” correspond to the same worldmaking expression *contra* syntax mirroring. And the worlds $w_5 = \{\varphi\}$ and $w_6 = \{\varphi[a/a]\}$ are the *same* world for every expression a . So, “Hesperus” and “Phosphorus” can be synonymous.

I think this is already a significant clarification in this dialectic. Unfortunately, the original arguments were too ambitious. They would prove widespread violations of compositionality and the triviality of synonymy. It can be enough for there to be unwanted violations of compositionality or unwanted breaks of synonymy for the conclusion that impossible worlds should not be used in semantic theorising. A framework that forces on us unwanted requirements is not a good framework.

Responding to this weakened form of this arguments is significantly harder. There might be a way to meet it head on, to show how using impossible worlds can be fully compositional and capable of respecting the synonymy relations we want. I do not know of such a direct way. But I think that getting clear on how impossible worlds can be used in practice can afford a way out for the impossible worlds theorist.

2.2.2 A Pragmatic View of Impossible Worlds

There is an hidden assumption in the objections above. It is assumed that if there *exist* two impossible worlds which together represent a counterexample to a desirable feature of a formal semantics then impossible worlds cannot be used to provide such a formal semantics. I think this is unwarranted.

The assumption can be stated as follows:

Ontological Determination: If a formal semantics uses some objects that satisfy an ontological story S , then it uses all of the objects that satisfy that ontological story.²³

Why would it be the case that using one impossible world commits us to using all of them? Impossible worlds are used to interpret formal semantics. For this purpose it is important that they exist and that the kind of things they are can stand in appropriate relations to the interpretations of other elements in the semantics. These are the kinds of properties that are closed under the subset relation: a proper subset of impossible worlds also exists and can also stand in the appropriate relations.

²³For a remark by Fine that can be read as an implicit endorsement of this thesis, see Fine (2018:2): “How then is the range of the impossible worlds to be restricted? Which of the putatively impossible worlds is a *genuine* impossible world?”. The relevant reading is one where “genuine” is meant in some ontological sense.

I think that the determination of which circumstances are used in a formal semantics should not be exhausted by the ontological story we tell about them. I defend the following:

Pragmatic Determination: The objects a formal semantics uses is determined by the special characteristics of the problem to be solved as long as there is an ontological story S that guarantees that (i) These objects exist; (ii) They can stand in the appropriate relations to the interpretations of other elements in the semantics.

The picture is this. A satisfactory ontological story characterises a certain class of objects that meets (i) and (ii). This is all that should be demanded from the ontology. After this, it detracts nothing from the ontological status of the circumstances used that we impose a *restriction* on them. We use only those objects that satisfy the restriction and not those that do not.

Obviously, this is not to say that any restriction is acceptable. The cost that is associated with a restriction is not ontological but explanatory. If we use *all* of the objects characterized by a certain ontological story we do not need to explain why we want to impose a restriction. If the kind of things the objects are imposes certain formal requirements in our semantics, and if these requirements are desirable, then the ontology itself is doing explanatory work. Those formal requirements might be considered *ad hoc* if not for the fact that they are guaranteed by the ontology. The important claim is that this is *extra* work done by the ontology, not *necessary* work.

We can thus use certain restrictions on the space of impossible worlds when we want to present a formal semantics. By imposing such a restriction we are forced to justify it somehow: the justification cannot be ontological. The heart of the idea of pragmatic determination is that the specifics of the problem to be solved by the formal semantics can give us enough materials to do this.

A lot of restrictions are possible, some better than others. Some restrictions will seem *ad hoc*. Some other will use an inappropriate language making the semantics circular and non-explanatory. Such problems must be addressed in a case by case basis.

We can now see the conceptual space for a solution to the compositionality and synonymy objections. These objections are domain-relative: they are arguments against the use of impossible worlds in *semantics*. As such, they are motivated by some problem specific considerations. There are reasons why we want a semantics for natural language to be compositional. Maybe a formal semantics for other problems should not be compositional after all. Call the reasons used to defend the thesis that a semantics for a natural language must be compositional C . I claim that the impossible worlds theorist can restrict the worlds she uses to the ones that are compositional, in the sense that they include sentences that, when translated to the natural language in question, do not represent failures of compositionality. When asked for the motivation for this restriction, she can just reply: “Because of C ”.²⁴

²⁴Of course, various restrictions will be compositional. Deciding which one to use requires

You might think that the possible worlds theorist has an advantage here. The ontological explanation is somehow better than the problem-relative one. It is possible that this is just a cost that the impossible worlds theorist has to live with. But I doubt even that. It seems to me that the possible worlds theorist is building the restriction into the ontological story. Unless you are a Lewisian modal realist, there is a specification in your ontological story where you explicitly eliminate the impossible worlds. It seems reasonable to me that this hampers the explanatory advantage of the ontological justification. The general rule seems to be that if your ontological story has a requirement that stipulates directly what the circumstances can make true, then no explanatory advantage can be gained in relation to the ontological story without the requirement but after restriction.

The idea is that, in a very real sense, it is more principled to say that what the possible worlds theorist is doing is presenting an ontological story that applies to impossible worlds and then restricting them away. Since the ontological justification seems definitely circular, the justification the possible worlds theorist must use will turn out to be as pragmatically motivated as the one used by the impossible worlds theorist.

In the next section I will present the particular restrictions on the space of impossible worlds I propose to use in the case of counterfactuals and sketch the main issues it raises.

2.3 A Restriction for the Semantics of Counterfactuals

I propose the following restrictions to the space of impossible worlds that is to serve as the circumstances in the semantics of counterfactuals. First, all impossible worlds will be logically possible. Secondly, at all impossible worlds, if a and b are rigid designators of the same object in English, then S and $S[a/b]$ are translated as the same worldmaking sentence. That is, substitutivity of identicals is preserved at every impossible world.

The same space of impossible worlds is used by Kment (2014). One difference is that he treats both of my restrictions at the level of the ontology of impossible worlds. The worlds are stipulated to be logically consistent - as I have indicated above I think it is in the interest of transparency to treat this as a restriction rather than part of the ontological story. And his ontology of worlds uses the simple lagadonian strategy of making each object stand for its own name, thus guaranteeing the substitutivity of identicals. Though I have also included this feature as part of my ontological story I include it here as a restriction for clarity.

The second big difference is that Kment does not provide any extensive discussion of the problems such a strategy faces in the form of “revenge” forms of vacuism. When we restrict the space of impossible worlds we are using, some

extra argument - argument that must fall under the domain of first order semantics. That is irrelevant to the main point I am making.

propositions might be true at no impossible world. A counterfactual with such a proposition as an antecedent is predicted to be vacuously true. Moreover, since his book predates Williamson's (2017) influential critique of non-vacuism, he does not explain how his theory can respond to Williamson's objections. In this thesis, I hope to contribute towards solving both of those deficiencies.

The main motivation for adopting these restrictions has to do with the problems non-vacuism faces. Let me explain how these problems lay the ground for the chapters that follow.

In the chapter on Counterfactual Logic we will see that non-vacuists can be accused of providing a logic that is too weak. Allowing that there are points of evaluation in our models that make arbitrary sets of propositions true, without any regard to logical relations between propositions, threatens to invalidate a lot of logical principles involving counterfactuals that seem valid. I will argue that maintaining the requirement that impossible worlds are logically possible solves this problem. The only difference between the resulting counterfactual logic and the standard vacuist counterfactual logic is that some worlds can represent metaphysical, though not logical, impossibilities. At the level of a propositional modal language, the differences are so minimal that the resulting logic can be considered robust enough.

The problem with this approach, when compared with a non-vacuist semantics that uses a plenitude of impossible worlds, is that there is a form of vacuism lingering. Each restriction of the space of impossible worlds makes it the case that some propositions will be true at no world in the semantics. And this is the source of vacuism. With respect to the restriction to logically possible worlds, the problem of *counterlogical* vacuism arises. If an antecedent is logically impossible, then the whole counterfactual is vacuously true. I will argue that the cost of allowing counterfactual vacuism is much smaller than allowing full counterpossible vacuism. Though there is no precise way of proceeding with the cost-benefit analysis, my impression is that solving the problem of counterfactual logic is reward enough for having to live with counterlogical vacuism. In any case, the very conceptual possibility of this solution shows that having a weak counterfactual logic is not a necessary part of non-vacuism.

One upshot of this discussion is that, in some sense, logic has a stonger influence in the semantics of counterfactuals than metaphysics does. Though it is hard to make precise sense of this idea, I will also argue that this situation makes good methodological sense.

An important point is that the connection between counterfactuals and metaphysical modality runs both ways. Though this might not be a great cost for the semantics of counterfactuals, it could be a significant cost for our understanding of metaphysical modality. If this were so, a good reason for keeping the standard vacuist semantics would be that it provides a good foundation on which to rest our understanding of metaphysical modality. In the chapter on counterfactuals and metaphysical modality I discuss two services vacuism can be thought to provide for our theory of metaphysical modality: that of providing a modal epistemology and that of explaining what metaphysical modality is in the first place. I will argue that the epistemological cost is minimal: the

non-vaculist can provide a modal epistemology that is reasonably successful provided that the vacuist modal epistemology is successful as well. The real cost comes in understanding what metaphysical modality is. But, I will argue, it is hard to say how big of a cost that is.

Finally, the chapter on Counterfactuals and Worldly Opacity discusses the second part of my restriction. Williamson (2017: Section 4) argues that counterfactuals should be transparent, that they should respect the substitutivity of identicals. He thinks that this must be the case because, if they were not, there would be a phenomenon of worldly opacity. And this phenomenon is on the border of being unintelligible.

I argue once again that the problem is serious and that the non-vaculist responses have been inadequate. I take it that this is enough to justify my second restriction that substitutivity of identicals must be possible at any impossible world. As before, this gives rise to a new form of vacuism. If an antecedent supposes that two identicals are distinct it cannot be true at any impossible world we are using and, therefore, the whole counterfactual is vacuously true. I argue that the costs of this vacuism are even easier to live with than the case of counterlogical vacuism. In any case, the very conceptual possibility of this solution shows that treating counterfactuals as violating the substitutivity of identicals is not a necessary part of non-vacuism.

Chapter 3

Counterfactual Logic

In this chapter we will consider the interplay between the logic governing counterfactual conditionals and the debate between vacuists and non-vacuists. Section 1 presents the standard counterfactual logic and shows why it is committed to vacuism. We then see which changes we should make to the semantics in order to get the non-vaculist semantics of Berto et al. (2017). Section 2 puts forward the *weak logic argument* against non-vacuism - non-vacuists cannot develop a robust logic for counterfactuals. We then evaluate the standard non-vaculist response to this problem and argue that it is unsatisfactory. Section 3 proposes a different solution, a non-vaculist logic that uses only logically possible worlds. I argue that it allows us to solve the problem of weak logic in a non-vaculist setting. The problem with this semantics is that it predicts that all counterlogicals are vacuously true. I then put forward different strategies that might allow us to live with counterlogical vacuism. I close with some methodological considerations based on this semantics.

3.1 Counterfactual Logic

3.1.1 What is a Counterfactual Logic?

I have been speaking of a semantic analysis of counterfactuals as a theory about what truth conditions these sentences have. This can make it seem that the job of a semantics for a particular class of expressions is finished when such a theory is provided. This would be wrong. Apart from the truth value intuitions speakers have about particular sentences, they also have intuitions about the *relations* between truth values of sets of sentences. One particularly relevant class of such intuitions are *validity intuitions*. When we have the intuition that a sentence, call it ψ , cannot be false when a set of sentences $\{\varphi_1, \dots, \varphi_n\}$ are true, we can be said to have the intuition that the inference from $\varphi_1, \dots, \varphi_n$ to ψ (or the argument which has $\varphi_1, \dots, \varphi_n$ as premises and ψ as a conclusion) is valid. A theoretical story about which inferences are valid on a certain language

that contains a target class of expressions counts as providing a *logic* for these expressions.¹

Normally, a counterfactual logic is developed by extending a propositional modal language with a connective representing the counterfactual conditional. This gives us the opportunity not only of characterizing which inferences involving counterfactuals are valid but also of studying the interactions of the counterfactual conditional with other logical expressions.

3.1.2 Standard Counterfactual Logic

Let us start, then, by presenting the counterfactual logic that emerges from the classic Lewis-Stalnaker analysis.²

We start with a propositional modal language. We have the truth-functional connectives \wedge , \vee , \neg and \rightarrow and the modal operators \diamond and \square . The notion of well formed formula is defined in the usual way. A model for this propositional modal language is just an ordered pair $M = \langle W, V \rangle$, where W is the set of (metaphysically) possible worlds and V is a valuation function that assigns the value 0 or 1 to each propositional atom at every possible world.

The truth conditions for sentences of this language at a world w of a given model M can be defined recursively as follows:

- $M, w \Vdash p$ iff $V_w(p) = 1$
- $M, w \Vdash \neg\varphi$ iff it is not the case that $M, w \Vdash \varphi$
- $M, w \Vdash \varphi \wedge \psi$ iff $M, w \Vdash \varphi$ and $M, w \Vdash \psi$
- $M, w \Vdash \varphi \vee \psi$ iff $M, w \Vdash \varphi$ or $M, w \Vdash \psi$
- $M, w \Vdash \varphi \rightarrow \psi$ iff $M, w \Vdash \neg\varphi$ or $M, w \Vdash \psi$
- $M, w \Vdash \diamond\varphi$ iff $\exists w' \in W, M, w' \Vdash \varphi$
- $M, w \Vdash \square\varphi$ iff $\forall w' \in W, M, w' \Vdash \varphi$

We want to extend this language to include a connective, \Rightarrow , representing the counterfactual conditional. The models for this enriched language are now triples $M = \langle W, \{R_A : A \in \Phi\}, V \rangle$. Φ is the set of well formed formulas of the language. For every formula φ , R_φ is a binary relation on W . $wR_\varphi w'$ is to be interpreted as “World w' is one of the closest worlds to w where φ is true”, where closeness is interpreted in terms of similarity, as explained in the first chapter.³

¹See Winter (2016: Ch.2) for an introduction to how semantics can deal with the notion of entailment.

²The semantics I present is in the style of Priest (2008: Ch. 5) and Berto et al. (2017).

³Note that, in the way I am presenting the model theory of counterfactual logic, there is no explicit relation of closeness in the model. The notion of similarity only enters into the interpretation of the relations R_φ . If one wished, the models could contain a set of contexts and a function from this contexts to closeness orderings, thus making such orderings explicit. The way I present the semantics has the advantage of making my discussion simpler. I will not be making any point whose logical sophistication would require the more complex semantics.

Given this interpretation, we must have the following:

(Adequacy) If $wR_\varphi w'$, then $M, w' \Vdash \varphi$.

This condition simply says that if a world w' is one of the closest worlds to w where φ is true, then φ must be true at it.

We now add a truth-condition for the counterfactual conditional:

- $M, w \Vdash \varphi \Rightarrow \psi$ iff for all w' such that $wR_\varphi w'$, $M, w' \Vdash \psi$

Finally, logical validity is just defined as truth-preservation at all worlds of all models:

- $\Sigma \models \varphi$ iff for all M and all w such that $M, w \Vdash \psi$, for every $\psi \in \Sigma$, $M, w \Vdash \varphi$.⁴

We have said enough to prove vacuism:

(Vacuism) If $\models \Box(\neg\varphi)$, then $\models \varphi \Rightarrow \psi$, for any ψ

Proof: $\Box(\neg\varphi)$ is true iff φ is false at every world. This means, given (Adequacy), that the relation $wR_\varphi w'$ is empty. This implies that the universal quantification “for all w' such that $wR_\varphi w'$ ” is vacuously true.

Note one thing about the proof above. The important part when proving vacuism is that the antecedent was not true at any point in the model. It just so happens that, in a context where the points are interpreted as possible worlds, a proposition’s being impossible is sufficient for this to be the case. It is the semantic machinery of the worlds analysis of counterfactuals, with its universal quantification over the closest worlds at the level of the metalanguage, that creates the conditions for vacuism. In a while we will consider expansions to the space of worlds given such that a proposition’s being impossible will no longer be sufficient to make it untrue at every world. But the phenomenon of vacuism remains as long as there are some propositions that are false at every point of every model.

When it comes to the interaction of counterfactuals with the background logic ruling the propositional modal language being extended, the following is the central result, and one that will play a central role in the discussion of this chapter:

(Closure) If $\models \psi_1 \wedge \dots \wedge \psi_n \rightarrow \chi$, then $\models (\varphi \Rightarrow \psi_1) \wedge \dots \wedge (\varphi \Rightarrow \psi_n) \rightarrow (\varphi \Rightarrow \chi)$

Proof: If $\models \psi_1 \wedge \dots \wedge \psi_n \rightarrow \chi$, then the material conditional $\psi_1 \wedge \dots \wedge \psi_n \rightarrow \chi$ is true at every world of every model, which means that at every world where $\psi_1 \wedge \dots \wedge \psi_n$ is true, χ is also true. If the conjunction of counterfactuals $(\varphi \Rightarrow \psi_1) \wedge \dots \wedge (\varphi \Rightarrow \psi_n)$ is true

⁴I will follow the convention that dropping one of the parameters to the left of the turnstile is interpreted as an universal quantification over that parameter.

at a world w , then all worlds w' such that $wR_\varphi w'$ are such that $\psi_1 \wedge \dots \wedge \psi_n$ is true, and, therefore, so is χ . Since χ is true at all worlds w' such that $wR_\varphi w'$, the counterfactual $(\varphi \Rightarrow \chi)$ is true.

This is a very important result. Its practical importance is the following. Assume that you are developing a counterfactual supposition A . You want to check which propositions would be true if A was true. It seems that you can use logical deduction to help you do this. If some propositions are true under the counterfactual supposition that A and if C is a logical consequence of these propositions, then C is also true under the counterfactual supposition that A . As a special case, if you have a logical deduction of C from the assumption that A , you also know that the counterfactual $A \Rightarrow C$ is true. The validity of (Closure) implies both that logic can be used to as a method to expand the set of propositions true under a counterfactual supposition and that logical deductions always correspond to true counterfactuals.

Finally, some more validities can be guaranteed by imposing further principles on the acceptable similarity metrics. The strategy is to argue that the fact that we have an interpretation for the relation R_φ is sufficient to constrain the intended models for the counterfactual logic. For example, the following principle seems reasonable:

(Weak Centering) If $w \Vdash \varphi$, then $wR_\varphi w$

What (Weak Centering) says is that if the world of evaluation of a counterfactual makes the antecedent true, then that world is among the closest worlds where the antecedent is true. In terms of similarity, any given world is at least as similar to itself (though there can be ties).⁵ This seems like a reasonable restriction given our intuitive grasp on the notion of similarity. If it is, we have less acceptable models for our counterfactual logic. And with less models there often come more validities. In this case, the following comes out valid:

(Modus Ponens for Counterfactuals) $\models (\varphi \Rightarrow \psi) \rightarrow (\varphi \rightarrow \psi)$

Proof: Assume that $(\varphi \rightarrow \psi)$ is false at a world w , that is, that φ is true and ψ is false. Since φ is true at w , this implies that $wR_\varphi w$, via (Weak Centering). But we assumed that ψ is false at w . So, the counterfactual $(\varphi \Rightarrow \psi)$ is false. This shows that $(\varphi \Rightarrow \psi) \rightarrow (\varphi \rightarrow \psi)$ is true at all worlds of all models.

We can distinguish between two types of validities: those that are valid given only the formal structure of the Lewis-Stalnaker semantics and those that are valid when the interpretation of the elements in the semantics comes into play.

⁵The reason that there can be ties is that not all aspects of similarity matter in every context. A red ball is as similar to itself as a red circle if the only relevant aspect of similarity is redness. If you think that ties cannot happen, then you think that when the world of evaluation makes the antecedent true then it is the *unique* closest world where the antecedent is true, what is normally called strong centering. Of course, strong centering implies weak centering.

This distinction will matter when we discuss the merits of the SIC-strategy in section 2.3.

This concludes my brief presentation of the standard counterfactual logic. We can now see which changes Berto et al. propose to prevent the validity of vacuism.

3.1.3 A Non-Vacuist Semantics

In this section I will present the non-vacuist semantics of Berto et al. (2017). Note that, strictly speaking, avoiding vacuism with regards to metaphysically impossible antecedents requires only the addition of one impossible world, a point where an impossible proposition is true. This minimal solution would not be satisfactory, of course. It would be wildly arbitrary which impossible proposition is allowed to be true at a world. But, as we will see, there is a lot of space in between this minimal solution and the one adopted by Berto et al.

The first modification is to the space of worlds used. Models for the language will now be 4-tuples $M = \langle W, P, \{R_A : A \in \Phi\}, V \rangle$. $P \subseteq W$ is the set of possible worlds, which makes $I = W - P$ the set of impossible worlds. The truth conditions are now stated differently for possible and impossible worlds. At possible worlds, the valuation function only assigns truth-values to the propositional atoms, the valuation being extended to the whole language via the recursive clauses above. The only difference is that the quantifiers in the clauses for the modal operators need to be restricted to possible worlds.

- $M, w \Vdash \Diamond\varphi$ iff $\exists w' \in P, M, w' \Vdash \varphi$
- $M, w \Vdash \Box\varphi$ iff $\forall w' \in P, M, w' \Vdash \varphi$

At impossible worlds, however, the valuation function assigns a truth-value directly to every formula, independently of its syntactic complexity:

- If $w \in I$, $w \Vdash \varphi$ iff $V_w(\varphi) = 1$

Berto et al. assume a plenitude of impossible worlds in the following sense. For any distinct formulas φ and ψ , there is an impossible world w such that $w \Vdash \varphi$ but $w \not\Vdash \psi$.⁶

The evaluation of counterfactuals at impossible worlds works in the same way as for any other formula: they get assigned a truth-value directly by the valuation function. The evaluation of counterfactuals at possible worlds remains the same with the following difference: impossible worlds are allowed to be among the closest worlds to a given possible world. In fact, this is the only way in which impossible worlds change the truth value of formulas at possible worlds.

⁶As we saw in chapter 2, whether or not there are that many impossible worlds when we take the interpretation of the language seriously is another matter. Depending on the expressive power of the object language, this can be quite hard to achieve. Let us bracket this problem in the context of this chapter.

This is enough to make falsify vacuism. A proposition φ can be impossible, by being false at every world in P , and still be true at an impossible world w_i . This means that a counterfactual $\varphi \Rightarrow \psi$ can be false if $w_i \not\models \psi$ and w_i is the unique closest world relative to R_φ .

In fact, give the plenitude of impossible worlds, there is no form of vacuism left. Every formula is true at some world. For every φ and w , there is at least a world w' such that $wR_\varphi w'$. The universal quantifications on the truth-conditions of counterfactuals are never empty. Therefore, they are never vacuously true.

The notion of logical validity also stands in need of revision. It will now correspond to truth-preservation at all *possible* worlds at all models.

- $\Sigma \models \varphi$ iff for all M and all $w \in P$ such that $M, w \Vdash \psi$, for every $\psi \in \Sigma$, $M, w \Vdash \varphi$

This is our first explicit non-vaculist semantics.

With such an explosion of acceptable models of our language, we stand to lose some validities. And this is indeed the case. (Closure) is no longer valid:

Counterexample to (Closure): In this semantics, due to the fact that logical validity only takes notice of possible worlds, we have that $\models (p \wedge \neg p) \rightarrow q$ for any propositions p and q . But the counterfactual $(p \wedge \neg p) \Rightarrow q$ can be false at a world w . It is enough that all impossible worlds w' such that $wR_{(p \wedge \neg p)} w'$ are such that q is not true at them. Since there is no reason for $R_{(p \wedge \neg p)}$ to be empty, since contradictions are true at impossible worlds, there are models where all the closest impossible worlds make q false.

The background logic is dissociated from the counterfactual conditional. The presence of impossible worlds makes no difference to the fragment of the language without the counterfactual conditional due to the disjunctive nature of the definition of logical validity - it cares only with what is true at possible worlds. But counterfactuals are influenced by the impossible worlds. Having a plenitude of impossible worlds forces us to have some logically anarchic impossible worlds. At those worlds, a formula and its logical consequences can come apart. Therefore, the logical connection between two formulas guarantees nothing regarding the truth of counterfactuals involving them.

In the next section we will see how the vacuist can turn this fact into an argument against the non-vacuist. We will then consider the response Berto et al. give to that argument.

3.2 Weak Logic

3.2.1 The Weak Logic Argument

Adding impossible worlds, then, has a real impact on the non-vacuist's counterfactual logic. Generally speaking, since impossible worlds can be among the

closest worlds and any connection between two formulas can be severed at an impossible world, the logic that emerges is very weak.

This can be turned into an argument against the non-vacuumist. Williamson puts the problem as follows:

We may also wonder what logic of counterfactuals [non-vacuumists] envisage. If they reject elementary principles of the pure logic of counterfactual conditionals, that is an unattractive feature of their position. (Williamson 2007: 174)

Call this the *weak logic argument*. It is based on a perceived lack of adequacy or mismatch between the non-vacuumist counterfactual logic, on one hand, and (1) Ordinary speaker's validity intuitions; and (2) Ordinary speaker's deductive behavior, in the other.

I gave an intuitive gloss of what I mean by "validity intuition" above. Sometimes we have the intuition that a sentence simply *follows* from a given set of sentences. We can have such an intuition even if we have no particular intuition about the truth value of the sentences involved. An intuition of validity is an intuition about the *relation* between the truth-value of sentences, not about the actual truth-value of the sentences themselves.

By the expression "deductive behavior" I mean something like the following. In ordinary discourse we can take certain attitudes regarding particular assertions. Roughly speaking, we can assent to them, we can deny them or we can suspend our judgement on them.⁷ It seems clear that there are certain normative principles that rule the way in which we can assent, deny or suspend judgement. A relatively plausible one is the one that says that we cannot assent to a proposition p while denying a logical consequence of p . Insofar as this is a good principle, rationality would demand that we obey it. If it is pointed out to us that we are in a position where the collection of our assertions and denials violates it, we must recognise that something needs to be changed (even if we do not know immediately what). This assigns a particular role to logic in governing our linguistic practice. Counterfactuals, of course, are no exception. Now consider the following dialogue:

A: If I had spent more money in my campaign, I would have won the election. The only reason I lost was that my opponent had much more resources than me.

B: But, wait. If you had spent more money, your opponent would also have spent more money, right? After all, he was financially supported by all the oil tycoons! If you think that resources were the key to this election, spending more of your hard-earned money would not have made a difference.

A: I guess you are right... I need to understand what was my mistake, though.

⁷See Ripley (2011) and Friedman (2017) for a discussion of these different attitudes.

Here is a natural way to understand this interaction. A agrees that the following two things are true under the counterfactual supposition of A’s spending more money: a) The candidate with more money would still have won; b) B would still have spent more money than A. The conclusion that B would still have won follows logically from a) and b). So A is forced to agree he needs to retract his initial assertion, it is not the case that if A had spent more money, she would have won the election. It seems that A and B are adhering to (Closure) as stated above. Though it could have occurred to A to refute one of the two counterfactuals B is considering (maybe the important thing for A was to have funding above a certain threshold, therefore making a) false under the counterfactual supposition), it does not seem like rejecting the conclusion that “If A had spend more money in the campaign, B would still have won” *after* accepting a) and b) was an option at all. And the natural way to understand this is that there was a logical inference involved and that logical consequences of what has been assented to must not be denied.

If this is true, if ordinary intuitions regarding what follows from what are violated and if certain normative principles governing the assertions people assent to in ordinary discourse lose their justification, then this is a cost for the non-vacuaist. How can the non-vacuaist respond?

One option would be to say that the cost of violating intuitions of validity is so much smaller than violating intuitions of truth-value that the cost can just be accepted. Validity is, after all, a somewhat theoretical notion, one that needs to be explained patiently to students taking a first-year logic class. This can be seen to erode some of our confidence in validity intuitions. And even if the connection of logical validity and what can be assented and denied is a particularly nice explanation of our linguistic behavior, it might not be the only one.

Given that the issue of the relative weight of different factors in the cost-benefit analysis is a particularly murky one, it is not easy to assess this reaction. My sense is that this can vary with different people’s intuitions. For the purposes of this chapter, the important part about this strategy is that it agrees that there is a cost associated with having a weak logic. In the endgame, this cost should be compared with the cost of whatever alternative strategies there are. Since I have no clear idea on how to do this, I will leave this worry to one side.

The traditional reaction of non-vacuaists, first suggested in Nolan (1997) and adopted by Berto et al. (2017), is different. They want to build a counterfactual logic using impossible worlds that is more robust than what initial appearances suggest. They do this by imposing a condition on admissible similarity metrics. We will present and evaluate their efforts in the next sections.

3.2.2 SIC-Based Strategy

The restriction on admissible similarity metrics non-vacuaists propose is the so-called *Strangeness of Impossibility Condition*:

(SIC) If $x \Vdash \varphi$, for some $x \in P$, then $wR_\varphi w'$ implies that $w' \in P$

This amounts to saying that if a p is possible, then any world among the closest p -worlds is a possible world. In terms of similarity, it says that any possible world is more similar to another possible world than to any impossible world (Nolan 1997: 550).

Before imposing SIC, the only influence impossible worlds had was contained to counterfactuals. After SIC this influence is reduced even further. Impossible worlds only matter when the antecedent of a counterfactual is impossible. We can prove the following:

(P-Validity) Let φ be a formula and let the set $\{A_1, \dots, A_n\}$ contain the antecedents of all counterfactuals occurring in φ . Given (SIC), if $\models_V \varphi$, then $\diamond A_1 \wedge \dots \wedge \diamond A_n \models_{NV} \varphi$ (where \models_V and \models_{NV} are the relations of logical consequence in the vacuist and non-vacuist logics, respectively).

Proof Sketch: Recall that logical validity is truth preservation at all possible worlds at all models. The truth value of the modal operators and truth functional connectives at a possible world is only sensitive to what goes on at possible worlds, leaving the evaluation of counterfactuals as the only area where impossible worlds can make a difference. The result follows by induction on the complexity of φ , noting that even this influence is blocked by the premise that the antecedents are possible together with SIC.

In particular, the following replacement of (Closure) is valid:

(P-Closure) If $\models \psi_1 \wedge \dots \wedge \psi_n \rightarrow \chi$, then $\diamond \varphi \models (\varphi \Rightarrow \psi_1) \wedge \dots \wedge (\varphi \Rightarrow \psi_n) \rightarrow (\varphi \Rightarrow \chi)$

And this allows the non-vacuists to use the counterfactual logic of the vacuist as a springboard for their own counterfactual logic. If some principle is valid in the vacuist's logic it will be valid in the non-vacuist's logic after taking the possibility of the antecedent of every counterfactual involved in the principle as an hidden premise. This makes for an all-purpose reaction to accusations of weak logic.

This strategy is initially appealing. It provides an answer to the challenge at hand and it points to a principled way to construct a counterfactual logic when impossible worlds are around. Here is what Berto et al. have to say about it:

(P-Closure) holds of the logic we have specified, if we assume SIC. And quite generally, with SIC in place, as long as the antecedents of all the conditionals we are dealing with are possible, we can simply ignore the impossible worlds. So the valid inferences of merely-possible-world semantics are recoverable enthymematically by adding suppressed premises of the form $\diamond A$. Adding impossible worlds loses us nothing. (Berto et al. (2017), p.8)

All of this being said, ultimately I do not think this strategy works.

3.2.3 Against SIC

I will present four reasons why the SIC-Based strategy does not work: 1- It makes for an implausibly gerrymandered semantics; 2- The logic is still too weak, by non-vacuaist's lights; 3- A principle like SIC is hard to establish; 4- SIC is intuitively false. Let me present each of these objections in turn.

First, accepting SIC predicts that there is a big difference between the logical behavior of counterfactuals with possible antecedents and counterpossibles. While counterfactuals with possible antecedents obey the exact same logic that the vacuaist takes them to obey, counterpossibles are nearly logically anarchic. As we saw on the introductory chapter, hybrid semantics, in the sense of postulating very different mechanisms of semantic evaluation for counterpossibles, are undesirable. The non-vacuaist must claim that this is not the case here. But this requires some justification. Here is how Williamson puts the point:

[Non-vacuaists] treat it as a virtue of their account that it preserves a strong logic of counterfactuals with possible antecedents. However, from an abductive perspective, that puts increased pressure on their reasons for rejecting the standard rules in the first place [...]. Those reasons had better be robust enough to justify the sacrifice of a strong and simple theory such as Stalnaker or Lewis's logic of counterfactuals, in favour of one patched up with messy adjustments and repairs. In particular, given that counterfactuals have a compositional semantics, we should be suspicious of the idea that their behaviour is radically different on the rare occasions when the antecedent is impossible. For how are we supposed to have come to be using so oddly contoured a conditional? (Williamson 2017: 208)

The justification Williamson is looking for should have the form of an explanation of why the counterfactual supposition of a metaphysical supposition can take us "beyond logical bounds", as Berto et al. (2017: 7) put it.

This brings me to my second objection. Why is the line drawn precisely there? Why is it *metaphysical* impossibility that affects the logical behavior of counterfactuals? I do not see any such strict connection between weakness of logic and metaphysical impossibility as plausible. In fact, it seems to me, the non-vacuaist's counterfactual logic with SIC is still too weak from the non-vacuaist perspective.

Let us focus on (P-Closure). Intuitively, it amounts to saying that you can use logic to figure out what is true under a counterfactual supposition, as long as the supposition is metaphysically possible. This does not cover the deductive behavior of metaphysicians when arguing using counterfactuals. Metaphysicians will sometimes proceed by accepting a metaphysical proposition counterfactually in order to show that it has unpalatable consequences of some kind. In their arguments they can use deductive reasoning based on classical logic. However, if it turns out that the counterfactual supposition is false, it will be impossible. Therefore, (P-Closure) cannot validate these deductions of metaphysicians. This is the kind of weakness of logic that is relevant for Williamson's argument

to go through. Moreover, accounting for the way in which metaphysicians use counterfactuals is one of the initial motivations for non-vacuism.

Third, SIC is a very hard principle to establish. Restrictions on similarity metric come in two varieties. If they are intended to apply to some contexts but not all, they can be said to be *restricted*. If they are intended to apply to all contexts, they can be said to be *unrestricted*. SIC is unrestricted, as it must be in order to be the basis of a counterfactual logic. It amounts to saying that metaphysical possibility is the *maximally important* measure of similarity in *every* context of evaluation. Compare this claim to the principle of (Weak Centering) we considered above. It seems that (Weak Centering) follows from our understanding of the notion of similarity. Whatever other problems using the notion of similarity might bring, we are at least entitled to this formal constraint on our models by using it. In contrast, why would the notion of similarity behave in accordance to SIC? It seems like we would need some convincing to be in a position to accept SIC. And it seems plausible that the only way to establish an unrestricted similarity principle would be to show how it follows from the notion of similarity we are using. The reason better not be because it gives us a robust counterfactual logic, otherwise this seems just like an *ad hoc* move. We need an argument. And no such argument has been given.⁸

Someone can reply to the objection of the last paragraph as follows. The notion of similarity used in the semantics of counterfactuals is not exactly the ordinary notion of similarity. It is partly a theoretical construction designed to produce the right results with regards to the truth-values of particular counterfactuals in context. As such, SIC may yet be true if it is the best systematization of our truth-value intuitions. Though I am unsure about the merits of this argument in the present context, it faces a bigger problem: our intuition tells us that there are some counterexamples to SIC.

A counterfactual is a counterexample to SIC if it has a possible antecedent and an impossible consequent. If such a counterfactual is true, then all the closest worlds where the antecedent is true must be impossible worlds - they make the impossible consequent true, after all. But, since the antecedent was possible, this is a direct violation of SIC. Consider the following example by Daniel Nolan:⁹

Suppose I have been playing a game with a boy called Oliver, where we arrange balls into a square grid, then tip the balls into a bag, then count the balls that come out. [...] While playing it, I introduce Oliver to the idea of a square number, in the obvious way. On a

⁸You might think that Kment (2006; 2014) has offered such an argument. Kment analyses metaphysical modality using the notion of similarity: a proposition is metaphysically possible if, and only if, it is true at all the members of a certain similarity sphere around actuality. It seems that SIC is built into the analysis. This, however, is too quick, even if the analysis is correct. Kment (2006:271) intends this to be valid only in *standard* contexts. As such, the sphericity principle you can derive by accepting his analysis is not unrestricted and therefore insufficient for purposes of building a counterfactual logic.

⁹See Nolan (1997: 550 and 569) and Vander Laan (2004) for more examples in the same spirit.

particular occasion, we come up with a count of 63 balls from the bag. ‘If the bag had 63 balls in it, 63 would have been a square number’ seems like an appropriate thing for me to say in explaining why I think we miscounted: and I think a true thing to say, in that context. But, of course, it is possible for that bag to contain 63 balls, and impossible for 63 to be square. So at least in some contexts of utterance of counterfactuals, SIC fails. (Nolan 2017:29)

There seems to be a clear reading of the counterfactual in question where it has a possible antecedent and an impossible consequent. The non-vacuaist that wants to keep SIC in place must be involved in the negative project of rejecting these linguistic intuitions. This is dialectically uncomfortable - the point of the addition of impossible worlds, points where impossibilities can be true, seems to be precisely that of avoiding situations like this.

We can conclude that the SIC-based counterfactual logic is built using very shaky foundations. I take it that this would represent a significant cost for the non-vacuaist. In what follows I will defend a non-vacuaist semantics which can provide a counterfactual logic that is robust enough in another way.

3.3 A Possible Solution

3.3.1 Using Only Logically Possible Worlds

My proposal is to use only logically possible worlds in giving our semantics for counterfactuals. There are many ways to find that what a world represents is impossible. If using logic is enough to find that what w represents is impossible, then w is not part of the space of worlds I want to use in my semantics.

These worlds must be fully compositional with respect to the language of classical propositional logic. Since these worlds are logically possible, they must be maximal and closed under the relation of logical consequence. When the truth value of all propositional atoms is fixed, when all of them are either true or false, the truth value of all syntactically more complex formulas follows logically. Since an instance of a failure of compositionality with regards to a logical language is logically impossible, these impossible worlds are compositional.

A consequence of this is that (Closure) is now valid again. If it is the case that $\psi_1 \wedge \dots \wedge \psi_n \rightarrow \chi$ is a logical truth, it will be true at every world in the model. If it was not, then logic would be enough to determine that that world was impossible. This is enough to make the proof of (Closure) given above go through.

More generally, the logic of the non-modal propositional fragment of the vacuaist’s logic is just the same. Possible worlds and the impossible worlds I am using cannot be distinguished by propositional logic alone. If they could, then the impossible worlds would be logically impossible. This means that any validity of the vacuaist’s logic that involves no modal operator is kept intact. If you forget the modal operators, this counterfactual logic *just is* the counterfactual logic of the vacuaist.

Modal operators, when interpreted as concerning metaphysical modality, can distinguish between possible and impossible worlds. The truth conditions for the modal operators, if they are to have the intended interpretation, must be restricted to the former. This means that some validities involving modal operators will be lost, most notably vacuism itself.

But if we reinterpret the modal operators as representing *logical* modality, everything can stay the same as in the vacuist semantics. Every validity in that logic which involves reference to metaphysical modality will now involve reference only to logical modality.

Metaphysical modality, then, is dissociated from the semantics of counterfactuals. I think that this is not a big loss for the logic of counterfactuals. The logic sketched above is quite robust. Metaphysical modality is not an ordinary notion, but a technical term in philosophy. If having a semantics of modality that respects speakers' intuitions better while providing a strong counterfactual logic requires that some metaphysical impossibilities be represented, then so much the worse for metaphysical modality. I think that speakers' intuitions of validity and their deductive behavior are not significantly affected by the transition of using metaphysical modality to using logical modality.

It might be that this is a problem for our theory of *metaphysical* modality. Maybe the association between counterfactuals and modal operators was more important to the modal operators than for the counterfactuals. We will consider whether that is so in the next chapter, but let us bracket that problem for now.

This semantics is non-vacuist. Antecedents expressed by "If Bill Gates were my father" or by "If I were an iPhone" are now true at some worlds, albeit impossible ones. If a proposition is a metaphysical impossibility that is not a logical impossibility it will be true at some worlds and the respective counterfactuals will not be vacuous. There does not seem to be anything logically inconsistent about either mathematical platonism or nominalism. One is true at some worlds and the other is true at some others. This can be iterated for all metaphysical debates where the opposing positions are logically consistent. This is a significantly large class of propositions.

When we move closer to logical impossibility, things are not so simple. Consider analytical truths, where analyticity is conceived as what is sometimes called *Frege analyticity*.¹⁰ A sentence is Frege analytic if it is a logical truth or it can be turned into a logical truth by the substitution of synonymous expressions. Examples are "Bachelors are unmarried" or "Vixens are female foxes". What about counterfactuals with the negations of such sentences as antecedents?

Since they are not logically impossible, they will be true at some world. But in that world the substitution of synonymous expressions cannot be valid, otherwise it would represent a logical falsehood. If you think that synonymy facts are part of a larger class of semantic facts that can be true or false at a world, it would follow that these worlds are worlds where the actual semantic facts change. If the counterfactual is uttered in a context where the semantic facts are to be held fixed, the antecedents will not be true at any world. I

¹⁰See Boghossian (1997).

will treat this issue together with the main problem for my semantics in the following sections.

Countermathematicals are hard to analyse. This is due to the fact that saying something concrete about them involves taking some stand or other on the philosophy of mathematics and the semantics of mathematical discourse. If logicism is true, then mathematically impossible propositions would not be true at any world. If logicism is false, then countermathematicals will not be vacuous. Even in this case, it is possible that mathematical propositions collapse into the case of analytic propositions when the synonymy facts are substituted by mathematical definitions. I will largely ignore countermathematicals in what follows and hope that what I say in the other cases is sufficient to see how the story about them would go assuming a certain philosophy of mathematics.

I hope to have given some idea of how my semantics is non-vacuiist while still providing for a strong counterfactual logic. I now turn to the main challenge it faces.

3.3.2 Counterlogical Vacuism

The semantics we are considering has a problem: a form of vacuism is still valid. If a counterfactual has an antecedent that is logically impossible, then that counterfactual is vacuously true. Call this *counterlogical vacuism*.

My positive reason for being a counterlogical vacuism is that it provides a strong counterfactual logic. I think that counterfactual thinking is logically constrained. Because I think that, I have the intuition that the logic of counterfactuals should be relatively robust. It would be a cost if it were not. By accepting counterlogical vacuism I am trading one cost for another. The trade is profitable if the new cost is smaller than the old one. Though these matters are always to adjudicate, I will give some reasons to think that the price is right in what follows.

There are good reasons to be a vacuist. The only way to eradicate all forms of vacuism is to have any proposition whatsoever be true at one world or other. This forces us to include in our models worlds which are wildly logically unruly. And these worlds must make a difference for the truth conditions of counterfactuals. This transfers the logical unruliness from the impossible worlds to the counterfactuals themselves.

The SIC-based strategy has an artificial feel to it. First, the impossible worlds are contained - they are always farther away in the similarity metric than any possible world. Then, a counterfactual logic is proposed where most validities have a hidden premise that all antecedents are possible. Counterfactual thinking is only logically constrained in the domain of the metaphysically possible.

The current solution is, I think, more principled. It accepts the costs of keeping logically unruly impossible worlds out of our models. In this way, it can provide a robust counterfactual logic without any gerrymandering whatsoever.

I will now turn to some defensive moves. This position is intermediate between that of Berto et al. and Williamson. I will try to convince you that the

problem of counterlogical vacuism is much smaller the problem of full counterpossible vacuism. I will do this by making it seem as small as I can.

3.3.3 Defensive Moves

I will claim that counterlogical vacuism is relatively easy to live with. It is clear that there are much less problematic instances than with full counterpossible vacuism. This is due to the fact that there are less logical impossibilities than there are metaphysical impossibilities. As we saw in the introductory chapter, any logical impossibility is metaphysical impossible but not vice-versa. So, if the problem with a vacuism is proportional to the quantity of ordinary intuitions it violates, counterlogical vacuism does better.

Kment (2014) also defends a theory of impossible worlds where all counterlogicals are vacuously true. Though he does not provide a full defense of counterlogical vacuism, here is what he has to say about the difference between counterlogicals and countermetaphysicals:

I have tried to support the claim that counterfactuals with impossible antecedents can have non-trivial truth-conditions by considering several examples where this seems plausible. Most people I asked agree with my judgment that these counterfactuals are good candidates for non-trivial truth or falsity. However, in my examples the antecedents—e.g., “Thatcher is my mother”—were always logically consistent. People’s responses become more varied when presented with counterfactuals whose antecedents are logically inconsistent, such as “If it were raining and also not raining, then” Such counterfactuals are quite different from the likes of “If Thatcher were my mother,” since their antecedents contradict themselves. Some of my subjects felt that that made it hard for them to understand what scenario they were supposed to envisage and reason about hypothetically, and therefore made it difficult for them to assign truth-values to the counterfactuals in a non-trivial way. I sympathize with this reaction. (Kment 2014: 73)

I also think that this difference makes some strategies have a better chance of being successful when only counterlogical vacuism is in question. I will divide counterlogicals in different classes about which I have something to say. The division will not be exhaustive by any means, but it will give an idea about the extension of the problem of counterlogical vacuism and about the strategies one can use to live with it. The division is as follows: 1- Unentertainable antecedents; 2- Change of Subject antecedents; 3- Countermetalogicals; 4- Philosophy Seminar counterlogicals.

The main reason I think we can live with counterlogical vacuism is that I think that a lot of counterlogicals fall into the first class. Recall from the first chapter Lewis’ unentertainability defense for vacuism. Certain antecedents are unentertainable. We just do now know what to do with them. We have no idea what would counterfactually follow from such a supposition. Our process

of counterfactual evaluation breaks down. And this break down is interpreted as vacuous truth by the semantics. This is exactly the situation where I think vacuism makes most sense. And I think a lot of counterlogicals are like that. Because they are very weird, we do not utter them often, of course. And this means that they do not play a central role in philosophical discussions. But that does not make them any less relevant when it comes to reducing the problematic class of vacuous counterfactuals. Consider the following:

- (1) If the ball were round and not round, we could play football with it

How on earth are we going to develop that counterfactual supposition? Can we play football with a contradictory ball? I just do not know what to say. “If that were so, everything you like would be true!”, as Lewis (1973: 24) said.

Such a defense is much weaker in the case of metaphysical impossibilities, which we can usually imagine in sufficient detail for counterfactual evaluation. As I remarked in the introductory chapter, entertainability is likely to be context dependent. With a sufficiently rich background, maybe we can entertain what would be the case if the ball were round and not round. My point is that in most ordinary contexts we would not.

I intend this to be the main independent defensive strategy for the counterlogical vacuist. For the next two classes of counterlogicals, my strategy will be pragmatic, as the general vacuist strategy was. I will claim that we can reinterpret the logically impossible antecedent into a new, logically possible one. And the resulting counterfactual still respects the intentions of the speaker. Using this strategy represents a very clear cost - we do not want to go around reinterpreting what ordinary speakers say only because it clashes with the predictions of a semantic theory. I do not deny that this is so, even in the case of counterlogicals. But I think that logic is special enough to lend credibility to *some* reinterpretation in *some* cases.

Let us start with what I called Change of Subject antecedents. Quine famously claimed that when a philosopher changed the logic of a discussion she thereby changed the *subject* of the discussion. Here is what he says regarding the debate between a classical and a paraconsistent logician:

My view of this dialogue is that neither party knows what he is talking about. They think they are talking about negation, “ \neg ”, “not”; but surely the notation ceased to be recognizable as negation when they took to regarding some conjunctions of the form “ $p \wedge \neg p$ ” as true, and stopped regarding such sentences as implying all others. Here, evidently, is the deviant logician’s predicament: when he tries to deny the doctrine he only changes the subject. (Quine 1970: 81)

If you believe that contradictions can be true, then you just changed the meaning of conjunction or the meaning of negation or both. Quine’s view is now widely disregarded.¹¹ It seems that philosophers can, in fact, have logical disagreements

¹¹See Warren (2018) for an overview of the discussion sympathetic to Quine’s position.

(I will turn to that in due course). But Quine might be right in *some* situations. Consider a counterfactual like

- (2) If the disjunction behaved as intuitionists thought it did, then $A \vee \neg A$ would have been false

The situation the counterfactual is asking us to consider does not seem to be logically impossible at all. It seems natural to read this counterfactual as regarding a change of the meaning of disjunction. The disjunction cannot behave as the intuitionists thought it did. We know this based on classical logic alone, if that is the correct logic. But the meaning of \vee could have been different. And this is not a logically impossible situation.

The most convincing cases of counterlogicals have been thought to be like the following:¹²

- (3) If intuitionistic logic were correct, excluded middle would have been valid

It seems that (3) ought to be false. But, given counterlogical vacuism, it is vacuously true. I think that the reason why this kind of example is *prima facie* convincing is that we can deduce the negation of the consequent from the assumption of the antecedent in a clear, mathematically precise way. Developing counterfactual suppositions presupposes some rules on how to do so. This is what makes an antecedent entertainable in the first place. And the antecedent of (3) seems very clearly entertainable: we know exactly what to do with it.

I think that the reason we know exactly what to do with it shows the way out of the problem. (3) is not really a counterlogical in most contexts. Its antecedent does not go against classical logic. It is a counter*metalogical*.

Simplifying a bit, metalogic is about either model theory or proof theory. If we read the antecedent is this way, it does not ask us to consider a logically impossible situation at all. In some contexts, it might be making some claims about classes of models. The class of models used to give a model theory to intuitionistic logic is just a collection of formal structures. Considering these models, we then make a claim about them: they are such that excluded middle is valid. And this claim is false. Therefore, the whole counterfactual is false. The models for intuitionistic logic are formal objects. They are not logically impossible, in fact, you might think they *actually* exist. A similar story could be told about proof theory and about which syntactic transformations are allowed in the object language.

Now, in some contexts this reading of the antecedent of (3) does not seem plausible. Maybe the speaker of (3) intends its antecedent to be a supposition about the structure of reality, not about models. A structure that logic alone stops reality from having. I can think only of one class of contexts where this reading is clearly forced upon us: the context of a philosophical debate in a philosophy seminar.¹³

¹²See Williamson (2017: 206).

¹³Obviously, explicitly stated speaker intentions might be sufficient to do this alone. My

Entering the philosophy seminar amounts to guaranteeing that no pragmatic treatment is allowed. When we utter something in such a seminar, we are committed to mean it *strictly and literally*.

If we say that there are tables in a philosophy seminar we are assuming an ontological commitment to tables. If we said the same thing outside, maybe a pragmatic treatment could save us from that ontological commitment. Maybe the assertion that there are mereological simples arranged tablewise can serve the same pragmatic purposes as the assertion that there are tables. And it has the added benefit that it is true even if there are not tables.¹⁴

Entering the seminar room sets us free from all pragmatics. If I ask you to suppose that intuitionistic logic is correct, and make no reference to models or proofs, then you are to take my literal word for it. I take it that this is the sense in which (3) is considered a genuine counterlogical.

And here I must bite the bullet. In this context, such a counterlogical is vacuously true. And its vacuous truth goes against our intuition. I do not think that this is such a big problem. Not many ordinary people find themselves in seminar rooms. If it took such a specialized class of contexts to consider a convincing genuine counterlogical, so be it.

We can still copy Williamson's strategy when it comes to *reductio ad absurdum* arguments to explain the behavior of philosophers. Uttering counterfactuals in the seminar room is demanding. It is not enough that they are true. There must be a connection between the antecedent and the consequent. In most cases it will be not a proof but an argument. The reason (3) would turn out to be unassertable is that there is no argument from its antecedent to its consequent.

All in all, I think I have said enough to conclude that counterlogical vacuism is easier to live with than full counterpossible vacuism.

3.3.4 Upshot

My main motivation for defending the semantics above is two-fold. On one hand, I think it might be the correct reaction to the problems facing full-blown non-vacuism. I hope I have said enough to convince you that this might be the case. On the other hand, I think it is very important to realise that there is conceptual space between the vacuism of Williamson (2017) and the non-vacuism of Berto et al. (2017). Being a non-vaculist regarding counterpossibles does not commit one to deny every form of vacuism. There were good reasons for being a vacuist in the first place. It seems that we should try to solve as much of the problem of counterpossibles as we can, while paying attention to the costs of disregarding these reasons.

goal is not to explain the whole range of counterlogicals away, but to make it more manageable. My thought is that there are not *a lot* (in some precisification of this expression) of contexts where ordinary speakers use counterlogicals non-vacuously and they cannot be explained away via one of these methods.

¹⁴See Dorr (2002: Ch.2) for a discussion of the influence of entering the seminar room over pragmatic treatments in the context of mereology.

With that being said, let me make some very speculative remarks about the significance of moving from full counterpossible vacuism to counterlogical vacuism.

There is a good reason for using circumstances in our semantic theories. Language is used to transmit information. And having information is, in a certain sense, the ability to discriminate between different circumstances.

The empirical success of a semantics which uses circumstances in this way depends on the modal space being used. If a sentence S is taken by ordinary speakers to allow a discrimination between the circumstances c_1 and c_2 , the semantics can only be faithful to the speakers' intuitions if the modal space on which it is based differentiates between c_1 and c_2 .

In this context, identifying the modal space to be used with the space of metaphysical possibilities involves a big risk. We do not know too much about metaphysics. Any discovery we make will shrink the space of metaphysical possibilities. Imagine that Spinoza was right when he said that any truth is necessary. If any truth is necessary, any falsehood is impossible, which means that there is only one possible world. Semantics that use a form of modality as a central notion are constrained by the structure of the corresponding modal space. A semantics based on a spinozist modal space is not going to go very far.

You might think that this is not a problem: Spinoza is almost certainly wrong. But let me tell you a different story.

Mary, the semanticist: Mary is a semanticist. She has a semantic theory that accounts for all the empirical data she knows of. It is a theory that uses a modal space. Mary does not know anything about metaphysical modality. And Mary is happy.

Now Jones comes along. Jones is a metaphysician. He comes bearing good news: the metaphysicians have finally made a discovery. It turns out that the metaphysical proposition φ is true. It also means that φ is metaphysically necessary and $\neg\varphi$ is impossible. Furthermore, any logical consequence of φ is also necessary and any logical consequence of $\neg\varphi$ is also impossible. The metaphysicians are sure of this.

Jones tells the good news to Mary. Mary is now anxious. If what Jones said is true, modal space has just shrunk. This means that the empirical predictions of Mary's theory are now different. Before, they accounted for all the data. Do they still account for all the data now?

Mary runs to check whether some new predictions are false. Sure enough, she finds that some sentences in her theory can only be true if we are allowed to discriminate between situations where φ is true or false. And speakers do have the intuitions that some of these sentences are true. The empirical adequacy of Mary's theory is gone. Mary is unhappy.

It seems to me that there is something methodologically wrong in this story. Semantics has something of a descriptive character. It aims to systematize

ordinary speakers' semantic intuitions. But semantic intuitions are no guide for metaphysical necessity. Therefore, clashes can happen. If the modal space the semanticist uses is the space of metaphysically possible worlds, metaphysics wins these clashes by default. And the descriptive adequacy of semantic theories is in jeopardy.

The situation is similar to one that Maddy (1992) describes in relation to the interplay between set theory and natural science. One of the best arguments for the existence of mathematical objects is the indispensability argument: mathematical objects are indispensable for natural science, therefore we should accept them. Part of what set theorists do is to determine what sets there are. If the only reason we have for the existence of mathematical objects were the indispensability argument, we would be left in the following situation:

We have reached this point: a methodological decision in set theory [...] hinges on developments in physics. If this is correct, set theorists should be eagerly awaiting the outcome of debate over quantum gravity, preparing to tailor the practice of set theory to the nature of the resulting applications of continuum mathematics. But this is not the case; set theorists do not regularly keep an eye on developments in fundamental physics. (Maddy 1992: 289)

Maddy thinks this is the wrong relation between the methodology of set theory and the advances of physics. I think that the same might be the case between the methodology of semantics and the advances of metaphysics.

Chapter 4

Counterfactuals and Metaphysical Modality

This chapter considers the relation between the semantics of counterfactuals and metaphysical modality. We want to know the extent to which the dissociation of the two implied by vacuism is a cost for our theory of the latter. Section 1 presents Williamson's logical equivalence between claims of metaphysical necessity and certain counterfactual conditionals. It shows how it can be used to develop a modal epistemology and to provide a clear conception of metaphysical modality. Section 2 shows how non-vacuism invalidates Williamson's logical equivalence. It then goes on to consider Berto et al.'s answers to the problems this implies and finds them wanting. Section 3 argues that, if Williamson's modal epistemology is successful in the first place, then the non-vaculist can develop a robust modal epistemology in the same spirit. The loss of the logical equivalence does not imply that the evaluation of counterfactuals cannot be used to acquire knowledge of modal claims. Section 4 argues that there is no way a non-vaculist can use Williamson's strategy to provide a conception of metaphysical modality. This is only a problem for the non-vaculist insofar as the notion of metaphysical modality plays an important theoretical role. But our present understanding does not allow us to reach an accurate verdict on this question.

4.1 Williamson's Logical Equivalence

Williamson (2007: 156) defends that the following two principles are intuitively correct:

$$\begin{aligned} \text{(Necessity)} \quad & \Box(p \rightarrow q) \rightarrow (p \Rightarrow q) \\ \text{(Possibility)} \quad & (p \Rightarrow q) \rightarrow (\Diamond p \rightarrow \Diamond q) \end{aligned}$$

(Necessity) says that if it is necessary that p cannot be true while q is false, then p counterfactually implies q . (Possibility) says that when p counterfactually

implies q , then p cannot be possible while q is impossible. These principles seem plausible. They have the effect of establishing that the counterfactual conditional is intermediate in strength relative to the material and the strict conditionals. And, in fact, both are valid in the Lewis-Stalnaker semantics.

If (Necessity) and (Possibility) are true, we can connect counterfactual logic with modal logic. Williamson (2007: 156-9) shows that the following logical equivalences can be derived:

- a) $\Box A \leftrightarrow (\neg A \Rightarrow \perp)$ ¹
- b) $\Box A \leftrightarrow (\neg A \Rightarrow A)$
- c) $\Box A \leftrightarrow \forall p(p \Rightarrow A)$ ²

What these results show is that, assuming (Necessity) and (Possibility), claims of metaphysical necessity are logically equivalent to claims about the truth of certain counterfactuals. We will call this result *Williamson's Equivalence*.

For all we have said until now, this is a merely technical result. But Williamson intends to put it to some philosophical use. The main subject of Williamson (2007: Ch.5) is how to use it as a basis for an epistemology of metaphysical modality.

4.1.1 Counterfactual Modal Epistemology

Williamson wants to use the logical equivalence above to explain how we can have knowledge of metaphysical modality, to provide a *modal epistemology*.

Modal epistemology is a problem in the first place because it seems that “metaphysical modality” is a technical term used exclusively in philosophical discourse. To provide an explanation of how we can acquire knowledge of a certain domain of truths, we must explain how human cognitive abilities are truth-tracking with respect to that domain. What cognitive ability do we have that tracks truth when it comes to the modal domain? Bealer (2002) defends that we have some sort of special kind of intuition that allows us to know when something is metaphysically possible or necessary. Setting aside problems about the clarity of the proposal, the crucial question is this: why would we have such an intuitive ability? We come to have the cognitive abilities we have based on our evolutionary history. But it is hard to see what sort of evolutionary advantage there is in knowing propositions involving metaphysical modality. Evolutionary advantage has to do with the ability to survive and reproduce. It is hard to believe that any modality broader than nomological modality is important for any of the activities that allow biological organisms to survive and reproduce. Is it just an evolutionary accident that we happen to have the sort of intuition that tracks modal truth?

Williamson argues that we can ground our knowledge of metaphysical modality in our knowledge of counterfactuals. He says the following:

¹ \perp here stands for an arbitrary propositional contradiction.

² Here some setup for propositional quantification is assumed.

[M]odulo the implicit recognition of this equivalence [a)-c)], the epistemology of metaphysically modal thinking is tantamount to a special case of the epistemology of counterfactual thinking. (Williamson 2007: 158)

The basic idea is the following. Assume we have an epistemology for counterfactuals, that it is clear that we have the cognitive ability to know counterfactual truths. We can then use this ability to extend our counterfactual knowledge into modal knowledge using the logical equivalence between the two. We have to assume only that knowledge transmits over known logical equivalence. Call this move the *epistemological reduction* of modal truths to counterfactual truths.

For such a reduction to be successful, we need to have an epistemology of counterfactuals in the first place. But the prospects here are much better than in the case of modal epistemology. Counterfactual thinking allows us to extract information from non-actual possibilities. We can learn from mistakes we haven't committed and we can explore alternative scenarios to reality. It seems plausible that this can be helpful.

Imagine that you find yourself in a situation where you choose a certain course of action and happily succeed in attaining your goal. If you evaluate as true the counterfactuals "If I had done X, I would still have attained my goal" and "If I had done Y, I would not have attained my goal", you gain useful practical knowledge for the future. Faced with a similar situation, you should choose courses of action similar to X and avoid courses of action similar to Y. This makes some *a priori* sense. But there is also empirical proof that knowledge of counterfactuals facilitates learning.³

This account would seem to be enough to provide a counterfactual modal epistemology. It makes evolutionary sense that we acquire knowledge of counterfactuals. And, as a matter of fact, we can describe a process we can go through to acquire counterfactual knowledge. Given Williamson's logical equivalence, we can then come to know modal truths. Though the story lacks detail, it seems to include the essential ingredients. Unfortunately, it needs to be complicated further.

As Jenkins (2008) points out, the logical equivalence between claims of metaphysical necessity and claims involving counterfactuals is not sufficient for Williamson's epistemological project. The evolutionary story makes sense for what we could call *ordinary* counterfactuals. But it seems reasonable that the counterfactuals that are equivalent to claims of metaphysical necessity are not ordinary. They involve contradictions, or propositions counterfactually implying their own negation, or universal quantifications over the antecedents of counterfactuals. We need an evolutionary and epistemological account of *these* kinds of counterfactuals. And it is unclear whether we have one.

We can answer this objection by making the account a little more complex.⁴ To see how, consider the following example (Williamson 2007: 171). We seem

³See Byrne (2016) for a very useful survey of the results of the psychological literature on counterfactuals.

⁴See Kroedel (2012) for an extensive discussion of this point.

to know a lot about logic. In particular, we take ourselves to have the following two kinds of knowledge, (1) knowledge of which empirical propositions follow from other empirical propositions, and, (2) knowledge of logical truths. It is clear why (1) can be an evolutionary advantage. Those organisms that have the ability to acquire this kind of knowledge will know more about the empirical world than those who do not. And knowing about the empirical world is important in order to survive and reproduce. But how about (2)? Logical truths do not tell us much about the empirical world at all. A possible reply is this. It is not easy, for beings with our basic biology and physiology and so on, to develop an ability to acquire knowledge of (1) without, at the same time, developing an ability to acquire knowledge of (2). It just so happens, by accident, that it is biologically easier for those abilities to come together than it is for them to be separated. The abilities relevant for (2) are an *evolutionary byproduct* of the abilities relevant for (1). And that is just lucky for our knowledge of logic.

The same answer can be given when it comes to counterfactuals. We have reason to believe that knowledge of ordinary counterfactuals is an evolutionary advantage. But there is no easy way for beings like us to have the cognitive abilities relevant for ordinary counterfactuals without thereby coming to have the cognitive abilities that allow us to have knowledge of the weird counterfactuals related to metaphysical necessity. Whether or not this story is actually true (and I have no way to know if it is), it does seem like a *prima facie* plausible explanation.

This closes my initial exposition of Williamson's modal epistemology. It allows us to ground our knowledge of esoteric modal truths in the cognitive capacities we develop to deal with mundane and practically relevant claims. We will have the opportunity to examine in more detail how the story goes in practice in section 3.

4.1.2 A Conception of Metaphysical Modality

I do not think that the only thing to come out of Williamson (2007: Ch.5) is a modal epistemology. I will argue that he also provides an *explanation* as to what metaphysical modality *is*. And this aspect should not be downplayed in our discussion.

As remarked above, “metaphysical modality” is a technical term in philosophy. As such, you would expect that philosophers would be able to explain it clearly, and in a uniform way, to outsiders. Much in the same way that a mathematician could explain technical terms in mathematics, like “bijection”, to an intelligent person given enough time. Unfortunately, you would be wrong.

There is no agreement as to what metaphysical modality is among philosophers. Some people, inspired on Fine (1994), think that something is metaphysically necessary if and only if it follows from the essences of things. But what is an essence? Some others think that metaphysical modality should be treated as an unanalysed primitive, as in Wilsch (2017). But that does not provide much illumination. Others still, like Sider (2011), think that there is nothing special about metaphysical modality, as far as modalities go. It is just the one modality

which we happen to find particularly interesting given our parochial interests as philosophers - it provides an adequate spotlight to logic, mathematics and metaphysics, the disciplines *we* take to be important to describe non-contingent reality.

The best we can do given all this, as argued in Rosen (2006), is to explain a number of paradigmatic examples and hope that the layman develops some ability to deal with the concept based on them. In fact, I did just that in the introductory chapter. It might seem like there is nothing particularly wrong with this. There are a great many concepts of ordinary life that no one knows exactly how to define. For example, do you know what is a *dessert*? Would you be able to give necessary and sufficient conditions for some food to count as a dessert?

In fact, I think the situation is much more problematic than in the case of “dessert”. “Metaphysical modality” is a *technical* term. We *introduced* it and we do not know how to explain it. We cannot rely on ordinary linguistic practice to fill in the details.

In any case, it would be better if we had a clear conception of metaphysical modality. And we can find one in Williamson’s logical equivalence. Recall the equivalence c) above:

$$c) \Box A \leftrightarrow \forall p(p \Rightarrow A)$$

A natural way to explain metaphysical modality is as follows. A proposition is metaphysically necessary if and only if it would have been true no matter what else was the case. c) allows a precise rendering of this explanation using the counterfactual conditional. It seems reasonably clear to me that anyone that understands the counterfactual conditional can understand this explanation. If we really do need to use the notion of metaphysical modality, providing such a clear and intuitive explanation of what it is can be a very significant theoretical advantage. And this is one Williamson can reclaim for himself.

That Williamson is at least partly involved in this project can be seen from some passing remarks he makes.

According to [c)] something is necessary if and only if whatever were the case, it would still be the case (see also Lewis 1986: 23). That is a natural way of explaining informally what metaphysically necessity is. (Williamson 2007: 159)

Whoever has what it takes to understand the counterfactual conditional and the elementary logical auxiliaries \neg and \perp has what it takes to understand possibility and necessity operators. (Williamson 2007: 158)

This finishes my brief exposition of the philosophical uses of Williamson’s logical equivalence. Let us see next what the role of these considerations is in the debate against the non-vacuumist.

4.2 Non-Vacuism and the Logical Equivalence

It seems that non-vacuism breaks Williamson's logical equivalence. Against a): the proposition that Bill Gates is my father is metaphysically impossible, but it counterfactually implies no contradiction. Against b): the proposition that Bill Gates is my father does not counterfactually imply its own negation. Against c): The proposition that Bill Gates is not my father is necessary, but its negation does not imply it. As a matter of fact, any metaphysical impossibility that is not logically impossible would do as a counterexample to the three logical equivalences.

This happens because non-vacuists seem committed to denying (Necessity). It can be the case that $\Box(p \rightarrow q)$ is true without it being the case that $(p \Rightarrow q)$. Any strict conditional which has the proposition that Bill Gates is my father as an antecedent is true: that proposition is false at all possible worlds and material conditionals are true when their antecedent is false. But it is not the case that the proposition that Bill Gates is my father counterfactually implies everything. That is the very intuitive justification for non-vacuism.

Prima facie, this seems to be cost for the non-vacuist. The logical equivalence did some philosophical work. If that philosophical work turns out to be important, losing it is a theoretical cost.

Berto et al. (2017) try to find a way around this situation by claiming that a form of Williamson's logical equivalence can be maintained in a non-vacuist setting. In what follows I will argue that they do not succeed. In the following sections we will turn to the question of how much weight these costs should have in our cost-benefit analysis.

4.2.1 Berto et al.'s Solution

Berto et al. (2017: 17) propose a reinterpretation of \perp . As we have seen, in Williamson's system it just stands for an arbitrary contradiction. Berto et al. propose that the semantic value of \perp is given by the following condition: \perp is a 0-ary connective that is false at all possible worlds and true at all impossible worlds. After this reinterpretation, and given SIC which Berto et al. accept, a) above is true: something is metaphysically necessary if and only if its negation counterfactually implies \perp . If something counterfactually a contradiction, then all the closest worlds where it is true are worlds where a contradiction is true - impossible worlds. And the only way this can happen, given SIC, is for that proposition itself to be impossible. What is broken is the equivalence of a)-c): b) and c) remain false.

They argue (Berto et al. 2017:18) that this substitution does not detract anything from the epistemological value of a). An epistemology of modality with the theoretical advantages of Williamson's can be based on this equivalence. The only objection they consider is the following: now \perp is inherently modal; therefore we are smuggling in modal knowledge implicitly. They note that, for Williamson's reduction to work in a Lewis-Stalnaker setting, Williamson must not just assume that contradictions are false, or even known to be false, but

that contradictions are *impossible*. Therefore, Williamson is building in some modal knowledge into his account as much as Berto et al. are.

But the problem with Berto et al.'s account is different. They have reduced to zero any epistemological value that comes from recognizing the logical equivalence a). Evaluating the relevant counterfactuals of the form $(\neg A \Rightarrow \perp)$ is as epistemologically difficult as knowing whether A is metaphysically necessary. There is nothing to be gained from the epistemological reduction.

In Williamson's picture, you get to know that a proposition A is necessary using roughly the following method. You recognize the equivalence in a). Then you assume counterfactually that $\neg A$. You proceed to develop this counterfactual supposition until you manage to find a contradiction. You conclude that $(\neg A \Rightarrow \perp)$ is true and reason back to A 's necessity using a). What would the same process look like with the new interpretation of \perp ? What are you looking for under the counterfactual supposition that A 's negation is true? \perp seems to be little more than an impossibility sign post, as it were. There is no reason to think that we have developed the cognitive abilities to recognize such a thing in counterfactual thinking. Therefore, we can conclude that this is no good foundation on which to build a modal epistemology.

Berto et al. do not make any remarks concerning the conception part of Williamson's contribution. Let me just note that their proposed strategy does not solve this problem either. A plausible test for whether you have a good conception of a term is whether it is sufficient to explain your use of it to an intelligent layman. It seems to me that if she does not understand the concept of "metaphysical modality" she will not understand your explanation of what \perp means. Your explanation has a circular feeling to it. While it might be reasonable to expect that someone can understand what a contradiction is while not grasping the concept of "metaphysical modality", the same seems hard to believe for this new interpretation of \perp . Those that can be helped using this strategy need no help in the first place.

Generally speaking, I think we can conclude from the points above that Williamson's logical equivalence *is* tied strongly to his vacuism. Just reinterpreting the symbols used will not help.

4.3 A Non-Vaculist Counterfactual Modal Epistemology

It seems to me that losing Williamson's modal epistemology is not a very significant cost for the non-vaculist. First of all, there are other modal epistemologies in the philosophical marketplace.⁵ If any of these candidates turns out to be satisfactory, then the non-vaculist can use it. There might be some theoretical cost associated with a loss of unification, but the non-vaculist is not very likely to care. That being said, I will argue that (assuming that Williamson himself succeeds in this) the non-vaculist can develop a satisfactory counterfactual

⁵See Vaidya (2015) for an introduction to the general field of modal epistemology.

modal epistemology.

Williamson provides a modal epistemology via what I called an epistemological reduction of the epistemology of modality to the epistemology of counterfactuals. He uses his logical equivalence in order to do this. Jenkins (2008), as we saw, made the point that the logical reduction was not sufficient for the epistemological reduction. In this section I will make the converse point - the logical equivalence is not necessary for the epistemological reduction either.

Here is a way to gloss what we mean by the epistemological reduction of domain D' to domain D . First, we can obtain, via an independent epistemological process, warranted beliefs in propositions of D . Secondly, our warrant for beliefs in D is transferred to a warrant for our beliefs in D' . In what follows I will use a notion of *warrant* for a belief that is both *defeasible*, in that acquiring new information can sometimes defeat our warrant for a given belief, and *fallible*, in that sometimes we have a warrant to believe something false.⁶

In Williamson's case, the idea that we developed the cognitive abilities to have counterfactual knowledge is intended to justify the first step. The logical equivalence takes care of the second one. And it does so in quite a good way. It seems very reasonable that warrant transfers over (known) logical equivalence. But, conversely, there is no reason to think that logical equivalence is the only way to transfer warrant.

Imagine that you want to consider an epistemological reduction between facts about our perceptual appearances and facts about the external world. This seems like a good reduction to pursue. If it *seems* to me that p , and I am warranted in believing that it does seem to me that p , then I can gain warrant for believing that p is true based on this belief. But there is no logical equivalence (nor logical *implication*) between how things *seem* like and how things *are*. There should be at least conceptual space for an epistemological reduction without logical equivalence (nor logical implication).

Having such a non-logically based epistemological reduction does run some risks. If it logically followed from the fact that it seems to me that p that p is true, then there would be no *skeptical scenarios* to worry about. Such scenarios would be logically impossible, so I can never find myself in one of them. If skeptical scenarios are not ruled out by logic, there is a good chance that they are not ruled out at all. There seems to me nothing rulling out that my perceptions are orchestrated by an evil demon and, as such, have no relation to the external world at all.

In general, this should not be a problem. I can know that there is a tree outside if it seems to me that there is a tree outside as long as there is no reason for me to believe that I am not in a skeptical scenario and if I am not, in fact, in a skeptical scenario. This seems to be a very reasonable way to go about this particular epistemological puzzle.⁷

I propose that a similar epistemological reduction can be obtained in the case of counterfactuals and modal claims. Losing Williamson's logical equiva-

⁶See, for example, Pryor (2000) for a thorough justification of this use of the term.

⁷It is the position Pryor (2000) calls "dogmatism".

lence opens up the possibility of skeptical scenarios. As long as they are not widespread, and we have no reason to think we are in one of them, we can still come to know modal claims using Williamson's method. A single example is enough to refute a logical equivalence. But it is not enough to refute an epistemological reduction.

4.3.1 Skeptical Scenarios

Williamson describes two ways in which we can gain modal knowledge based on the evaluation of counterfactuals, one for possibility claims and one for necessity claims:

- (K \Box) If you have a warrant to believe that the counterfactual ($\neg A \Rightarrow \perp$) is true, you have a warrant to believe that A is necessary
- (K \Diamond) If you have a warrant to believe that the counterfactual ($\neg A \Rightarrow \perp$) is false, you have a warrant to believe that A is possible⁸

Williamson gives a summary of how we come to have the counterfactual beliefs we do as follows:

[O]ne supposes the antecedent and develops the supposition, adding further judgments within the supposition by reasoning, offline predictive mechanisms, and other offline judgments. The imagining may but need not be perceptual imagining. All of one's background knowledge and beliefs are available from within the scope of the supposition as a description of one's actual circumstances for the purposes of comparison with the counterfactual circumstances [...]. Some but not all of one's background knowledge and beliefs are also available within the scope of the supposition as a description of the counterfactual circumstances, according to complex criteria (the problem of cotenability). To a first approximation: one asserts the counterfactual conditional if and only if the development eventually leads one to add the consequent. (Williamson 2007:152-3).

As Williamson recognizes, this makes (K \Box) much more reliable than (K \Diamond). A counterfactual evaluation of a counterfactual of the form ($\neg A \Rightarrow \perp$) is a search for a contradiction. If you find what you are looking for you can straightforwardly conclude that the counterfactual is true. But what can you do if you do not find a contradiction? You can only be as confident that the counterfactual is false as you are confident about the quality of your search. You cannot conclude that there are no elephants in Amsterdam by looking outside your window. Establishing this requires a much more thorough search.

As it turns out, the non-vaculist has a relatively easy time with (K \Box). In fact, if the Strangeness of Impossibility condition were true, the non-vaculist can just *copy* it. Assume that we know that ($\neg A \Rightarrow \perp$) is true. The non-vaculist accepts,

⁸He does not put it in warrant-theoretical terms, of course: see above for my justification for this.

as Williamson does, that we know that contradictions are impossible. And, if SIC is true, so is (Possibility). This means that if $\neg A$ was possible, \perp would have been possible as well. Since contradictions are not possible, $\neg A$ must be impossible as well and A must be necessary, just as Williamson concludes.

As I argued in chapter 3, I do not think that SIC is true. I think that there are contexts where it fails. This means that SIC is not good enough for a counterfactual logic. But it may still be good enough for a modal epistemology. We have discovered our first skeptical scenario:

Skeptical Scenario #1: We try to use $(K\Box)$ in a context where SIC is false

The examples we gave of SIC being false, such as “If Gödel had thought that Fermat’s Last Theorem was false, then Fermat’s Last Theorem would have been false”, are quite weird and uncommon. Weird and uncommon counterexamples are sufficient to show that a logical principle is false. But they are not enough to send us into skeptical despair.

The case of $(K\Diamond)$ is harder for the non-vacuaist. The non-vacuaist is committed to thinking that some metaphysical impossibilities simply do not counterfactually imply contradictions in normal contexts. This would seem to reduce the hope of having a non-vacuaist counterfactual modal epistemology to nothing. As it turns out, this is too quick. To understand why, we need to discuss how contradictions appear in Williamson’s epistemological story in more detail.

Consider the proposition that the atomic number of gold is 79. This proposition is metaphysically necessary. But the non-vacuaist does not think that its negation counterfactually implies a contradiction. Williamson must disagree. Here is his explanation of how this comes to be:

Once we take something to be a necessary truth, of course we can use it in developing further counterfactual suppositions. But that does nothing to show that we have any special cognitive capacity to handle modality independent of our general cognitive capacity to handle counterfactual conditionals. If we start only with the latter, just as envisaged above, it will generate knowledge of various modal truths, which can in turn be used to develop further counterfactual suppositions, in a recursive process. For example, we need not judge that it is metaphysically necessary that gold is the element with atomic number 79 before invoking the proposition that gold is the element with atomic number 79 in the development of a counterfactual supposition. Rather, projecting constitutive matters such as atomic numbers into counterfactual suppositions is part of our general way of assessing counterfactuals. The judgment of metaphysical necessity originates as the output of a procedure of that kind; it is not an independently generated input. (Williamson 2007: 170)

Williamson’s explanation is puzzling. His claim that it is part of “our general way of assessing counterfactuals” that in the middle of developing a counterfactual supposition we consider its very negation as being true strains belief.

This is not the way we do it at all. Whatever the answer to the problem of co-tenability turns out to be, we have to forget that the negation of a counterfactual supposition is actually true when evaluating the relevant counterfactual. The key notion seems to be that of a “constitutive fact”. Constitutive facts are exceptions to this generalization. We hold them fixed when evaluating *any* counterfactual whatsoever.

This is Williamson’s story for how we find a contradiction when counterfactually supposing that the negation of a constitutive fact is true. Call a proposition that would be a constitutive fact if true a *constitutive* proposition. If a proposition is constitutive, we can only use (K \diamond) if we accept his claim about our ordinary methods of evaluating counterfactuals. Otherwise, if the negation of the constitutive fact is not assumed, we will fail to find a contradiction independently of how thorough our search is. And it seems very likely that the non-vacuaist does not accept this picture of ordinary counterfactual evaluation.

We have found our second skeptical scenario:

Skeptical Scenario #2: We try to use (K \diamond) to determine the possibility of a constitutive proposition

This means that we cannot use this modal epistemology to find the modal status of constitutive facts. I will address this question in the next section.

What about propositions which are not constitutive, are metaphysically necessary and do not counterfactually imply a contradiction? An example would be any proposition whose negation follows from a constitutive fact, such as the one expressed by “This watch is made of a material whose atomic number is 89”.

Here again Williamson’s technique for finding a contradiction involves a clash with the constitutive facts. But there is no reason why the non-vacuaist cannot follow him there. The difference is that Williamson thinks that the constitutive facts are held fixed due to the general nature of our process of counterfactual evaluation. The non-vacuaist denies this. In contrast, the non-vacuaist needs to be in a suitable context, one where constitutive facts are held fixed. We have found our third and last skeptical scenario:

Skeptical Scenario #3: We try to use (K \diamond) in a context where the constitutive facts are not held fixed

The second and third scenarios involve Williamson’s notion of constitutive fact. Understanding why these skeptical scenarios are not threatening requires some discussion of this notion.

4.3.2 Constitutive Facts

Williamson never explains what he means by “constitutive fact”. We can get a partial characterization via the theoretical role it is supposed to play. A constitutive fact is something that is kept fixed in every counterfactual evaluation. As non-vacuaists, this will not help us much since we think that *nothing* can play this theoretical role.

I am attracted to another gloss on constitutive facts. I think they should rather be called “basic modal facts”. I think that a characterization of basic modal facts comes with each conception of metaphysical modality. These are just the facts for which *truth is sufficient for metaphysical necessity*. In an essence-based account, essential facts play this role. If a true proposition is an essential fact then it is trivially a necessary truth. On Sider’s account, if some truth belongs to one of the domains in his list, then it is necessary. Philosophers very often talk like this: a proposition’s being a mathematical truth is enough for it to be necessary and its being a mathematical falsehood is sufficient for it to be impossible. There is no complicated detour over modal epistemology: truth and subject matter suffice. And, for Sider, truth and subject matter is what fixes basic modal facts. I think this pattern generalizes. In particular, Williamson’s conception of constitutive facts just follows from his conception of metaphysical modality based on counterfactuals.

I think this suggests that we do not need a *modal* epistemology for basic modal facts. For basic modal facts, truth coincides with metaphysical necessity. And it does so analytically, as it were, based on what we mean by metaphysical necessity.

This means that a modal epistemology for basic modal facts is just a non-modal epistemology for whatever science studies basic modal facts. What that science is will vary across different conceptions of metaphysical modality. It would be foolish to use a modal epistemology to find out whether Lewisian modal realism is true, for example. Only someone with a conception of metaphysical modality which does not provide any way of identifying basic modal facts would have any trouble with their epistemology.⁹

This discussion is supposed to show that the second skeptical scenario is non-threatening. It represents no epistemological loss that we cannot use counterfactuals to know whether basic modal facts are necessary or possible. We have other ways to do it.

This links to our third skeptical scenario. If we are evaluating counterfactuals in order to obtain modal knowledge and we are to be as successful as Williamson, we have to make sure we keep fixed the basic modal facts - we need those to generate the relevant contradictions. I think it is very plausible to think that we can do this in most contexts - basic modal facts are so central to our conception of metaphysical modality that we do not vary them when we want to find out whether something is metaphysically possible. In some contexts, we find it natural to keep psychological, scientific or semantic facts intact. For our modal epistemology to work, we just need to find it plausible that, in a context where

⁹You might think that it is the job of a modal epistemology to explain *why* the basic modal facts have their status. The first thing to notice is that this is a lot to demand from an epistemology and it might be theoretically fruitful to distinguish between this foundational project and the general epistemology of non-basic modal claims. Compare the epistemology of mathematics with the difficult field of the epistemology of mathematical axioms. Secondly, my aim in this section is to provide an epistemology that shows that losing Williamson’s strategy is not a cost for the non-vacuaist. And I very much doubt whether Williamson has solved the epistemological problem of basic modal facts with his conjectures about how we always import them to our counterfactual evaluation, even when considering their own negation.

we want to find out whether something is possible or necessary, we keep the basic modal facts fixed.

What we are doing, in effect, could be understood as a relabelling of \perp , just as Berto et al. did. The difference is that we are not just postulating that it is true at every impossible world. It now represents whatever forms the core of our very notion of metaphysical impossibility - the negation of what we have been calling “constitutive facts”. And this promises to bring back some epistemological gain from the evaluation of counterfactuals with \perp as a consequent.

Imagine that \perp stands not for an arbitrary contradiction, but for the negation of known basic modal facts. We know that contradictions are impossible, so those are included. But we can also develop our counterfactual supposition to contradict essential facts, or mathematical facts, or even metaphysical facts if we know any.

The upshot is this. If you are in a context where SIC holds and the context is such that basic modal facts are kept intact, then you can use the evaluation of Williamson’s counterfactuals to derive knowledge about the modal status of non-basic modal propositions. Given the details of Williamson’s modal epistemology, I think that my proposal will be found acceptable by those convinced by Williamson’s epistemology, but not by his vacuism, in the first place.

4.4 The Conception Problem

Williamson gave us an epistemology and a conception of metaphysical modality. As I have argued, I do not think losing the epistemology is a prohibitive cost for the non-vacuiist. There are other alternative ways to build a modal epistemology and I have even suggested that one building on Williamson’s might work. I find that the losing Williamson’s conception of metaphysical modality is a bigger problem, though I must confess not to know how big.

Given non-vacuism, metaphysical modality is not what Williamson thinks it is. There is no way around this. There are metaphysical necessities that would have been false under certain counterfactual suppositions. This means that the non-vacuiist should search the philosophical marketplace in search for a different candidate - much as I have recommended in the case of epistemology above. The problem is that this marketplace is not as good.

Fine’s essence-based conception of metaphysical modality might be the better developed alternative. But theorizing about essences is not to everyone’s taste. And, in any case, it would be a surprising consequence of being a non-vacuiist that we must. What other options are there?

Sider’s (2011) conception of metaphysical modality is thoroughly deflationist. He thinks that the reason metaphysical modality has the extension it has is dependent on our parochial philosophical interests. There is nothing metaphysically significant about metaphysical modality. It is not one of the fundamental building blocks of reality. There need be no mentions of metaphysical modality

when writing the ultimate book of the world.¹⁰

A recent paper by Justin Clarke-Doane (2017) echoes this despair. Clarke-Doane runs through all the ordinary ways in which we are inclined to say that something is necessary and he finds no way of explaining why the extension of metaphysical necessity is what the philosophers take it to be.

In general, I think some form of deflationism about metaphysical modality is natural for a non-vacuumist. Non-vacuumism should at least be committed to some degree to the deflation of the importance of metaphysical modality *for semantics*. My version of non-vacuumism certainly wears this on its sleeve. It would be natural to extend this deflationist feeling to other areas of philosophy, including metaphysics itself.

But this can give the vacuumist the ingredients for an argument. If it can be shown that we need metaphysical modality for our philosophical theorising, and if the vacuumist's conception of it is the most theoretically fruitful, then this is a significant benefit for vacuumism and a significant cost for non-vacuumism. Given semantic consequentialism, in some ways of evaluating costs and benefits this could tip the balance of the debate.

For what is worth, this is my biggest reservation in being a non-vacuumist. I have the intuition that Williamson's conception of metaphysical modality is the best one on offer. I am led by non-vacuumism to reject it - it is a beautiful idea that does not work. But I find it natural to then be a modal deflationist. If I were to find that being a modal deflationist puts me at a significant theoretical disadvantage compared to other theorists, maybe my pull towards non-vacuumism would be weakened.

In the present state of the debate, I do not think we can judge whether this is a significant cost or not. Only recently, with papers such as Clarke-Doane (2017) and Williamson (2016), has the question of deflationism about metaphysical modality been taken seriously.¹¹ Only after more work has been done on figuring out exactly what the notion of metaphysical modality can do for us and, conversely, what the cost of being a modal deflationist is, can we reach a more informed verdict about the magnitude of this particular cost.

¹⁰For a similar diagnosis about metaphysical modality, see Cameron (2009).

¹¹Not that there weren't precedents in the past, as in Quine (1955) or Sidelle (1989), but the general philosophical opinion was unconvinced.

Chapter 5

Counterfactuals and Worldly Opaqueness

In this chapter we will be dealing with the question of whether the substitutivity of identicals should be valid for counterfactuals. The reason this topic is important for us is that counterpossible non-vacuism has been taken to provide motivation for the denial of the substitutivity of identicals in the semantics for counterfactuals.

Section 1 lays the terminological groundwork for our discussion. Section 2 characterizes the nature of Williamson's (2017: Section 4) challenge to non-vacuists which defend that the substitutivity of identicals fails. Section 3 critically evaluates the efforts of Berto et al. (2017: Section 3.2) to respond to this challenge and finds them unsatisfactory. Section 4 defines my stance on this issue, that we should restrict the space of impossible worlds used in our non-vaculist semantics in such a way as to make the substitutivity of identicals valid. The restricted form of vacuism that follows is a smaller theoretical problem than the challenge of explaining how worldly opaqueness makes sense.

5.1 Preliminaries

Let us start by fixing some terminology needed to make sense of the discussion in this chapter.

A key idea is that of *hyperintensionality*. A lot of things can be said to be hyperintensional in philosophical parlance: theories, sentences, contexts, philosophical domains, and so on. The sense that will matter to us is the one concerning *sentence positions* and, by extension, what I will call *semantic constructions*, the things that are obtained by substituting the constituents of a sentence for variables ranging over expressions of the appropriate type. Saying that a sentence position is hyperintensional is saying something about what substitutions can be performed in that position keeping the truth value of the sentence intact.

To understand what hyperintensionality means, we must start by understanding what “*extension*” means. Different expressions are associated an extension by semantic theory. A proper name is associated an object, a predicate is associated a set of objects that actually satisfy the predicate and a declarative sentence is associated a truth value. The notion of extension can also be extended to other parts of the language. The matter is, to some degree, a theoretical one. Let us assume that the kind of thing an extension is is fixed in this way. The only examples that will matter to us are the ones I just gave.

In a possible worlds framework, we can define the *intension* of an expression. It is the function from possible worlds to the extension of the expression in that possible world. It is important to note that what matters is the extension that the expression in question has given its semantic content in the *actual* world. Assuming that there is something that fixes this semantic content, call it the *metasemantic facts*, and that those can vary from world to world, it can happen that the same expression would have a different semantic content relative to the metasemantic facts of a different world. Example: the extension of “Saul Kripke” will be the individual Saul Kripke at every possible world, even at a world whose speakers use the expression “Saul Kripke” to refer to a dog, something (necessarily) non-identical with Saul Kripke.

Now consider a sentence position. It is occupied by a certain expression which has a certain extension. Let us consider the possible grammatically correct substitutions that we can perform on that position. If all the substitutions that substitute the original expression for one with the same extension preserve the truth-value of the original sentence, the sentence position is said to be *extensional*. If all the substitutions that substitute the original expression for one with the same intension preserve the truth-value of the original sentence, the sentence position is said to be *intensional*. If a sentence position is neither extensional nor intensional, it is *hyperintensional*.

These terms are generalized to semantic constructions by generalizing over some or all of the actual constituents of the sentence in question. For example, the sentence “John believes that the ball is red” can be made to correspond to the semantic construction “*A* believes that the *F* is *G*” or the slightly more abstract “*A* ψ ’s that the *F* is *G*” or something of the same sort.

The other substitutional property of sentences and semantic constructions that will be important is the so-called substitutivity of identicals. In this context, it is something of a misnomer. What is substituted are not objects but expressions. And the substitutions in question are not of identical *expressions*. The way we will use it will concern the substitutivity of *rigid designators*. A rigid designator, following Kripke (1980), is an expression that designates the same object at all possible worlds where that object exists.¹ A sentence position respects the substitutivity of identicals if the substitution of a rigid designator for another which designates the same object preserves truth value. We will furthermore assume that proper names in natural language are rigid designators.²

¹We will not care about what the rigid designator does when that object is not around.

²See Cumming (2013) for an introduction to the philosophical questions surrounding proper

It is easy to see why we can call the substitution of such expressions “substitutivity of identicals” – the object is so tightly connected with the designating expression that the distinction between the two is blurred. I will indulge in the misnaming myself and talk about “substitutivity of identicals”. The relevant literature does this and the terminology is very much fixed. But it is important to be clear at the outset about the way in which we will be using the term.

If a semantic construction respects the substitutivity of identicals we will say that it is *transparent*. If it does not, we will say it is *opaque*.

Finally, a key notion in this chapter will be the one of *worldliness*, of an expression’s being *worldly*. There is no precise definition of the term in the literature. For dialectical reasons, I will use the gloss given by Williamson:

[Worldly Expressions] are about the very objects, properties, relations and states of affairs their [constituent expressions] are about, not the ways in which agents represent those objects, properties, relations and states of affairs. (Williamson 2017:209-10)

The formulation is as precise as the notion of (partial) aboutness being used. I do not have a general theory of aboutness to give.³ I take it that we have some intuitions about the notion of “aboutness” – those that would be the data such a theory would have to respond to. We will have to let us be guided by them in the discussion that follows.

One important point to note is that a worldly expression *can* be about representations. This is due to the fact that representations are often objects. A word is something in the world, whatever that something might be. If there is some constituent in a worldly expression that is about a representation *as an object*, then it is partly about representations. It is clear that this does not say anything against the characterization above. We should not let this point be a source of confusion in what follows.

We have now said enough to formulate Williamson’s objection to non-vacuism.

5.2 Worldly Opaqueness

Counterfactuals are usually taken to be worldly – to be (wholly) about the things their constituent expressions are about and not about how these things are represented by representational agents. The counterfactual “If Jones had kicked the ball, he would have scored the winning goal” is about Jones, kicking, the ball and the goal that would have won something.

I will assume that this is right for the purposes of my discussion, as do Berto et al. (2017:9). It definitely *seems* right. One of the ways to demarcate the slippery distinction between indicative conditionals and counterfactuals has to

names, on which Kripke (1980) is a classic.

³Yablo (2014) is the main responsible for bringing the topic of aboutness back into philosophical attention. See Hawke (2017) for a presentation of the available theories of aboutness in a common framework in which they can be compared.

do with the epistemic character of the first and the non-epistemic character of the second.⁴

One bad reason for thinking that counterfactuals are epistemic has to do with their context-sensitivity. It definitely seems like the truth value of a counterfactual depends on what agents care about. In some contexts, “If Caesar had been the commander in the Korean war, he would have used nuclear weapons” seems true while in others “If Caesar had been the commander in the Korean war, he would have used catapults” seems true instead. And plausibly this has to do with whatever the people in the conversation think relevant to hold fixed: whether the facts about Caesar’s psychology or about the weaponry he had at his disposal when he was alive.

But this does not show that counterfactuals are not worldly. These facts about representational agents matter in fixing the relevant similarity metric in the context. When we come to evaluate the counterfactuals, we still care only about the object that the expression “Caesar” is about and about what that object does in some set of non-actual situations.

It is open to the non-vacuumist to develop a theory of counterfactuals where they turn out not to be worldly. But, for the purposes of my discussion, I will set that option to one side.⁵

Moreover, it seems that counterfactuals are transparent. Consider the following examples by Williamson (2007: 174-6), numbers as in the original:

(34) If the rocket had continued on its course, it would have hit Hesperus.

(36) If the rocket had continued on its course, it would have hit Phosphorus.

It seems to us that those counterfactuals should have the same truth value, given that Hesperus is identical to Phosphorus – hitting Hesperus is just the same thing as hitting Phosphorus.

But, given non-vacuism, some examples come up that threaten the transparency of the semantic construction associated with the counterfactuals (numbers again as in Williamson 2007: 174-6):

(32) If Hesperus had not been Phosphorus, Phosphorus would not have been Phosphorus.

(33) If Hesperus had not been Phosphorus, Hesperus would not have been Phosphorus.

It is clear that (33) can be obtained from (32) via an instance of substitution of co-designating rigid designators. It is also clear that (33) should be true since it

⁴Weatherson (2001), for example, uses this idea, claiming that the difference between counterfactuals and indicative conditionals is analogous to the difference between the necessary and the *a priori*. See also Stalnaker (1975) for an account of indicative conditionals that treats them with an epistemic flavor.

⁵Berto et al. identify Lycan (2001) as an account of counterfactuals according to which they are partly epistemic. Discussing Lycan’s theory is outside of the scope of this thesis, but see Weatherson (2002) for some criticisms.

is just an instance of reflexivity. But some non-vacuists, Berto et al. included, want to claim that (32) can be false. Here is their reasoning:

(32), on the other hand, is implausible; although Hesperus and Phosphorus are identical, if they had not been nothing follows about the self-distinctness of one of them. (Berto et al. 2017: 8)

This would mean that counterfactuals would be opaque. Connecting this with the assumption above, the result is that counterfactuals would be an example of a semantic construction that is both worldly and opaque.

The problem is, claims Williamson, that there cannot be any such thing as *worldly opaqueness*. The discussion in Williamson (2017: Section 4) is not very clear on what the problem with worldly opaqueness is supposed to be. I will reconstruct it as an intelligibility worry: we just have no idea of what it would mean for an worldly expression to be opaque.

I do not wish to claim that this was the force Williamson intended his argument to have. I do think it is the most plausible way to go. I just do not think that the notion of worldliness is precise enough for there to be an explicit *argument* from worldliness to transparency. I fear that any way to make the term more precise would risk straightforwardly begging the question against those that think there can be worldly opaqueness. The situation is not untypical of philosophical theorizing: we have to carry along a notion regarding which we have some intuitions but no explicit characterization.

The reason we might take the notion of worldly opaqueness to be unintelligible is roughly this. A sentence is worldly if it is about the things its constituent expressions are about. Two co-designating rigid designators are about the same thing, the object they designate. Therefore, two sentences that are the result of a substitution of identicals are about the same things. Fixing a context of evaluation, how can two worldly expressions with the same syntactic structure be about the same things and still differ in truth value?

This is not supposed to mean that it cannot be the case: it is not an argument to the conclusion that worldly expressions must be transparent. The dialectical import of an accusation of intelligibility is essentially a demand for explanation. It is a device to shift the burden of proof. As such, it is as strong as candidate explanations are weak. In the next part of this chapter we will consider the style of explanation adopted by Berto et al.

5.3 Defensive Moves

Berto et al. (2017: 3.2) propose two defensive moves against Williamson's claim that opaque worldly constructions are unintelligible. The moves are:

- 1- Claim that there have been cases of worldly opaqueness defended in the literature, namely in Nolan (2014) and in the literature on metaphysical grounding.

- 2- Claim that they have provided a model of worldly opaqueness using impossible worlds – the existence of a precise formal model is taken to ease intel-

ligibility concerns.

I think both of these moves ultimately fail. But seeing what they do will help us understand the problem of worldly opaqueness better, and to see which restrictions a putative solution is subject to. My own stance on this issue will be that the problems are too serious to ignore and too hard for me to solve at the moment. For these reasons, the solution to the problem of counterpossibles I defend will preserve transparency.

5.3.1 Peer-based Intelligibility

Accusations of unintelligibility are hard to deal with methodologically. They cannot amount to a *direct* argument against the view: after all, we cannot argue against something we do not understand. Unintelligibility considerations also have a certain *ad hominem* flavor to them – if there are actual philosophers defending the views taken to be unintelligible, they are defending views which they do not understand. If you believe, as you should, that philosophers are not the kind of people that defend views they do not understand, at least not *en masse*, then you think that the fact that a reasonable number of philosophers defend a particular view is good evidence that the view is intelligible after all.

While this argument-form has its problems, I think that it is *prima facie* plausible. And Berto et al. use it to defend the phenomena of worldly opaqueness. Here is what they say:

There are hyperintensional contexts that are not in any way ‘about representational features’ (see Nolan 2014), and counterfactuals may well be among these. Hyperintensionality without appeal to representation is invoked in many discussions of metaphysical grounding; [...] The claim that hyperintensionality as such requires being about representational features would need serious support; and this Williamson does not offer. (Berto et al. 2017:9)

This reply rests on a confusion. What Williamson claims is unintelligible (or what I claim to be unintelligible on his behalf) is the phenomenon of worldly *opaqueness* not worldly *hyperintensionality*. Berto et al. seem to think that these phenomena coincide:

That is, we take it that counterfactuals create hyperintensional contexts, contexts in which substitutivity of identicals is not valid. (Berto et al. 2017:9)

Williamson (2017: 175) however, holds that this is “highly implausible”. The reason given there has two premises: that hyperintensionality occurs only in constructions that are “about representational features” (that is, constructions that are broadly epistemic or intentional, like ‘It is a priori...’ or ‘Alice believes...’); and that counterfactuals are not about representational features in this way. (Berto et al. 2017:9)

But these phenomena do not coincide. In fact, none of the examples reviewed in Nolan (2014: Section 4) as instances of worldly hyperintensionality violate the substitutivity of identicals.⁶

And metaphysical grounding is also generally taken to be transparent. Claiming that the consensus on the grounding literature is that grounding is opaque would definitely be false. There is, however, one proponent of such a view, Wilson (2014). Wilson has been criticized in Bennett (2017: 3.2.2). It will be instructive to sketch the general lines of that debate in order to understand better what is at stake in ours. But first let me present a relatively uncontentious case of hyperintensionality without opaqueness to conclude the point that Berto et al.'s appeal to existing literature is misguided.

The example is taken from Williamson (2017: 210ff). Consider the relation $A\#B$ in which two sentences stand if and only if they express the same russelian proposition. A *russelian proposition* is, roughly, a structured entity which is expressed by declarative sentences and whose constituents are the objects and relations referred to and whose structure roughly mimics the syntactic structure of the sentence.⁷ The sentence “ $2+2=4$ ” expresses the russelian proposition that can be represented by $\langle [=], ([+]([2], [2]), [4] \rangle$, which has as its constituents the numbers 2 and 4 and the relations *plus* and *identity*. “ $\#$ ”, as defined above, is hyperintensional. “ $2+2=4$ ” has the same intension as “ $2+3=5$ ”, true at all possible worlds. But they do not express the same russelian proposition - one has the number three as a constituent and the other does not. But $\#$ is transparent. If you take any russelian proposition and substitute one of its constituents for an identical constituent, you get the *same* russelian proposition. This much should be enough to highlight the illegitimate inference from hyperintensionality to opaqueness.⁸

Let us now turn to the debate between Wilson and Bennet in the arena of metaphysical grounding. The discussion concerns the relation of being “more fundamental than”. It seems clear that such a relation must be irreflexive, in much the same way as the relation of “being taller than” must be irreflexive. But Wilson (2014: 573) thinks that there can be cases where: 1- a is more fundamental than b ; and 2- $a = b$. An example would be the case of identity-based physicalism. Such a physicalism would say that a mental state is identical to a physical state. But any physicalism must claim that the physical is more fundamental than the mental.

If all of the above is true, the relation of being “more fundamental than” had better be opaque. Otherwise, the physical would be more fundamental than itself (and so would the mental). For this reason, Wilson (2014: 573) postulates opaqueness for the “more fundamental than” relation.

Karen Bennett's (2017: 43) reaction to Wilson is very much similar to the

⁶They are: explanation, essence, grounding, intrinsicity, property identities and dispositions.

⁷King (2007) is a very influential defense of this theory of propositions. Whether or not it is the *correct* theory of propositions is immaterial to the point I am making.

⁸But an instance of failure of substitutivity of identicals is, indeed, an instance of hyperintensionality, since two co-referring rigid designators have the same intension.

one Williamson has in relation to the non-vacuists. This should show that the issues we are dealing with in this chapter are general enough to be replicated in other debates in different areas.

Bennett’s main complaint is the following:

The claim that the predicate ‘more fundamental than’ creates an opaque context needs to be supplemented with a story about why it does so—a story, if you will, *about the metaphysical underpinnings of the semantic phenomenon*. (Bennett 2017: 44, my emphasis)

One explanation, which Bennett goes on to consider, is that representations are involved. But this cannot be the solution to this problem. The phenomenon of fundamentality is not a representational phenomenon. It has to do only with how things are in the world, not with the way we represent them to be. If the mental is identical with the physical then they are the same *thing*. Bennett goes on to conclude:

Perhaps she has in mind a different explanation of, or metaphysical underpinning for, the putative opacity. But in the absence of one, it is not promising or plausible to retain reflexive [instances of “more fundamental than”]. (Bennett 2017: 45-6)

The objection is that the opacity in question must have some kind of metaphysical underpinning which must not be representation-based. Just postulating opacity in a strictly worldly matter raises intelligibility worries. What is needed is a “story [...] about the metaphysical underpinnings of the semantic phenomenon”. And that is exactly what is required of the non-vacuist which thinks the counterfactuals are opaque.

In the next section, we will consider a way in which Berto et al. try to provide one and see why it should be found wanting.

5.3.2 Impossible Worlds Model

Another typical reaction when facing an unintelligibility accusation is to provide a model for the theory in question. What a good model provides is a guarantee of consistency. If something is guaranteed to be consistent, the case for its intelligibility improves immensely.⁹ The crux of the matter, often, is whether the model represents accurately the phenomenon for which it is intended to lend intelligibility to. Models are easy to come by. Only after a careful interpretation that maps the model to the phenomena under study can we be confident that the argument has any force.

Here is Berto et al’s statement of their view:

Then *a*’s not being *b* is a way things just cannot be: an impossibility. In particular, it can be so at an impossible world. There is nothing

⁹Incidentally, this shows that there seems to be link between consistency and intelligibility, something that matters to my arguments in favor of counterlogical vacuism from chapter 3.

particularly epistemic about this, any more than there is about a world which hosts a physical impossibility, such as (supposing Einstein was right) something accelerating through the speed of light. A world is partially characterised by a set of sentences. These tell you exactly what the world is like—whether it is possible or impossible. And if it be retorted that if $a = b$, and this statement really is about a and b , the failure of substitutivity would be impossible, the reply is ‘Of course’! (Berto et al. 2017: 9)

The thought seems to be this. We have a model as to how a worldly statement “ a is F ” can be opaque. That is, we have a model of how “ a is F ” and “ $a = b$ ” can be true while “ b is F ” is false. There is this worldly thing, an impossible world, which represents a as being F while representing b as not being F . Insofar as the model is consistent, this show that the unintelligibility claim is, to a certain degree, misplaced.

The real question is, of course, whether this explanation sheds any light on the “metaphysical underpinnings of the semantic phenomenon”, to use Bennett’s phrase. I do not think it does. To show why I think that, and because these issues are very complex and unclear, I will analyze first two examples of models that can be considered to be unsatisfactory for their own purposes and then argue that the same considerations can be used against Berto et al.’s model. The first is a very simple and silly toy example of my own. The second, is Akiba’s (2015) critique of Barnes and Williams’s (2011) theory of metaphysical indeterminacy.

To start with my toy example. Let us say we have a predicate F in a first order language for which Fa is true, $a = b$ and Fb is false. The way in which we know whether Fx is true for any term x is by assigning an extension to F , a set of objects such that F applies to them, and then checking whether the denotation of x is in that set. Since a and b denote the same object, they both are or both are not in the extension of F . A natural way to solve this problem is to enlarge the domain of the model to include *proxies* for the constant terms of the model. For any x , there is a corresponding proxy $P(x)$. And the reason why Fa is true is because $F(P(a))$ is true - because $P(a)$ is in the extension of F . And the reason why Fb is false is because $P(b)$ is not in the extension of F .

We have succeeded in altering the usual semantics of first order logic in such a way in which we can model the opacity of F . But it seems clear that we have not managed to present a good model of the *worldly* opacity of F . If F is worldly, Fa should be about what a is about, that is, the referent of a . The device of creating proxies cannot preserve this intuition under any natural interpretation – what can these proxies be if not representations of the objects they are proxy for? The model in question *models* but does not *explain* how F can be both worldly and opaque. If only solving the problem turned out to be so easy!

Now to the same problem in a much more complex setting. Barnes and Williams (2011) presented a model of metaphysical indeterminacy. The idea of there being indeterminacy in reality is notoriously hard to wrap one’s head

around. Lewis (1986: 213), to take one significant example, thinks such indeterminacy to be “unintelligible”. Providing a model for this phenomena seems to exactly parallel the dialectical move being made in the case of counterfactuals that is worrying us. They use ersatz possible worlds to provide the model. Akiba (2015) claims the model fails precisely because the representation these worlds provide cannot make the model explanatory enough to provide a satisfactory account of its target phenomenon.

Here is Barnes and Williams’ model in a nutshell, described in a passage in Barnes (2010: 613):

For the purposes here I take precisifications to be possible worlds – not just *like* possible worlds, they *are* possible worlds. The set of precisifications will be the set of possible worlds closest to the actual world. . . . Importantly, these must be ersatz possible worlds – abstract representations of ways things could be – for I will appeal to the distinction between the actual world and the actualized world that is familiar from ersatz theories of possible worlds but is missing from Lewisian concrete realism. The distinction arises for the ersatzist precisely because possible worlds are abstract representations, and yet the world that we are literally a part of is not an abstract representation – it is a concrete individual. So there is the actual world – a mereological sum of concrete objects – that is not one of the possible worlds. And there is the actualized world – the abstract world that represents things as being as they are as opposed to representing things as being as they are not. My proposal is this: that every possible world is fully precise, but that if there is ontological indeterminacy it is indeterminate which of the possible worlds is the actualized world – that is, it is indeterminate which world, out of the many worlds that represent things to be a precise way, is the one that represents the way the actual world is. (Barnes 2010: 613)

The idea is that there can be metaphysical indeterminacy when it is indeterminate which of various possible worlds, taken as representational entities, effectively represents the actual world. Their theory needs to use ersatz worlds because it depends on there being a difference between the actual world, the concrete thing where you and I live in, and the actualized world, a mere representation of this wonderful place.

Predictably, they come under criticism for this aspect of their view: how can truths about representations be an adequate model for a phenomenon that is strictly metaphysical in essence? Here is Akiba’s complaint:

We gain nothing by saying, instead of simply saying that ‘*p*’ is true in [the actual world], that a ‘*p*’ world is the actualized world. [...] The reason seems to be in the two quite different notions of possible world we’ve distinguished and, consequently, the two notions of precisification. [O]n the semantic notion of possible world, possible worlds are nothing but abstract mathematical entities such as

sets of interpretations (ultimately, sets). On the ontic, or realistic, notion, possible worlds are worlds that contain concrete, often material, objects, including chairs, tables, humans and events. These are completely different notions. When Barnes and Williams adopt ersatzism, they are affected by this possible source of equivocation. Ersatz possible worlds are exactly what the name says: they are not possible worlds of the second, realistic, kind; they are mere substitutes of the genuine (i.e. realistic) things – specifically, they are sets of sentences or representations. (Akiba 2015:573)

Akiba’s (2004) own theory uses concrete, Lewisian possible worlds. These provide the kind of representation that is adequate for a strictly worldly matter. Though Barnes (2010: 617) criticizes this theory, its fault is not that representations play a central role in the explanation of a metaphysical phenomenon.

After this long digression, we come back to Berto et al.’s model that is our central concern. They think that an impossible worlds model can be used to explain worldly opacity in quite a simple way. There is a world, though it obviously has to be an impossible one, that represents an instance of worldly opacity as being the case. It seems to me that their answer is unsatisfactory for the same reasons that my silly toy model and Barnes and Williams’ model for metaphysical indeterminacy are unsatisfactory. The truth of worldly sentences will be made, by the model, dependent of representational matters. No illumination of the “metaphysical underpinnings of the semantic phenomenon” is achieved in this way.

The first point to be emphasized is that what impossible worlds are, and the way they represent propositions as being true, must matter. It will not be enough to remain agnostic on this matter, as Berto et al. do. As we have seen, it is not so easy to have impossible worlds where Hesperus and Phosphorus come apart in the context of linguistic ersatzism. The simple lagadonian strategy that stipulates that each object is to stand for its own name will not work. Consider Jago’s (2014) alternative, using the DTHAT operator. The strategy is based on the fact that different descriptions can successfully pick out the same object in the actual world. Once we rigidify these descriptions, we obtain a designator for the object in every possible world. But they can come apart at impossible worlds. This will plausibly make the property “Actually seen in the evening” part of the DTHAT description of Hesperus and the property “Actually seen in the morning” part of the DTHAT description of Phosphorus. This makes the fact that “Hesperus is bright” at an impossible world dependent on something other than the object Hesperus – it also depends on the interaction of actual representational agents with Hesperus.

This is not an argument against all possible ontologies of impossible worlds that can be deployed to provide the model Berto et al. need. The burden is on them to provide such an account. However, I want to point out a significant restriction on this project: the so-called *exportation problem*.¹⁰

¹⁰See, for example, Jago (2014: Section 4.3).

Akiba (2004; 2015) could use a concrete conception of possible worlds to address his own concern with Barnes and Williams' model. The way that Lewisian possible worlds represent satisfied the requirement for a metaphysically explanatory notion of representation, given Akiba's standards. The impossible worlds theorist cannot mirror such a move. There cannot be Lewisian impossible worlds.

The reason is that if an object is represented as existing by a Lewisian world then that object exists *simpliciter*. The way a Lewisian world represents an object as existing is precisely by having such an object as a mereological part. So, if a Lewisian impossible world represents a round square as existing then a round square exists. But round squares cannot exist - that is why they are impossible objects. Therefore, neither can Lewisian impossible worlds.

The key to this argument is the inference from "*o* exists at world *w*" to "*o* exists". We can call such an inference *exportation*. An ontology of possible worlds can live with exportation. One might not like that non-actual objects are taken to exist in the very same metaphysical sense as actual objects do, but there is no incoherence in that position. In contrast, an ontology of impossible worlds where exportation holds simply does not work. These impossible worlds are impossible objects. They inherit their impossibility from the impossible objects they represent as existing.

The challenge is to provide an ontology of impossible worlds whose representation relation explains worldly opacity and which, at the same time, avoids the exportation problem. I do not know if the challenge can be met but, if it can, I have no idea how. I conclude that Berto et al. have not provided us with a satisfactory model of worldly opacity and that such a model will not be easy to come by.

5.4 Counterfactuals are Transparent

I will use no impossible worlds where actual identicals come apart. Substitutivity of identicals is valid by design. If *a* and *b* are two rigid designators which designate the same object, then an impossible world represents that "*a* is *F*" if and only if it represents that "*b* is *F*".

There is some motivation for this in that it eases the construction of the impossible worlds needed by my theory (as explained in chapter 2). But the main motivation is that I think that if counterfactuals are worldly (something I do not wish to dispute in this context) then substitution of identicals is valid. And this restriction is the most straightforward way of doing so in this context.

As is always the case when a class of impossible worlds is thrown out of our semantics for counterfactuals, this restriction will carry with it a certain form of vacuism. In this case, every antecedent which involves an impossibility based on the necessity of identity is vacuously true.

My strategy for dealing with this vacuism should be familiar by now.

First of all, sometimes what we mean by an antecedent with the form "If Hesperus hadn't been Phosphorus" is really a proposition having to do with the

reference of the names “Hesperus” and “Phosphorus” - it is a counterfactual about representations. And it is entirely possible that Hesperus and Phosphorus designated different things – at other possible worlds, whatever metasemantic facts fix the semantic value of “Hesperus” and “Phosphorus” might be different in ways that change their reference.

As it could be expected, this strategy will not be able to cover all cases. Sometimes someone can actually wish to counterfactually suppose that two identicals are distinct. What I want to claim here is that that scenario is very often unentertainable. And unentertainability is a good justification for vacuism.¹¹

It seems to me that this is an even more widespread phenomenon here than in the case of counterlogicals. We just have no idea about what would be the case if a thing somehow was multiplied into two things which are not identical with each other in a counterfactual scenario. In fact, I find it quite hard to think of an ordinary context where this would be entertainable.¹²

I think that living with this form of vacuism is a much smaller theoretical problem than explaining how an worldly construction can be opaque. It seems to me that the semantics of counterfactuals should either be transparent or some theory in which they are partly about representations should be developed. Since the participants in this debate assume that the latter is not the case, I think they should go with the former option.

In any case, the very conceptual possibility of this option shows that non-vacuism is not tied to a denial of the substitutivity of identicals, or to the project of defending the intelligibility of worldly opacity. One can very well be a non-vacuist and still think that counterfactuals are transparent.

¹¹This strategy was discussed in the first chapter and used to size down the problem of counterlogical vacuism in the third chapter.

¹²As Kripke (1980) points out, we can *think* we are entertaining this impossibility when in fact we are entertaining the possible shift in semantic meaning of the rigid designators we referred to above.

Chapter 6

Conclusion

In this thesis, we dealt with the problem of counterpossibles. The Lewis-Stalnaker semantics, the standard semantics for counterfactuals, makes all counterfactuals with a metaphysically impossible antecedent vacuously true. This goes against our intuitions - some counterpossibles seem false. But there are also good theoretical reasons for being a vacuist.

We had the opportunity to clarify the debate in various ways. I argued that the current discussion implicitly assumes semantic consequentialism - a methodological picture where a semantic theory can trade off adequacy with speakers' intuitions for other theoretical gains. This is the only framework where the debate between vacuists and non-vacuists makes sense.

We then saw which costs the vacuist takes the non-vacuist to be committed to. The vacuist thinks that non-vacuism implies using a dubious tool for semantic theorizing - impossible worlds -, accepting a very weak logic of counterfactuals, worsening our theoretical understanding of metaphysical modality and predicting that counterfactuals violate the substitutivity of identicals.

I hope to have clarified the nature of these challenges. Furthermore, I tried to show how the existing literature fails to address them successfully.

Finally, I sketched a non-vacuist position which trades off a better solution to the challenges facing non-vacuism for certain limited forms of vacuism. I think a moderate form of non-vacuism like this can turn out to be the best reaction to the problem of counterpossibles. But, more importantly, it shows that non-vacuism by itself is not invariably committed to any of the costs pointed out by vacuists. This allows for a richer understanding of the conceptual space of answers to the problem of counterpossibles.

Bibliography

- Akiba, K. (2004), “Vagueness in the world”, *Nous* 38 (3):407–429.
- Akiba, K. (2015), “How Barnes and Williams have failed to present an intelligible ontic theory of vagueness”, *Analysis* 75 (4):565-573.
- Anderson, A.R. (1951), “A note on subjunctive and counterfactual conditionals,” *Analysis* 12: 35–8.
- Balcerak Jackson, M. (2016), “On the Epistemic Value of Imagining, Supposing and Conceiving”, in Kind, A. and Kung, P. (eds.), *Knowledge Through Imagination*, Oxford University Press, 41-61.
- Barnes, E. (2010), “Ontic Vagueness: A Guide for the Perplexed”, *Nous* 44 (4):601-627.
- Barnes, E. and Williams, R. (2011), “A Theory of Metaphysical Indeterminacy”, in Karen Bennett & Dean W. Zimmerman (eds.), *Oxford Studies in Metaphysics volume 6*, Oxford University Press, 103-148.
- Barwise, J. and Perry, J. (1981), “Situations and attitudes”, *Journal of Philosophy* 78 (11):668-691.
- Bealer, G. (2002), “Modal Epistemology and the Rationalist Renaissance”, in Gendler and Hawthorne (eds.), *Conceivability and Possibility*, Oxford University Press, 71-126.
- Beall, J.C. and Restall, G. (2005), *Logical Pluralism*, Oxford University Press.
- Bennett, K. (2017), *Making Things Up*, Oxford University Press.
- Berto, F. (2010), “Impossible worlds and propositions: Against the parity thesis”, *Philosophical Quarterly* 60 (240):471-486.
- Berto, F. (2013), “Impossible worlds”, *Stanford Encyclopedia of Philosophy*.

- Berto, F., French, R., Priest, G. and Ripley, D. (2017), “Williamson on Counterpossibles”, forthcoming in *Journal of Philosophical Logic*, 1-21.
- Boghossian, P. (1997), ”Analyticity”, in B. Hale and C. Wright (eds.), *A Companion to the Philosophy of Language*, Blackwell, 331-368.
- Brogaard, B. and Salerno, J. (2013), “Remarks on Counterpossibles”, *Synthese* 190 (4):639-660.
- Byrne, R. (2016), “Counterfactual Thought”, *Annual Review of Psychology* 67: 135-157.
- Cameron, R. (2009), “What’s Metaphysical about Metaphysical Necessity?”, *Philosophy and Phenomenological Research* 79 (1):1 - 16.
- Chalmers, D. (2002), “Does Conceivability Entail Possibility?”, in Gendler and Hawthorne (eds.), *Conceivability and Possibility*, Oxford University Press, 145-200.
- Clarke-Doane, J. (2017), “Modal Objectivity”, forthcoming in *Nous*.
- Cumming, S. (2013), “Names”, *Stanford Encyclopedia of Philosophy*.
- Dever, J. (2012), “Compositionality”, in Fara and Russell (eds.), *The Routledge Handbook to the Philosophy of Language*, 91-102.
- Divers, J. and Melia, J. (2002), “The analytic limit of genuine modal realism”, *Mind* 111 (441):15-36
- Donellan, K. (1966), “Reference and definite descriptions”, *Philosophical Review* 75: 281-304.
- Dorr, C. (2002), *The Simplicity of Everything*, PhD Dissertation, Princeton University.
- Fine, K. (1975), “Critical Notice of Lewis, Counterfactuals”, *Mind* 84 (335):451 - 458
- Fine, K. (1994), “Essence and Modality”, *Philosophical Perspectives* 8:1-16
- Fine, K. (2012), “Counterfactuals Without Possible Worlds”, *Journal of Philosophy* 109 (3):221-246
- Fine, K. (2018), “Constructing the Impossible”, Manuscript in https://www.academia.edu/11339241/Constructing_the_Impossible

- Friedman, J. (2017), “Why Suspend Judging?”, *Nous* 51 (2):302-326.
- Gibbard, A. (1981), “Two Recent Theories of Conditionals”, in William Harper, Robert C. Stalnaker & Glenn Pearce (eds.), *Ifs*, Reidel, 211-247.
- Gillies, A. (2007), “Counterfactual Scorekeeping”, *Linguistics and Philosophy* 30: 329–360.
- Glanzberg, M. (2013), “Truth”, *Stanford Encyclopedia of Philosophy*.
- Goodman, N. (1955), *Fact, Fiction and Forecast*, Cambridge, MA: Harvard University Press.
- Grice, H. (1989), *Studies in the Ways of Words*, Harvard University Press.
- Hawke, P. (2017), “Theories of Aboutness”, forthcoming in *Australasian Journal of Philosophy*.
- Hawthorne, J. and Gendler, T. (2002), *Conceivability and Possibility*, Oxford University Press.
- Iatridou, S. (2000), “The grammatical ingredients of counterfactuality”, *Linguistic Inquiry* 31: 231–270.
- Jago, M. (2014), *The Impossible: An Essay in Hyperintensionality*, Oxford University Press.
- Jago, M. (2015), “Hyperintensional propositions”, *Synthese* 192 (3):585-601.
- Jenkins, C. (2008), “Modal Knowledge, Counterfactual Knowledge and the Role of Experience”, *The Philosophical Quarterly* 58: 693-701.
- Kaplan, D. (1978), “Dthat”, In Peter Cole (ed.), *Syntax and Semantics*, Academic Press, 221-243.
- King, J. (2007), *The Nature and Structure of Content*, Oxford University Press.
- Kment, B. (2006), “Counterfactuals and Explanation”, *Mind* 115 (458):261-310.
- Kment, B. (2014), *Modality and Explanatory Reasoning*, Oxford University Press.
- Kripke, S. (1980), *Naming and Necessity*, Harvard University Press.
- Kroedel, T. (2012), “Counterfactuals and the epistemology of modality”, *Philosophers’ Imprint* 12: 1–14.

- Lewis, D. (1973), *Counterfactuals*, Oxford: Blackwell.
- Lewis, D. (1979), “Counterfactual dependence and time’s arrow”, *Nous* 13 (4):455-476.
- Lewis, D. (1986), *On the Plurality of Worlds*, Oxford: Blackwell.
- Lycan, W. (2001), *Real Conditionals*, Oxford University Press.
- McGrath, M. and Frank, D. (2018), “Propositions”, *Stanford Encyclopedia of Philosophy*.
- Menzel, C. (2016), “Possible Worlds”, *Stanford Encyclopedia of Philosophy*.
- Nolan, D. (1997), “Impossible Worlds: A Modest Approach”, *Notre Dame Journal of Formal Logic* 38 (4):525-572.
- Nolan, D. (2014), “Hyperintensional Metaphysics”, *Philosophical Studies* 171 (1):149-160.
- Nolan, D. (2016), “Modal Fictionalism”, *Stanford Encyclopedia of Philosophy*.
- Nolan, D. (2017), “Causal Counterfactuals and Impossible Worlds”, in Helen Beebe, Christopher Hitchcock & Huw Price (eds.), *Making a Difference*, Oxford University Press.
- Peters, J. and Westerstahl, D. (2006), *Quantifiers in Language and Logic*, Clarendon Press.
- Plantinga, A. (1974), *The Nature of Necessity*, Clarendon Press.
- Priest, G. (1997), “Sylvan’s Box”, *Notre Dame Journal of Formal Logic* 38 (4):573-582.
- Priest, G. (2008), *An Introduction to Non-classical Logic*, Cambridge University Press.
- Pryor, J. (2000), “The Skeptic and the Dogmatist”, *Nous* 34 (4):517–549.
- Quine, W. (1955), “Three Grades of Modal Involvement”, *Journal of Symbolic Logic* 20 (2):168-169.
- Quine, W. (1970), *Philosophy of Logic*, Englewood Cliffs: Prentice Hall.

- Ripley, D. (2011). "Negation, Denial, and Rejection", *Philosophy Compass* 6(9): 622-629.
- Rosen, G. (2006), "The Limits of Contingency", in Fraser MacBride (ed.), *Identity and Modality*, Oxford University Press, 13-39.
- Schulz, K. (2014). 'Fake Tense in Conditional Sentences. A Modal Approach', *Natural Language Semantics*, 22(2): 117-144.
- Shoemaker, S. (1998), "Causal and metaphysical necessity", *Pacific Philosophical Quarterly* 79 (1):59-77.
- Sider, T. (2011), *Writing the Book of the World*, Oxford University Press.
- Stalnaker, R. (1968), "A Theory of Conditionals", in Nicholas Rescher (ed.), *Studies in Logical Theory*, Volume II, American Philosophical Quarterly Monograph Series, 98-112, Blackwell.
- Stalnaker, R. (1975), "Indicative Conditionals", *Philosophia* 5: 269-86.
- Stalnaker, R. (1976), "Possible Worlds", *Nous* 10 (1):65-75.
- Vander Laan, D. (2004), "Counterpossibles and Similarities", in Frank Jackson & Graham Priest (eds.), *Lewisian Themes: The Philosophy of David K. Lewis*, Oxford, UK: Clarendon Press, 258-275.
- von Fintel, K. (2001), "Counterfactuals in a Dynamic Context", in *Ken Hale: A Life in Language*, chap. 3, MIT Press.
- Warren, J. (2018), "Change of Logic, Change of Meaning", *Philosophy and Phenomenological Research* 96 (2): 421-442.
- Weatherson, B. (2001), "Indicatives and Subjunctives", *Philosophical Quarterly* 51 (203):200-216.
- Weatherson, B. (2002), "Real Conditionals: Book Review", *Philosophical Review* 111 (4):609-611.
- Williamson, T. (2007), *The Philosophy of Philosophy*, Wiley-Blackwell.
- Williamson, T. (2013), *Modal Logic as Metaphysics*, Oxford University Press.
- Williamson, T. (2016), "Modal Science", *Canadian Journal of Philosophy* 46 (4-5):453-492.

Williamson, T. (2017), “Counterpossibles in Semantics and Metaphysics”, *Argumenta* 2 (2):195-226.

Wilsch, T. (2017), “Sophisticated Modal Primitivism”, *Philosophical Issues* 27 (1):428-448.

Wilson, J. (2014), “No work for a theory of Grounding”, *Inquiry : An Interdisciplinary Journal of Philosophy* 57 (5-6):535-579.

Winter, Y. (2016), *Elements of Formal Semantics*, Edinburgh University Press.

Yablo, S. (2014), *Aboutness*, Princeton University Press.