UNIVERSITÀ DEGLI STUDI DEL PIEMONTE ORIENTALE
"AMEDEO AVOGADRO"

DIPARTIMENTO DI STUDI UMANISTICI

Corso di Dottorato di Ricerca in Filosofia

Ciclo XXXVII

Titolo tesi

Worldly Hyperintensionality
Propositions, Counterfactuals and Truthmakers

SSD M-FIL/05

Tutor
Prof. Matteo Plebani

Dottorando
Giorgio Lenta

Coordinatore
Prof. Luca Fonnesu

Worldly Hyperintensionality:

Propositions, Counterfactuals, and Truthmakers

Giorgio Lenta

2025

UNIVERSITÀ DEGLI STUDI DI GENOVA
UNIVERSITÀ DEGLI STUDI DI TORINO
UPO
UNIVERSITÀ DEGLI STUDI DI PAVIA

Worldly Hyperintensionality: Propositions, Counterfactuals, and Truthmakers

Giorgio Lenta

Università degli Studi di Genova - Consorzio FINO

Supervisor:
Prof. Matteo Plebani

Reviewers:
Prof. Francesco Berto
Prof. Johannes Korbmacher

Contents

Chapter 1

Introduction

A wide variety of philosophical disciplines are currently witnessing what Nolan (2014) called a "hyperintensional revolution". As noticed by Berto and Nolan (2023), the primary drive for this change of paradigm can be identified with the progressive recognition of some important limitations in the standard intensional models, which have been massively employed to account for several notions at the heart of our philosophical theorizing.

Among these notions, we find the subject matters of this dissertation. More precisely, the present work explores aspects of two progressively narrower instances of non-representational hyperintensionality: propositions and counterfactual conditionals. Thus, introducing the content of this work requires some preliminary clarification about the general notion of hyperintensionality, and its specific variant that is most relevant for the upcoming discussion.

The earliest use of the concept of hyperintensionality is Cresswell (1975)'s notion of a hyperintensional *context*: a place in a sentence where substitutivity *salva veritate* of (classical) logical equivalents fails. Consider two logical equivalents $\neg p$ and $\neg\neg\neg p$, and suppose that Anna believes only the former (maybe she does not realize that they are equivalent because she has troubles in parsing the syntax of expressions with multiple negations). If we replace the former with the latter in a true ascription like "Anna believes that $\neg p$", we obtain a *prima facie* false sentence. Therefore, propositional attitudes like belief are widely recognized to create hyperintensional contexts at least in Cresswell's sense.

However, hyperintensionality in its contemporary use typically encompasses a broader notion of *necessary* equivalence, which usually includes logical, mathematical and metaphysical equivalence. Indeed, we can illustrate failures of substitutivity *salva veritate* within the same belief context of the example above, where instead of substituting two logical equivalents we substitute two *metaphysically* necessary equivalents, such as "Hesperus is Hesperus" and "Hesperus is Phosphorus". Again, Anna may very well believe the former without believing the latter, despite their being both true as a matter of metaphysical necessity.

In the present work, I will adopt an analogous definition of hyperintensionality that

is explicitly formulated against the background of the standard intensional framework, namely, Possible Worlds Semantics (PWS). Specifically, I will stick to the definition of a hyperintensional context as one in which substitutivity *salva veritate* of *intensional* equivalents fails. Obviously, this will make the notions of hyperintensional *concepts* or *operators* elliptical for concepts or operators that create hyperintensional contexts.

This definition has the double advantage of helping to establish formally precise boundaries for its application and to better illustrate the above-mentioned limitations of standard models: in order to know whether a context is hyperintensional, we just need a formally precise definition of an intension for each relevant type of linguistic expression. Luckily, PWS comes in handy for this purpose, since we can standardly define the intension of each expression in a PWS model as its extension across possible worlds. More precisely, the intension of a singular term is a function from possible worlds to individuals; the intension of a predicate is a function from possible worlds to sets of individuals; and the intension of a sentence is a function from possible worlds to truth values (or more succinctly, the set of possible worlds at which the sentence is true).

Recently, a different way to spell out and deal with the problem of hyperintensionality emerged. While it is standardly understood in terms of failures of substitutivity within certain sentential contexts, so that, strictly speaking, hyperintensionality is a property of such contexts only, one may wonder how a *logical* notion of hyperintensionality should look like. In other words, one may wonder how to characterize a hyperintensional *logical consequence* relation. This started a new debate that has been the focus of Leitgeb (2018), Odintsov and Wansing (2021) and Bonaguidi (2022), and that established a distinction between "weak" and "strong" hyperintensionality.

In the present work, I will stick to the standard conception of hyperintensionality for sentential contexts, and leave related questions about logical consequence aside. Thus, I will not differentiate between weak and strong hyperintensionality. There is, however, another distinction in kinds of hyperintensionality that is particularly relevant for the present work: the one between *representational* and *worldly* hyperintensionality.

To briefly illustrate the distinction, we can go back to our previous examples. Regardless of the scope of a given definition of hyperintensionality (i.e. regardless of what kind of equivalence we care about), we observed that the hyperintensional contexts mentioned above involve propositional attitudes. These are standardly understood as *representational* contexts, since they exist in virtue of ascriptions of intentional *mental* states such as belief, knowledge, supposition, and so on. In light of this, we can characterize a notion of representational hyperintensionality: failure of substitutivity *salva veritate* of intensional equivalents that has its source in features of representation.

Can there be other, non-representational sources (and thereby non-representational *kinds*) of hyperintensionality? Obviously, if the answer were no, the "representational" label would be redundant: hyperintensionality would always derive from features of representation. However, recent developments in formal semantics and metaphysics suggest quite the opposite: there seem to be several cases of hyperintensionality that do

not depend on representations at all.[1]

In other words, as extensively discussed by Nolan (2014), the progressive recognition of the need for fine-grained content identification in analyzing many key concepts (even in the absence of propositional attitudes) provides solid ground for defending a notion of *worldly* hyperintensionality: failure of substitutivity *salva veritate* of intensional equivalents that has its source in features of the world.

The debate around the sources of hyperintensionality is also the subject of Chapter 2 of the present dissertation, in which I illustrate and defend the distinction between worldly and representational hyperintensionality through concrete examples and the criticism of a conceptualist reaction to it due to Byrne and Thompson (2019). Moreover, I provide a new framing of the issue through the comparison with a similar distinction between representational and worldly *intensionality*. The upshot of the chapter can be summarized as follows: establishing whether or not a putative case of hyperintensionality is representational should not be affected by the prior adoption of a specific interpretation of the background semantic framework, and a more data-driven approach to formal semantics is to be preferred. An earlier version of this chapter has been published in Lenta (2023).

Chapter 3 explores what I deem to be the broadest cases of non-representational hyperintensionality: semantic content and propositions (non-representational here simply means that alleged distinctions in meaning revealed only by propositional attitude reports and other distinctly representational ascriptions are not considered relevant). The discussion is centered around one of the most developed truth-conditional accounts of hyperintensional content: the compositional Impossible Worlds Semantics of Berto and Jago (2019). This is shown to be either semantically inadequate (it delivers a too fine-grained picture of content), or not compositional (at least not in the way intended by the authors), alongside further more general issues.

Chapter 4 explores some issues surrounding propositional quantification. This is taken to be a genuine linguistic phenomenon that should not need a "special" treatment in the formal setting (against a somewhat popular view within the program of higher-order logic), in line with the data-driven approach of the dissertation. However, this choice leads to some well-known paradoxes affecting some of the most popular accounts of hyperintensional propositions. In particular, I show how a variant of the Russell-Myhill paradox affects the Impossible Worlds Semantics discussed in Chapter 3. Then, I move on to the discussion of Kaplan's paradox and its hyperintensional variants. I argue that the best way to avoid this kind of paradoxes without giving up on some key formal notion (or on hyperintensionality) is to adopt a Truthmaker account of propositions. An earlier version of part of this chapter (Sections 4.3 – 4.7) has been published in Lenta (2024).

Chapter 5 restricts the focus of the dissertation on a specific class of hyperintensional

1. Hyperintensional resources are nowadays employed to account for a large variety of non-representational notions and aspects of reality such as essences, properties, dispositions, metaphysical grounding, counterfactual dependence and causation, just to mention a few. To borrow Berto and Nolan (2023)'s words: «almost any area of metaphysics where necessity or counterfactuals have played a role is a candidate for hyperintensional approaches».

propositions: counterfactual conditionals. I first highlight how the debate on hyperintensionality in counterfactuals has been developed along two distinct axes: the *counterpossible non-vacuism* axis (many counterfactuals with impossible antecedents seem false, despite the fact that standard intensional semantics makes all of them *vacuously* true); and the *formal invalidities* axis (some inferences involving counterfactuals that are generally valid according to standard intensional semantics must be rejected). The former axis is explored in the rest of Chapter 5, by considering the debate on strong emergence in the metaphysics of chemistry as a case study. I argue that, if the thesis of strong emergence is true, then some counterpossibles must be false (and others non-vacuously true), under few reasonable assumptions. This would provide further independent motivation for non-vacuism, along the lines of recent literature about the use of non-vacuous counterpossibles in certain theoretical contexts. The chapter is partly based on a short paper presented at the first Metaphysics of Chemistry Workshop, which was held at UCLouvain in July 2024. The paper is ought to be included in a forthcoming edited volume on the metaphysics of chemistry.

Finally, Chapter 6 explores the second of the previously mentioned axes: hyperintensionality in counterfactuals motivated by formal invalidities. In particular, the chapter deals with the complexities surrounding the principle of Simplification of Disjunctive Antecedents and related issues with determinable antecedents. Here, the discussion of Kit Fine's Truthmaker Semantics becomes the starting point for a new combined account that incorporates some features of the standard Stalnaker-Lewis approaches. The resulting framework gets the best of both worlds and satisfies all of our theoretical *desiderata*: it complies with the linguistic evidence, delivers a hyperintensional picture of counterfactuals and gives rise to an adequate logic. The account is further refined in the final part of the chapter, where I show how it can be modified to deliver non-vacuous truth-conditions for counterpossibles, and how it can be "exactified" to allow for even greater expressive power. The chapter is the result of a collaboration with Johannes Korbmacher.

Chapter 2

Sources of Hyperintensionality

2.1 Background

Hyperintensionality – the failure of substitutivity *salva veritate* between intensionally equivalent expressions – is nowadays one of the most debated topics in logic and philosophy of language.[1] Despite its being a feature of a wide variety of concepts and sentential contexts, most of the literature on this phenomenon has been focusing primarily on its *representational* instances: as widely recognized, hyperintensional contexts are paradigmatically triggered by propositional attitudes such as belief, knowledge, supposition, and so on. Thus, hyperintensionality is sometimes characterized as a phenomenon due to some limitations in our cognitive or conceptual faculties.[2]

Recently, the idea of hyperintensionality as an entirely representational phenomenon has been challenged. According to Daniel Nolan, hyperintensionality is often a genuine worldly matter that does not depend on our representations at all. More precisely, Nolan (2014) argues for extending the attribution of hyperintensional features to phenomena outside the realm of representation, introducing a notion of *worldly* hyperintensionality.

To clarify this point, Nolan suggests that a good way of classifying phenomena amounts to detecting a principled need for certain kinds of expressive resources to explain them. For instance, alethic modalities systematically require *intensional* language to be accounted for, whereas propositional attitude typically need something more fine-grained. In light of this, Nolan proposes a categorical shift from the language to the phenomena that the language captures. We get a rough picture of his classification from the following passage (p. 152):

> Running might then count as an extensional phenomenon, if we think an account of running can be adequate if it only mentions actual arrangements of legs and bodies.

1. Two sub-sentential expressions are intensionally equivalent if and only if they have the same extension at all possible worlds, and two sentences are intensionally equivalent if and only if they are true at the same possible worlds. A concept is said to be hyperintensional when it creates hyperintensional contexts, namely, places in a sentence where substitutivity *salva veritate* of co-intensional expressions fails.

2. See for instance Stalnaker (1984), Byrne and Thompson (2019).

> Necessity would then naturally be classified as an intensional phenomenon, needing expressions like 'necessarily', that create intensional positions, to capture it. Belief might naturally be classified as a hyperintensional phenomenon, needing an expression like 'believes that . . . ' to adequately capture it. Most theories of phenomena will consist of a variety of sentences, so if we want to assign each phenomenon one label from 'extensional', 'intensional' and 'hyperintensional', it might be best to assign 'hyperintensional' if hyperintensional language is needed to capture a phenomenon, 'intensional' if intensional but not hyperintensional language is needed, and 'extensional' only if no intensional nor hyperintensional language is needed.

As Nolan himself notices, this kind of classification might raise some doubts. One might believe that an allegedly extensional phenomenon like running is actually intensional, or even hyperintensional, on a more careful reading. On the other hand, one might stress that necessity can and should be accommodated in a purely extensional way, for instance by means of quantification over possible worlds. Nonetheless, Nolan believes that even if a fully extensional account for intensional or hyperintensional phenomena was actually available, this would not threaten the idea that the accounted phenomena are *themselves* intensional or hyperintensional under his interpretation; just like the possibility to offer a theory of possible worlds in an extensional language does not undermine the idea that possible worlds are themselves intensional entities. However, as we shall see in more detail, one could still reject the existence of worldly hyperintensionality regardless of Nolan's classification. In other words, one might still deny that hyperintensionality can have a non-representational *source*, no matter what kind of language (or semantic framework) we employ to talk about or analyze certain phenomena.

The aim of the present chapter is to defend a genuine distinction between worldly and representational hyperintensionality, and to show that the issue of its source, when properly framed, should not be affected by preferences towards specific semantic frameworks. Therefore, rejecting worldly hyperintensionality merely on the basis of such preferences is not a viable strategy. Moreover, the present chapter aims to provide reasons for skepticism towards a generalized conceptualist picture for all (or even most) kinds of hyperintensionality.

In order to achieve this, I will first provide a new framing of the notion of worldly hyperintensionality through the comparison with that of *worldly intensionality* (Section 2.2); then, I will present one of the main arguments against worldly hyperintensionality in the existing literature, developed by Byrne and Thompson (2019) (Section 2.3); finally, I will offer two arguments against Byrne and Thompson's view (Section 2.4). As a result, I will conclude that the comparison with intensional phenomena – upon which Byrne and Thompson also rely – actually supports the existence of a genuine distinction between worldly and representational hyperintensionality.

2.2 Intensionality, Hyperintensionality and their Sources

The need for semantic resources capable of drawing distinctions among intensionally equivalent contents is widely recognized.[3] However, the view that hyperintensionality can derive directly from the world, rather than merely from our representations of it, is more controversial. According to Nolan, the difficulty in accepting this idea has its roots in the common features of the most discussed cases of hyperintensionality (those involving propositional attitudes), which can always be traced back to some limitations in our cognitive or conceptual faculties.

If propositional attitudes exhausted the whole range of hyperintensional phenomena, that would be enough to conclude that hyperintensionality is always a representational matter. In such a scenario, having hyperintensional contexts in our natural languages could indeed be explained with our inability to have perfect access to all of the intensions associated with each meaningful expression and its corresponding representational content. Another reason why someone might be tempted to regard hyperintensionality as merely representational is that her preferred hyperintensional semantics *identifies* contents with representational entities. This view has been advocated by Byrne and Thompson (2019), as we shall see in the next section. But first, further clarification on the issue is required.

Nowadays, it is widely accepted that the cases of hyperintensionality triggered by representational attitudes are just a subset of the whole range of hyperintensional phenomena. Following Nolan (2014, p. 158):

> if we find ourselves appealing to hyperintensional distinctions in our theory of non-representational matters, that would be good evidence that hyperintensional phenomena are not just representational.

For instance, metaphysics ordinarily deals with notions like grounding, essence, properties and dispositions, all of which seem to require hyperintensional resources to be accounted for.[4] Since metaphysics is supposed to be the inquiry on the fundamental structure and nature of reality, and the subject matters of metaphysical theories are typically taken to be non-representational, showing that some fundamental feature of reality requires hyperintensional explanations – for instance in the form of models employing fine-grained entities – hints to the conclusion that hyperintensionality can itself be non-representational, by having its source in some independent feature of the world. Moreover, a number of philosophers have recently argued that hyperintensional explanations often show up even in natural sciences, where subject matters are most clearly non-representational.[5]

Still, such a shift from the language to the world must be approached with some caution. The above-mentioned cases share the features of being *about*

3. See Berto and Nolan (2023) for an overview of cases where hyperintensional semantic resources are needed.

4. See Nolan (2014) for an in-depth discussion.

5. See, for instance, Jenny (2018), Tan (2019) and Wilson (2021).

non-representational subject matters and of requiring hyperintensional resources to be accounted for. But this might not suffice for concluding that, in such cases, hyperintensionality arises from the world rather than from our representations of it. After all, hyperintensionality has been characterized in linguistic terms as the failure of substitutivity *salva veritate* between co-intensional (linguistic) expressions. Moreover, one might argue that theories can themselves be characterized as linguistic items, regardless of the (possibly) non-representational nature of their subject matters. Therefore, one might find it reasonable to consider theories as representational objects and, in turn, hyperintensionality as a representational feature of some theories. But then, how are we supposed to flesh out the idea that some instances of hyperintensionality have their source outside of the (merely representational) linguistic realm?

To develop an answer, we can start by drawing a comparison with a similar issue concerning intensionality. It is widely known that modal expressions can be often read according to either the *de dicto* interpretation or the *de re* interpretation. In light of this, we may ask: what is the extra-linguistic ground that guides the choice of a certain reading as the correct one? Consider how we ordinarily deal with ambiguous cases of alethic modal talk:

(1) The number of my fingers is necessarily even.

Let x range over the natural numbers and N stand for a magnitude representing "the number of". We can read (1) *de dicto* as:

(1d) $\quad \Box\exists x(N(myfingers) = x \wedge even(x))$

According to (1d), it is necessary that I have an even number of fingers (which is clearly false). In other words, (1d) ascribes a modal property to a linguistic item (a sentence, as long as we do not want to commit to any specific account of semantic content). A more charitable reading of (1) – on the assumption that I actually have ten fingers – should be *de re*:

(1r) $\quad \exists x(N(myfingers) = x \wedge \Box\, even(x))$

According to (1r), the *object* denoted by "the number of my fingers" in the actual world – i.e. the number 10 – necessarily has the property of being even, which is true. In light of this, it seems that we can easily make a case for a notion of *worldly intensionality*: apparently, some statements involve intensional notions because they capture features that reality instantiates objectively, and not simply because intensionality is a feature of our way of talking about (and thereby representing) certain aspects of reality.[6]

This sort of distinction gets reflected in the different analyses we provide in a formal

6. While extensionalists like Quine would deem (1r) as nonsensical, it is nowadays standard to treat *de re* modal ascriptions as perfectly meaningful readings. Indeed, such readings are needed if we want to capture modal facts appropriately – see the *locus classicus* Kripke (1980).

language for the different uses (and truth-conditions) of the same modal expressions in a natural language: worldly (*de re*) and representational (*de dicto*). Notice that this does not amount to a mere distinction in subject matter: one may reasonably hold that (1d) and (1r) have the same subject matter (after all, they are two distinct readings of the same sentence), but different truth conditions in virtue of the different binding of the necessity operator, corresponding to a different source of intensionality.[7]

Likewise, some statements may involve hyperintensional notions because they capture features that the world instantiates objectively, and not simply because we "represent the world hyperintensionally". We must avoid the mistake of collapsing the distinction between reality and the language we use to talk about it, both in the sense of ascribing properties of our representations to the world, and in that of ascribing properties of the world to our representations.

In order to clarify this point, consider the case of metaphysical grounding. Just like it seems implausible that the number 10 is necessarily even *because we represent intensional modality like that*, or because *it behaves semantically like that*, it seems equally implausible that the existence of {Socrates} is grounded on the existence of Socrates *because we represent grounding like that*.[8] In grounding cases, pairs of sentences like "{Socrates} exists" and "Socrates exists" behave hyperintensionally – we cannot substitute one for the other *salva veritate* despite their being intensionally equivalent – but the grounding relation holds regardless of one's representational attitudes towards the notions involved.

To test this intuition, suppose that hyperintensional distinctions are merely the product of our cognitive or conceptual limitations. Of course, we might have some propositional attitude – let us say belief – towards P, without having the same attitude towards its intensional equivalent Q, because we are able to deduce P but not Q from our belief base: for instance, Q is syntactically too complex, or some conceptual limitations prevent us from grasping the content of Q, even if we grasp the content of P. But in putative cases of worldly hyperintensionality, such as the just sketched grounding example, what would be the relevant cognitive limitations? One can make a distinction between the existence and identity conditions of {Socrates} and those of Socrates even when one is seemingly on top of all the relevant notions (one understands well who Socrates was, masters set theory perfectly, etc.) and, additionally, one can parse the syntax of all the relevant sentences. In other words, what *cognitive mistake* is an agent making, when he tells apart the existence/identity conditions of Socrates and {Socrates}, and thereby distinguishes "{Socrates} exists" from "Socrates exists"? This represents a serious challenge for those who tie hyperintensionality to the cognitive or conceptual limitations of agents.[9]

7. If one has doubts concerning whether two sentences with the same subject matter can have different truth values, consider "it is raining" and "it is not raining". These must have different truth values. However, negation is typically taken to be topic-transparent, so we can reasonably identify the subject matter of both sentences with *'whether it is raining'*. For more on this, see Berto (2022).

8. {Socrates} is Socrates' singleton, that is, the set whose only member is Socrates.

9. Thanks to an anonymous referee for highlighting this point. For a recent reply that appeals to alleged unreliable heuristics at play in distinguishing intensional equivalents, see Williamson (2024).

To borrow another example from Nolan, consider the following counterpossible (i.e., a counterfactual conditional with an impossible antecedent):

(2) If there were a piece of steel in the shape of a 36-sided platonic solid, it would have more sides than any piece of steel in the shape of a dodecahedron.

As highlighted by Berto and Nolan (2023), (2) seems to be about pieces of steel and their shapes: «which blocks of steel have which shapes is not just about us and our representations. It would not be even if steel could take shapes that it in fact cannot». Again, the truth of (2) does not seem to depend on our representational attitudes or merely theory-laden (in virtue of the vacuity of counterpossibles within a standard intensional semantics): (2) strikes us as intuitively true, even before subscribing to any specific semantic framework. We cannot account for this fact by intensional means only, for instance by employing the standard framework of Possible Worlds Semantics. But even though we need an hyperintensional setting, this does not imply that there is something distinctively representational going on with (2).[10]

Summing up, an illuminating way of framing the question about the source of hyperintensionality seems to be through the comparison with the question about the source of intensionality. For instance, the ascription of a modal property like (1) is intensional because it requires intensional notions (e.g. possible worlds) to be properly analyzed. Intensionality may well be characterized as a linguistic phenomenon, but the circumstances that bring about intensionality at the linguistic level do not always derive from our representations. In cases of *de re* modality, the intensional features of the language employed to analyze the ascription of a modal property derive directly from the actual possession of that property by the object of which it is ascribed. Intensional modality can be a feature of non-representational objects, not exclusively of the expressions that we use to talk about them. In this sense, intensionality can be worldly. Similarly, hyperintensionality can be worldly, as long as the features that require hyperintensional explanations are directly instantiated by non-representational portions of reality, just like the above-mentioned cases suggest.

As we shall see in the next section, the conceptualist reaction to worldly hyperintensionality parallels the extensionalist reaction to the so-called intensional revolution that took place between the 1960s and 1980s: a time that witnessed the rise of the intensional devices that are nowadays ordinarily employed across multiple fields of philosophical inquiry. Back then, many philosophers were unwilling to take intensionality seriously outside of the merely linguistic-representational domain. As remarked by Nolan, a common strategy adopted by extensionalists consisted in paraphrasing intensional concepts by mentioning them in quotation, in order to reinterpret modality as a feature of sentences only. This allowed them to analyze every

10. Counterfactuals are often understood as capturing dependence facts about reality, and not merely about our representations or epistemic states. This gets reflected in the widespread use of counterfactuals to account for non-representational notions such as causation (Lewis (1973a), Woodward and Hitchcock (2003)), alethic modality (Williamson (2010)) and even non-causal metaphysical explanation (Reutlinger (2017)). I will explore this feature of counterfactuals in more detail in chapter 5.

intensional phenomenon as if it were ultimately representational. According to the extensionalist picture, when we assert that Jim is necessarily human, we are not ascribing a property to a non-representational individual, namely Jim himself. Instead, we are actually ascribing a property of being necessary to the sentence "Jim is human". In other words, within the extensionalist picture there was no place for *de re* modality. Nowadays, this would constitute a quite unorthodox view, as Nolan (2014, p. 155) appropriately points out:

> while extensionalism still has some adherents, I think it is fair to say that most philosophers working on necessity, belief and so forth, find these extensionalist maneuvers clumsy, inadequate to the data, and philosophically perverse.[11]

What about the conceptualist reaction to worldly hyperintensionality? In what follows I will discuss one of the main criticism that has been raised against the idea that hyperintensionality can have its source in the non-representational realm.

2.3 Against Worldly Hyperintensionality

Byrne and Thompson (2019) reject Nolan's idea of worldly hyperintensionality on conceptualist grounds. Although they accept that hyperintensional language is needed to accurately describe the world, they reject the view that hyperintensionality can genuinely derive from it: ≪ hyperintensionality derives from features of representations≫ (p. 153).

The authors defend their claim by sketching a Fregean picture of representational hyperintensionality analogous to one that, they argue, can be employed for analyzing intensional notions. Fregean accounts hold that linguistic expressions – at least those embedded in non-extensional contexts – do not refer directly to worldly entities such as objects, properties and states of affairs. Instead, reference is mediated by the *senses*, or *modes of presentation* of the objects denoted by the linguistic expressions.[12] Hence, the content of natural language expressions are not worldly entities that do not depend on our representational faculties, but *ways* in which we conceptualize such entities: a kind of representation.

While Byrne and Thompson do not subscribe to any specific *formal* framework, they provide some examples to defend their intuition. To see their strategy at work, consider the following sentences:

(3) a. It is necessary that 10=10.

11. A similar kind of resistance, this time on intensionalist grounds, exists nowadays towards hyperintensionality. Williamson (2013) can be considered the greatest example of this tendency. For a follow-up and further discussion in defense of hyperintensionality, respectively, see Williamson (2024) and Berto (2024).

12. Explaining what sort of entity Fregean senses are supposed to be, metaphysically speaking, is one of the toughest issues that a defender of such a view must face; Byrne and Thompson avoid this specific point.

(3) b. It is necessary that the number of my fingers=10.

From a Fregean standpoint, the difference in meaning (and the corresponding difference in truth value) between (3a) and (3b) does not depend on the denotations of "10" and "the number of my fingers", which are contingently the same. The distinction depends on the fact that they have two different *senses*, and substitution *salva veritate* of expressions with different senses can fail even in intensional contexts.

According to Byrne and Thompson, the same happens with ordinary cases of representational hyperintensionality, such as those triggered by propositional attitudes. Within this setting, propositional attitudes are relations that agents bear with senses. Suppose that Anna does not know that Brian Hugh Warner is Marilyn Manson.

(4) a. Anna believes that Marilyn Manson is a great artist.

(4) b. Anna believes that Brian Hugh Warner is a great artist.

(4a) and (4b) differ in truth value because they say that Anna bears a relation with two different senses, despite the fact that the worldly referents of those senses are necessarily the same object: for Anna, "Marilyn Manson" and "Brian Hugh Warner" are each associated with different descriptions, and this is enough to explain the (alleged) semantic difference between "Marilyn Manson is a great artist" and "Brian Hugh Warner is a great artist" at the hyperintensional level.

After claiming this, the authors summarize what they take to be Nolan's core thesis in the form of a conditional:

(N) If a subject-matter is not representational, then we cannot explain the fact that we need hyperintensional idioms to describe and explain it in conceptualist terms.

Then, they compare (N) with its intensional counterpart:

(M) If a subject-matter is not representational, then we cannot explain the fact that we need intensional idioms to describe and explain it in conceptualist terms.

Finally, they claim that if (M) can be successfully rejected, there is no reason to think that the same cannot be done with (N). Their strategy begins by considering that «a Fregean analysis of the non-representational intensional operator *per excellence* [i.e., necessity] is well-known» (p. 156), so they take a pair involving necessity – like (3a) and (3b) – interpreted in a conceptualist fashion, and propose it as evidence against (M). More precisely, they argue that what a sentence like (3a) says is that a certain sense has the property of being necessary. Therefore, truth preservation is not guaranteed if one of its constituents is substituted with another one having a different sense.

They conclude that, since senses are representational entities, intensionality ultimately arises from representation, even if the subject matter of a sentence like (3a) is non-representational. In other words, they seem to suggest that the extensionalist move of paraphrasing *de re* modality into *de dicto* modality – treating features of worldly

entities as features of representational entities – is the correct way of approaching the issue of the source of intensionality. By way of analogy, the same should happen with hyperintensionality (p. 157):

> if there's a plausible conceptualist account of an intensional locution about a non-representational phenomenon, and if plausible conceptualist accounts of intensional locutions can be extended to deliver plausible conceptualist accounts of hyperintensional ones, why should we expect these extensions to break down when the locutions are about non-representational phenomena?

I already presented in sec. 2.2 some reasons for rejecting the idea that *de re* intensionality should be systematically paraphrased into *de dicto* intensionality. However, in what follows I will argue that the argument from analogy proposed by Byrne and Thompson is flawed for a number of independent reasons.

2.4 Against Conceptualism

2.4.1 A matter of semantic taste?

Byrne and Thompson's focus on the possibility of extending an intensional Fregean semantics to its hyperintensional counterpart seems to implicitly assume that the problem of the source of hyperintensionality is ultimately a matter of theoretic preferences. But the mere choice of a given semantic framework for settling this kind of issue looks question-begging. If Byrne and Thompson's strategy for showing that hyperintensionality is always representational consists in nothing more than adopting a theory in which semantic contents are *themselves* identified with representational entities, it seems that they are just walking in circles. Independent reasons are required for preferring that particular framework – provided that it actually exists, which is far from obvious – over any other with equal expressive power.

Otherwise, they are claiming something fairly trivial, namely, that some cases of worldly hyperintensionality might be paraphrased in conceptualist terms, just like the extensionalists did with worldly intensionality. But *why* should we want that? One might offer an argument from ontological parsimony, by embracing some principle according to which a theory with a lighter ontological commitment is to be preferred to a theory with a heavier one. But this would simply shift the discussion from the metaphysical problem of the nature of semantic content to the epistemic problem of the trade-off between ontology and ideology.[13]

Furthermore, it is not clear why Fregean senses should be considered less metaphysically problematic than any other candidates for the role of meaning identification (such as sets of worlds or truthmakers in the sentential cases, and individual objects and properties in the sub-sentential cases). Again, independent justification for deeming certain ontologies more acceptable than others would

13. See Berto and Plebani (2015) for an in-depth discussion.

be required. Even though every semantic theory comes with its own ontological commitment to the entities chosen for representing meanings, appealing to the mere fact that a given theory may work at the formal level cannot be used to justify the acceptability of its ontology. One might reasonably reject a given theory by judging its ontological commitment as problematic, but it is far from clear whether and how one might assess a metaphysical claim – such as the one concerning the source of hyperintensionality – simply by adopting a certain semantic framework. Are there strong theoretical advantages in adopting a conceptualist reading, or are conceptualists simply forcing it upon non-representational content? The question becomes even more urgent if we consider how *de re* intensionality has become part of the orthodoxy, despite the extensionalists' efforts and creative paraphrases, as we have seen in section 2.2.

The crucial issue to settle in the present debate is not whether it is possible to have a conceptualist semantics for worldly hyperintensionality, but rather how hyperintensionality arises *regardless of our semantic-theoretic preferences*. Take again the case of *de re* intensional modality. It seems that our justification for accepting it often comes from arguments that do not rely exclusively on semantic considerations. This is most evident with cases of necessary *a posteriori* truths like "Hesperus is Phosphorus" or "water is H_2O", where the modal status of the sentences is not established by means of semantic arguments only.[14] If we have solid, independent reasons to accept *de re* metaphysical modality, we should expect our best semantics to capture it accordingly. Generally speaking, an appropriate semantic framework is supposed to model the phenomena (for instance, the instantiation of *de re* modal properties), and not the other way around: we should not attribute features of the framework to the phenomena that it models. If we grant this, Byrne and Thompson's conditional reconstruction of Nolan's view misses the point.

The weakness in Byrne and Thompson's reasoning can be further exposed by employing their own strategy against them. Recall Byrne and Thompson's core thesis: «hyperintensionality derives from features of representations» (p. 153). Suppose that the mere adoption of a certain semantic framework suffices for establishing that, as they seem to claim. If that were the case, one could develop a similar argument for defending the opposite thesis, namely, that hyperintensionality *does not* derive from features of representations. All one needs to do is to adopt a semantics for intensional phenomena paired with a non-representational ontology and stress that, by analogy, the same can be done for hyperintensional phenomena.

For instance, suppose that in order to deal with intensional phenomena one adopts a version of Possible Worlds Semantics (PWS) paired with a heavily non-representational modal ontology, such as the infamous Lewisian modal realism.[15] According to this view, possible individuals and possible worlds are concrete entities (as opposed to abstract or representational), by no means less real than the actual ones. This enables Lewis to provide a fully extensional analysis of *de re* modal notions, drawing a straightforward

14. For instance, Kripke (1980) famously provides a battery of metaphysical and epistemic arguments for necessary *a posteriori* truths, with direct consequences on the choice of an appropriate semantics.

15. See Lewis (1986).

non-representational picture of intensionality. In other words, if we adopt Byrne and Thompson's understanding of the relationship between ontology and semantics, the Lewisian framework would locate the source of all intensional phenomena in the non-representational realm, by assuming the concreteness of possible worlds and individuals.

At this point, one might easily replicate Byrne and Thompson's move for attributing the just sketched worldly picture of intensional content to hyperintensional content: it would suffice to extend Lewis-style realist PWS in a way that accommodates also hyperintensional distinctions. Funnily, in this case we would not even need to *suppose* the existence of such a framework: any Impossible Worlds Semantics paired with a form of extended modal realism would do the job. For instance, Yagisawa (2010) describes a modal space of real (as opposed to fictional or *ersatz*) *impossible worlds*, which represent all sorts of impossibilities by instantiating them directly. If we set aside the extremely controversial notion of real impossibilities and focus on the purely functional aspect of how such a framework analyzes content – just like Byrne and Thompson do with their hypothetical extended Fregean semantics – we can easily recognize that, by adopting it, we would have a fully non-representational account of hyperintensional content. For instance, the distinction between necessary equivalents like "Marilyn Manson" and "Brian Hugh Warner" would be captured by the fact that, at some impossible worlds, they do not denote the same individual, regardless of our ways of representing him.[16] Hence, by replicating Byrne and Thompson's reasoning, we would be able to conclude that hyperintensionality *does not* derive from features of representations: after all, we would have a fully non-representational account for hyperintensional content.

To rephrase them again: if there is a plausible *worldly* account of an intensional locution about a non-representational phenomenon, and if plausible *worldly* accounts of intensional locutions can be extended to deliver plausible *worldly* accounts of hyperintensional ones, why should we expect these extensions to break down when the locutions are about *representational* phenomena? If such an analogy would not be acceptable, since it would mistakenly locate the source of representational hyperintensionality in the world, then the same is true for the opposite analogy defended by Byrne and Thompson, since it mistakenly locates the source of worldly hyperintensionality in representation.

2.4.2 Extensibility and granularity

A further issue concerns the alleged extensibility of the Fregean intensional picture to its hyperintensional counterpart, which is far less obvious than Byrne and Thompson claim. Even if a fully Fregean account of intensional locutions is available, does that constitute genuine evidence for the possibility of an extended account which may work also for all cases of (*prima facie*) worldly hyperintensionality? Byrne and Thompson declare themselves «confident that such semantics could be formulated for most of them at least» (p. 155). However, they do not go further than sketching a counterexample to

16. In his discussion of belief ascriptions Yagisawa (2010, chap. 9) defends a similar view.

(M) and claiming that we can reject (N) on the same basis, without addressing the issue of what such an extended Fregean semantics should look like. Obviously, they do suggest an equivalence between the basic and the extended framework with respect to ontological commitment: in order to represent meanings, both frameworks shall include representational entities only (Fregean senses).

But the idea of a Fregean semantics that must be able to deal with *most* kinds of hyperintensionality still sounds a bit too optimistic. So far, several different frameworks have been developed for dealing with variably broad families of hyperintensional phenomena, but none of them had the ambition of providing a systematic account for all or even most of them. In the words of Leitgeb (2018), p. 308:

> no system so far has managed to set a general standard or to gain anything like a status of 'canonicity' for all hyperintensional operators whatsoever, and it seems unlikely that one system will ever be able to handle all hyperintensional operators simultaneously and remain informative with respect to the 'internal logic' of hyperintensional contexts.

The development of a unified semantics restricted to *representational* hyperintensionality only (i.e., a unified semantics for at least all propositional attitudes) already appears just as ambitious as unlikely. Moving to the *prima facie* non-representational realm, even the most explicit attempt towards a hyperintensional *Fregean* semantics to date (Skipper and Bjerring (2020)) is not suited for worldly hyperintensionality, as its proponents openly admit (p. 22):

> we are not claiming that our framework can be used to shed light on all hyperintensional phenomena that one might want to reason about. For instance, Nolan (2014) surveys a number of interesting and important issues in the area of hyperintensional metaphysics that lie beyond the scope of our framework. Moreover, we are not claiming that epistemic n-intensions can be used to shed light on all the semantic phenomena that one might want to reason about. [...] if we are interested in capturing an externalist notion of content – as championed by, for example, Kripke (1980) and Burge (1979a, 1979b) – epistemic n-content may not be of much use.

Again, the main reason for such limitations seems to be rooted in the very nature of certain hyperintensional phenomena, which in turn depends on their source. This is a general worry that arises regardless of any specific conceptualist account one might adopt.

Indeed, a well-known risk for any strongly representational theory of semantic content is that of being *too fine-grained*. Consider the following sentences:

(5) a. Marilyn Manson is an artist because Marilyn Manson is a singer.

(5) b. Brian Hugh Warner is an artist because Marilyn Manson is a singer.

No propositional attitude is involved in (5a) and (5b). A natural reading of 'because' here is the one capturing a hyperintensional determination or grounding relation. Since

'being an artist' is a determinable property that can be determined (grounded) by the instantiation of 'being a singer', and Marilyn Manson is identical to Brian Hugh Warner, (5a) and (5b) are not only true, but also at least ground-theoretically equivalent. Whatever semantics for this kind of construction turns out to be the correct one, it needs to capture at least this fact. Furthermore, it needs to be fine-grained enough to capture hyperintensional distinctions between sentences like "Socrates exists" and "{Socrates} exists" (recall the grounding example discussed in section 2.2), but also coarse-grained enough to guarantee the substitutivity *salva veritate* between "Marilyn Manson is an artist" and "Brian Hugh Warner is an artist" in contexts like (5a)–(5b). This is an instance of the so-called *granularity problem*.

Now we should ask: do Fregean approaches always cut contents with the appropriate granularity? As long as they are committed to the view that contents are in some sense representational, they will arguably hold that "Marilyn Manson is an artist" and "Brian Hugh Warner is an artist" have different senses. After all, one might believe the former without believing the latter, and this amounts to a difference in their representational content. But this means that (5a) and (5b) must express different contents as well, on pain of violating compositionality.[17] If that is the case, the answer to our granularity question is obviously no: Fregean approaches face the risk of being too fine-grained in many contexts. This raises further doubts concerning the extensibility of the Fregean picture from intensional to (most or all) hyperintensional locutions, and answers to Byrne and Thompson's question: «why should we expect these extensions to break down when the locutions are about non-representational phenomena?».

Notice that I am not claiming that Fregean accounts can *never* be extended to deal with *any* hyperintensional phenomena. On the contrary, their application to several representational contexts is certainly a viable option. My only aim here is to highlight that Byrne and Thompson do not provide compelling reasons to accept that such approaches can in principle be extended to all or even most hyperintensional phenomena, while they seem to require this in order for their argument to go through. More precisely, Byrne and Thompson seem committed to the following conditional: if there is a general representational treatment of hyperintensionality, then hyperintensionality derives from representation. The points raised in Section 2.4.1 are enough to reject the conditional as a whole, but even if we granted its acceptability, here I have pointed out the lack of evidence for its antecedent. Our available evidence suggests instead the opposite, namely, that it is not the case that Fregean approaches can be extended for a general representational treatment of hyperintensional phenomena. Hope is the last to die, but so far the burden of proof is still on the conceptualists.

17. Strictly speaking, the violation would concern the analogous principle of *reverse* compositionality, which says that if we start with an expression *E* and replace part of *E* with an expression having a different meaning, we end up with an expression having a different meaning than *E*.

2.5 Conclusion

The issue of the source of hyperintensionality in cases that do not involve propositional attitudes might not worry those who subscribe to a broadly externalist picture of semantic content, or those already sympathetic with Nolan's view that hyperintensionality comes in a variety of flavors, both representational and worldly. The present chapter aimed at showing the failure of the main conceptualist reaction to this view, and at providing insight on the notion of worldly hyperintensionality through the comparison with the notion of worldly intensionality.

In order to reject the view that hyperintensionality can be non-representational – that is, having its source in the world rather than in our representations of it – Byrne and Thompson's analogy with intensional locutions does not work. On the contrary, that very same analogy seems to support Nolan's remarks about worldly hyperintensionality. The distinction between *de re* and *de dicto* intensional modality captures a distinction between different sources of intensionality: worldly and representational. This is nowadays a platitude, so it seems reasonable to expect a similar kind of distinction also at the hyperintensional level, instead of postulating a single kind of hyperintensionality. However, Byrne and Thompson's argument fails because it relies on the idea that the issue should be settled simply by choosing an appropriate semantic framework, paired with a representational interpretation of its ontological commitments.

Against that, I showed how (i) by replicating Byrne and Thompson's strategy it is possible to defend the opposite of their claim (in an equally unsatisfying way), namely that hyperintensionality does not derive from features of representations; (ii) the existence of a Fregean semantics for intensional locutions does not guarantee the possibility of extending it to cover all or even most cases of hyperintensionality: the available evidence suggests instead that Fregean approaches will often turn out to be too fine-grained for that purpose.

Chapter 3

Hyperintensional Propositions I: Semantic Adequacy and Related Issues

3.1 Background

Once hyperintensional phenomena (be they representational or worldly) are identified, and a corresponding need for a hyperintensional semantics is recognized, the next step is to find an appropriate formal framework. As already discussed in chap. 2, it seems unlikely that a unique account will ever exhaust the whole range of hyperintensional phenomena. In line with this, the last part of the present work will focus on some specific instances of hyperintensionality. However, before exploring their peculiarities, I want to approach what is probably the broadest notion of this kind: *semantic content*.

It is sometimes claimed that *what is said* is a representational matter. Just to mention a recent example that is particularly relevant for the present discussion, the section on hyperintensional semantic content in the Stanford Encyclopedia entry on hyperintensionality (Berto and Nolan (2023)) is placed under the section labeled 'hyperintensionality in representation'. This is not surprising, especially if we consider that most of the classic examples in the literature involve conflicts in the evaluations of certain attitude ascriptions, presented as evidence for the difference in meaning among many intensionally equivalent expressions.

However, as long as semantic contents are understood from a broadly externalist perspective, they seem to be the kind of things whose identity and existence conditions should be largely independent of our representational attitudes. I already illustrated some evidence for this in sec. 2.4.2: representational content (which may very well be identified with Fregean senses or entities alike) sometimes provides an excess of fine-grain, even within some uncontroversially hyperintensional contexts. The case of metaphysical grounding ascriptions (where grounding is understood as a sentential operator), provides a strong reason to believe that the appropriate granularity for

this type of semantic content lies somewhere between intensional equivalence (too coarse-grained) and equivalence of representational content (too fine-grained). Is it the case that semantic content *in general* display this feature?[1] While what follows is not meant to provide a definitive answer to this question, I want to illustrate some further evidence that casts doubts on the opposite view, namely, that semantic content has the granularity of representational content. Indeed, if we let propositional attitudes (and thereby representational equivalence) to guide our search for a general account of semantic content, we get some very weird results.

To see this, suppose that we possess an adequate criterion for sentential synonymy (the technical details do not matter for now). That is, a criterion for co-hyperintensionality that perfectly captures sameness of meaning for sentences according to the linguistic data (at least for expressions that are not within the scope of propositional attitudes). Now, consider two sentences that undoubtedly express the same content:

(1) Mary loves Paul.

(2) Paul is loved by Mary.

Let us use $[A]$ to denote the semantic content of A. According to our ideal criterion, [1]=[2].[2] However, it is possible that John believes (1) – I will abbreviate this as $B_j(1)$ – without believing (2). Hence, as long as our criterion is at least truth-conditional – i.e. intensional equivalence is a *necessary* condition for its satisfaction – $[B_j(1)] \neq [B_j(2)]$.

This leaves us with two options: (i) an ideal criterion for sameness of sentential meaning may not preserve compositionality (and thus, it would not be very useful as a *semantic* criterion); or (ii) belief (and arguably, propositional attitudes in general) creates hyperintensional contexts that are even more fine-grained than sameness of sentential meaning, and that require a form of representational equivalence to grant substitutivity. Not only (i) seems very implausible, but there seem to be independent reasons to believe (ii): as discussed in section 2.4.2, identifying contents with representational entities (those that could play the role of "modes of presentation" of linguistic expressions for

1. An interesting class of hyperintensional expressions that might require a notion of content more fine-grained than grounding equivalence without explicitly involving representational notions are some intensional transitive verbs. See Graeme (2020) for an in-depth discussion.

2. A possible objection to this equivalence might be formulated from the perspective of two-component semantics (Berto (2022)) and theories of aboutness (Yablo (2014)). One may claim that (1) and (2) cannot have the same meaning, since they have a different subject matter – (1) is about *whether Mary loves Paul*, and (2) is about *whether Paul is loved by Mary*: this is why someone may believe the former without believing the latter. Is this the correct approach to topicality? Different accounts of subject matter may result in different takes on the issue. For instance, according to Fine's Truthmaker Semantics (Fine (2017b)), as long as (1) and (2) have the same exact truthmakers (which seems the most reasonable option here), they will have the same subject matter, corresponding to the fusion of their truthmakers. This becomes problematic only if we allow propositional attitudes to have a saying about what the correct granularity of subject matters is. Why should we allow this? Appealing to the representational nature of semantic content begs the question. Maybe, a different ontology of subject matters is required for different pieces of language (very fine-grained for propositional attitude reports, more coarse-grained for expressions not embedded in such attitudes). For more on topicality, see Berto (2022).

agents) delivers poor results in some worldly hyperintensional contexts. But this means also that any alleged counterexample to same-saying in natural language that appeals to propositional attitudes (and thereby to representational content) is not compelling. In particular, the fact that $[B_j(1)] \neq [B_j(2)]$ should not imply that $[1] \neq [2]$.[3]

The conceptual distinction between co-hyperintensionality for semantic content and what I dubbed as equivalence of representational content can be further illustrated by considering the phenomenon of referential opacity. Ever since the discovery of Frege's puzzles (Frege (1892)), sentences with the same meaning that differ only for an occurrence of a co-referring proper name have been employed to illustrate the opacity created by propositional attitudes. Consider again:

(3) Marilyn Manson is a great artist.

(4) Brian Hugh Warner is a great artist.

Since Marilyn Manson=Brian Hugh Warner, (3) and (4) have *prima facie* the same meaning. However, Anna may very well believe (3) without believing (4). Should we then conclude that (3) and (4) actually express two different contents? A positive answer amounts to maintaining that semantic content *is* representational content, so that even atomic sentences that do not contain definite descriptions and are not embedded within propositional attitudes can be referentially opaque, which seems absurd.

One of the reasons why these two notions (hyperintensionality and referential opacity) get sometimes conflated is linked to the tendency, discussed in chap. 2, to consider all cases of hyperintensionality as representational. Representational attitudes are referentially opaque. Clearly, all cases of referential opacity are hyperintensional (failure of substitutivity *salva veritate* of two co-referring rigid designators is failure of substitutivity *salva veritate* of two co-intensional expressions). This may lead to think that the converse (i.e. all cases of hyperintensionality are referentially opaque) is true as well. However, there seem to be plenty of cases of hyperintensional contexts that are *not* referentially opaque. Why should we think that a more general notion of non-representational semantic content should be different? An externalist account of semantic meaning, while hyperintensional, should not be referentially opaque: pairs of sentences like (3) and (4) *do* have the same meaning and our *prima facie* intuition is correct. Generally speaking, if we assign meanings to sentences according to a criterion of co-hyperintensionality, we must be careful in distinguishing same-saying from

3. What happens to compositionality for statements like $B_j(1)$? Maybe its meaning should *not* be a function of that of (1), even when meaning is not understood merely intensionally. Notice that I am not making the obvious claim that propositional attitudes are not truth-functional: I think that the *hyperintensional* content of $B_j(1)$ might be independent from the (non-representational) hyperintensional content of (1). We can come up with a semantics assigning different contents to $B_j(1)$ and $B_j(2)$ while assigning the same to (1) and (2). The challenge would be to spell out compositional clauses for the belief operator while staying adequate with respect to the linguistic data; an in-depth discussion of this issue lies beyond the scope of the present work. For a classic diagnosis of this kind of issues, see Salmon (1986). For a recent contextualist explanation, see Dorr (2014).

equivalence of representational content. In a slogan: belief is not a guide to meaning.[4]

In light of this, for the rest of the work I will focus on an objective notion of semantic contents, namely, as some kind of non-representational entities. This view is in contrast with a Fregean/conceptualist approach to the topic, which seems more consistent with an internalist picture of meaning and might turn out to be more adequate as an account of representational content. On the other hand, a long (and to some extent, still orthodox) tradition in philosophy of language is more in line with the non-representational view of semantic contents, in the shape of one of the most fruitful paradigms in analytic philosophy: the truth-conditional picture of meaning. Through the works of giants like Wittgenstein, Carnap, Kripke, Lewis and Stalnaker, this general paradigm proved itself as a solid foundation for a huge amount of technical and philosophical work. In accordance with this, the focus for the rest of the work will be almost exclusively on truth-conditional accounts of semantic content.

Traditionally, the role of semantic contents for sentences has been played by propositions. While the debate on their metaphysical status and the technical challenges associated with them is ongoing (I will discuss some of these in the next chapter), it is common to associate a distinct account of propositions to each fully spelled-out formal semantic framework. In the rest of this chapter, I will assess one of the most popular truth-conditional accounts of hyperintensional propositions, associated with a particular semantic framework: Impossible World Semantics (IWS).[5] In particular, I will focus on its viability as an account of semantic content, against the background of the just sketched externalist, non-representational picture of meaning. Chapter 4 will cover further issues surrounding IWS and its comparison with an alternative hyperintensional truth-conditional account: Truthmaker Semantics (TMS).

3.2 From Possible to Impossible Worlds

The standard truth-conditional picture of meaning finds its highest and fullest expression in Possible Worlds Semantics (PWS), according to which the content of a sentence is its intension, i.e. a function from possible worlds to truth values. An equivalent and more succinct way of phrasing this is to say that, according to PWS, the proposition expressed by a sentence is the set of possible worlds at which it is true. A widely recognized limitation of PWS is that it collapses the distinction among co-intensional (necessarily equivalent) propositions: necessary truths are identified with the set of all possible worlds, and impossibilities are identified with the empty set.

For instance, tautological sentences such as "if Mario Draghi is a human, then Mario Draghi is a human" or "either it is raining or it is not raining" are true precisely

4. For recent arguments (on intensionalist grounds) against the idea that semantic content is representational content, see Williamson (2024), chap. 4.

5. Alternative frameworks to deal with hyperintensionality that have been offered are the two-dimensional semantics of Chalmers (2008), the structured propositions approaches of King (King 1995, 1996, 2009), Soames (Soames 1985, 1987), and others. Recent approaches include Yablo (2014)'s work on aboutness and Berto (2022)'s two-component semantics.

at the same possible worlds: all of them. Thus, according to PWS, they all express one and the same proposition. Similarly, sentences that are true at no possible world such as "Mario Draghi is a dragon" and "it's raining and it's not raining" express the same proposition, represented by the empty set. Although it seems obvious that these sentences have different meanings, PWS cannot distinguish among them. In short, PWS lacks the resources to account for hyperintensional semantic distinctions.

While PWS still maintains its status as the dominant framework for a great portion of contemporary philosophical theorizing, among the most popular solutions proposed to overcome its limitations we find its natural hyperintensional refinement: *impossible* worlds semantics (IWS).[6]

Impossible worlds are worlds in which some impossible state of affairs obtains. For the current purposes, I will consider accounts of semantic content committed to the finest grain of distinction between necessary equivalents, in order to have a general account of propositions in terms of worlds, both possible and impossible. This amounts to including worlds that are not closed under any non-trivial logical rule, as well as worlds where metaphysically necessary equivalents such as Hesperus and Phosphorus are not one and the same object, along the lines of the accounts proposed by Yagisawa (2010), Jago (2015) and Berto and Jago (2019).[7]

The conceptual core of such approaches is the identification of a proposition p with the set of (possible and impossible) worlds at which it is the case that p. This allows for a straightforward explanation of a variety of hyperintensional phenomena. For instance, consider the paradigmatic case of propositional attitude reports: it is possible that Jim believes that Hesperus is bright without believing that Phosphorus is bright. Within an impossible worlds semantics, the mismatch in truth-value between "Jim believes that Hesperus is bright" and "Jim believes that Phosphorus is bright" is accounted for by the existence of some impossible worlds where Hesperus is not identical to Phosphorus, so that "Hesperus is bright" and "Phosphorus is bright" are identified with two different sets, and thus they express two distinct propositions. Generally speaking, an account of propositions of this kind is able to capture very fine-grained distinctions between intensionally equivalent contents.

The most pressing question that any advocate of IWS has to address is the one concerning the ontological status and nature of impossible worlds. While there exist realist options available in the literature (see the extended modal realism of Yagisawa (2010), according to which impossible worlds are just as real as the actual one), the least controversial and most popular route is to understand impossible worlds as some sort of abstract entity. Berto and Jago (2019) extensively discuss and review all the existing options, and conclude that the best variant of non-realist IWS is a form of *linguistic ersatzism*. In light of this, they propose their own strain of ersatz IWS, which currently

6. Accounts of this kind trace back at least to Rantala (1982) and Priest (1992).

7. The main differences between these accounts concern the ontological status of the worlds to which they are committed (in Berto and Jago's accounts worlds are representational *ersatz* entities, while in Yagisawa's account they are as real as the actual world); and the room for more than two truth-values (Yagisawa identifies propositions with tuples of sets of worlds instead of simply sets of worlds for this reason). The relative advantages of one approach over the other are not relevant for the present discussion.

stands as the most developed account of this kind available in the literature.

In what follows, I will introduce the impossible worlds framework proposed by Berto and Jago (2019) in more detail, in order to argue that it faces a serious dilemma: either it is inadequate with respect to phenomena of same-saying in natural language, or it is not compositional. I aim to show that, one way or another, IWS based on linguistic ersatzism is not viable as a formal semantics for natural language. While the focus of the present chapter will be Berto and Jago's account of how impossible worlds should be constructed, I will make clear that my objections should be broad enough to represent a challenge to any approach relying on similar techniques.

3.3 Building Impossible Worlds Compositionally

The basic ingredient of the semantics developed by Berto and Jago is the notion of *open world*, first introduced by Priest (2005). An open world is a world that is not required to obey any logical closure principle other than $A \vDash A$. In order to track hyperintensional distinctions among intensionally equivalent propositions, within such a framework $[P]$ – the semantic content of P – is identified with the set of open worlds at which P is true.

Prima facie, the introduction of open worlds amounts to rejecting some fundamental semantic facts. For instance, for some open worlds it is not the case that $A \wedge B$ is true if and only if A and B are both true. More generally, open worlds represent violations of a form of compositionality, according to which the semantic content of a sentence (i.e. the worlds at which it is true) should be a function of that of its constituents.

From a set-theoretical perspective, within an open worlds-based account we do not get the content of $A \wedge B$ by simply taking the intersection of $[A]$ and $[B]$. If, at some open worlds, the truth of $A \wedge B$ is independent from the truth of A and B, then $[A] \cap [B]$ will not contain some $A \wedge B$-world. Therefore, $[A \wedge B] \neq [A] \cap [B]$. Given that the former is not a function of the contents of A and B, it seems that the strong connection between semantic content and compositional truth-conditions no longer holds when open worlds are introduced.

However, the introduction of open worlds in the framework allows for a very fine-grained picture of semantic content, in compliance with our hyperintensional needs. The simplest (and most expressive) semantics of this kind is one that encompasses the full variety of open worlds to deliver the finest hyperintensional granularity. Now, in order to build the semantics, the impossible worlds theorist needs first to address the question about how open worlds represent contents – namely, how should such worlds be constructed?

As said earlier, the options available in the literature mirror those that have been proposed for addressing the same question about possible worlds.[8] Since the only constraint placed on open worlds is that $A \vDash A$, the kind of world-representation at stake here will be at least as fine-grained as the sentences of the object language through which we model natural language. Otherwise, as remarked by Berto and Jago (p. 177),

8. See Lewis (1986) and Berto and Jago (2019) for an overview.

> there would be distinct object-language sentences A, B such that a world represents that A iff it represents that B. But this gives us a closure principle $A\dashv\vDash B$ which all worlds must obey, contrary to what we established above.

Linguistic ersatzism – the view according to which worlds are sets of sentences – seems to be the best candidate to complete the task.[9] This choice is motivated by Berto and Jago by appealing to two facts.

The first concerns the granularity requirements mentioned above. Linguistic ersatzism identifies worlds with sets of sentences, and this is compatible with the view that any arbitrary set of sentences counts as a world. If that is the case, then world-representation will be as fine-grained as the language employed for constructing worlds itself, namely, the *worldmaking language*. In other words, for any pair A and B of worldmaking sentences, according to linguistic ersatzism there will be a world which represents that A but does not represent that B.

The second fact has to do with how ersatz worlds represent. According to Lewis (1986), ersatz possible worlds represent possibilities in a similar way to how language does. In particular, they represent possibilities without having to resemble them, but only in virtue of the structural features of the representation. The same obviously applies to ersatz impossible worlds.

Summing up, under the linguistic version of ersatzism worlds are sets of sentences of a worldmaking language. Such a language is nothing more than the specified language to which the sentences belong. More precisely, Berto and Jago suggest the possibility of employing a *Lagadonian* language, namely, an artificial language in which each individual, property and relation names itself.[10] By combining them in syntactically well-formed strings (provided that we also have worldmaking translations for the logical connectives and operators), we get the worldmaking translation for every sentence of our object language. Within this setting, a world w represents that A if and only if the worldmaking translation of A belongs to w.

Despite not being immune to well-known limitations, linguistic ersatzism seems to provide a fairly solid theoretical basis for developing a very fine-grained semantics from impossible worlds.[11] According to Berto and Jago, one further advantage of linguistic ersatzism lies in its ability to avoid *the compositionality objection.*

Compositionality, as previously mentioned, is the property in virtue of which the semantic content of a complex expression is a function of the contents of its constituents. A corresponding normative principle which is standardly assumed takes compositionality to be a minimal requirement of any adequate semantic framework. In other words, being compositional is typically considered a mandatory feature for a theory of meaning. However, as noticed above, recapturing compositionality within an impossible worlds framework looks far from easy. Indeed, the open worlds theorist needs to deal with the following claims:

9. Linguistic forms of ersatzism date back to Carnap (1947).
10. See Lewis (1986), Sider (2002).
11. See the problem of alien entities in Lewis (1986).

(P1) Impossible worlds are open worlds.

(P2) Open worlds-based approaches are not compositional.

(P3) An adequate semantics must be compositional.

To avoid the following conclusion:

(C) An adequate semantics cannot be based on impossible worlds.

In order to do this, Berto and Jago's strategy focuses on attacking (P2). Their idea is to provide a way to move recursively from the content of a sentence of the object language to the translation of that sentence in the worldmaking language. In particular, they aim at showing that a full hyperintensional account of meaning can be captured in terms of compositional models of ersatz IWS, as they explicitly state (p. 182):

> at least some open worlds-based approaches are compositional. This would show that a full theory of meaning can be given by including open worlds.

Here is their proposal in more detail: let A be a sentence of the object language, and A^* its translation in the worldmaking language. Within this setting, $[A]$ will be the set of worlds that have A^* as a member (where w is any world, and W is the set of all worlds):

$$[A] := \{w \in W \mid A^* \in w\}.$$

Truth-at-a-world is defined in terms of world-membership:

$$(\mathrm{T}_w)\ A \text{ is true at } w \text{ iff } A^* \in w.$$

In light of this, if we add A^* to every world in W, the set we obtain, $\{w \cup \{A^*\} \mid w \in W\}$, is precisely the set of worlds that have A^* as a member, namely $[A]$.

Furthermore, since $\{A^*\}$ is itself a world, the intersection of all the worlds in $[A]$ is the world containing only A^* as a member; that is, $\bigcap[A]=\{A^*\}$.

Summing up, by adding A^* to each world we get $[A]$, that is the content of A; while by taking $\bigcap[A]$ we get $\{A^*\}$, which contains only A^*, namely the worldmaking language translation of A.

This strategy is generalized for complex expressions in the following way: let O_1 and O_2 be any 1-place and 2-place operators, respectively. Let $x^\frown y$ be the concatenation of the worldmaking strings x and y. According to this notation, then, $O_1^\frown A^*$ and ${A^*}_1{}^\frown O_2^\frown {A^*}_2$ are worldmaking sentences.

Two functions f_1 and f_2 are then defined. The former works for one-place operators (O_1) and one sentence, the latter for two-place operators (O_2) and two sentences:

$$f_1(o, c) = \{w \cup \{o \frown x \mid x \in \bigcap c\} \mid w \in W\}$$

Where o is a one-place operator in the worldmaking language (for instance, the worldmaking language translation of the object language negation '¬*'), and c is the content of a sentence in the object language. Clearly, $\bigcap c$ will be some singleton of the form {A*}. Then, f_1(¬*,[A])=[¬A], namely the set of worlds at which ¬A is true (which are exactly the worlds containing the worldmaking sentence ¬* $\frown$ A*).

The same holds for f_2:

$$f_2(o, c_1, c_2) = \{w \cup \{x \frown o \frown y \mid x \in \bigcap c_1, y \in \bigcap c_2\} \mid w \in W\}$$

For instance, consider A, B and their conjunction $A \wedge B$. One can have f_2 taking ∧* (the worldmaking translation of '∧') with the contents of A and B as input to get [$A \wedge B$] as output. This is how Berto and Jago's account manage to satisfy the compositionality requirement: the functions allow us to move back and forth between the contents of the sentences in the object language and their corresponding translations in the worldmaking language.

3.4 A Dilemma

As pointed out by Berto and Jago, the compositionality objection is deflected by setting up a formal machinery that always allows one to move from semantic contents to worldmaking sentences and back, in order to obtain A* from [A] and *viceversa*.

However, such a result heavily relies upon the possibility of generating complex worldmaking language sentences from simpler ones in roughly the same way we do with sentences of natural languages. While this clearly strengthens the viability of IWS by recapturing the basic mechanism of compositionality at the fundamental world-building level, it may still not be enough to make IWS a good semantics for natural language.

As it is standardly assumed, any minimally adequate semantics for natural language must be compositional and capture semantic facts with the correct granularity. Each of these conditions is necessary, but not sufficient on its own. They may be reasonably considered jointly sufficient. However, as we shall see, Berto and Jago's strain of lingistic ersatzism makes them mutually exclusive.

More precisely, I will challenge the ersatzists with the following dilemma: either they employ their mechanism to make IWS compositional, or they make IWS semantically adequate for natural language. If the former path is taken, IWS will turn out semantically inadequate for natural language (i.e. it will not capture semantic facts about the language with the correct granularity); on the other hand, taking the latter path amounts to losing compositionality (their mechanism will deliver the wrong results). This should suffice to show that, at its current state, Berto and Jago's account cannot represent a viable semantics for natural language.

3.4.1 First Horn: compositional IWS is semantically inadequate

In order to show that Berto and Jago's account, while compositional, does not capture semantic facts with the correct granularity, we first need to go back to how the Lagadonian worldmaking languange works.

Recall that, by introducing worldmaking translations for each individual, property, relation and operator and building sentences out of them, to each well-formed string (or tuple) of worldmaking expressions will correspond a different worldmaking sentence. This partly reflects the need for a one-to-one correspondence between the sentences of the natural language (in context and after disambiguation) and the sentences of the worldmaking language.

To better see this, consider "$5 < 7$" and "$7 < 5$". Since they have the same constituents but clearly express two different things, their worldmaking translations should be two different objects. The only way to achieve this is by making worldmaking sentences sensitive to the order of their constituents. In other words, worldmaking translations must preserve syntactic structure.

In light of this, as pointed out by Jago (2015), one can use *Russellian tuples* to represent worldmaking sentences. If that is the case, IWS based on linguistic ersatzism turns out to be equivalent to most accounts that take propositions to be structured entities, as noticed by Jago (pp. 11-12):

> ersatz-world-building sentences are identical to, or at least very similar to, Russellian tuples. If they are not identical, then there is no theoretically important difference between Russellian tuples and ersatz-world-building sentences. [...] Moreover, the two notions of propositions are inter-definable and inter-substitutable. As semantic theorists, we can move freely between the two conceptions of what propositions are. The Russellian and the sets-of-worlds approaches are not genuinely distinct accounts of propositional representation.

If Jago is right, IWS inherits not only the virtues of structured accounts, but also their shortcomings. In particular, a widely known problem of such accounts is that they seem to individuate semantic contents too finely.[12] Consider again:

(a) Mary loves Paul.

(b) Paul is loved by Mary.

These two sentences say the same thing, so their content should be identical according to any adequate semantics for natural language. But as long as 'loving' and 'being loved by' are two distinct relations, there will be two distinct worldmaking translations for them. So, a^* will not be identical to b^*. But this means that there will be open worlds which contain only one of them (most noticeably, the singleton-worlds $\{a^*\}$ and $\{b^*\}$). In other words, $[a]$ will not be identical to $[b]$ within the ersatz picture, contrary to

12. See Pagin (2019) for an in-depth discussion of this issue.

the lingustic data: IWS, just like structured approaches, is too fine-grained to properly account for phenomena of same-saying.[13]

One might still regard the *prima facie* identity of semantic content in the just discussed example as somewhat contentious, for the reasons discussed in the previous section. If so, consider pairs of sentences whose synonymy is blatantly obvious, such as "it is rainy and windy" and "it is windy and rainy", or "5 = 7" and "7 = 5". As long as worldmaking strings are sensitive to the order of their constituents (as required by the theory), each of those pairs will express two distinct propositions.

While it may be useful to distinguish them within specific representational contexts (such as belief ascriptions, where IWS seems to work just fine), the fact that they have the same meaning and thereby express the same proposition will be uncontroversial to any competent English speaker (and this is accepted even by Berto and Jago themselves, as we shall see in the next section).

Summing up, the use of ersatz worlds built out of Lagadonian translations to represent semantic contents delivers the wrong results with respect to phenomena of same-saying in natural language. While this kind of constructions preserves compositionality, it does not capture semantic facts with the appropriate granularity.[14]

3.4.2 Second Horn: semantically adequate IWS is not compositional

Berto and Jago are well aware of the excess of granularity exhibited by the full version of their semantics when applied to natural language and phenomena of same-saying. In this regard, they write:

> there are (hyperintensional) notions of *semantic content* which, we think, require us to restrict the domain of worlds. [...] These notions require worlds more fine-grained than classical possible worlds, but not as fine-grained as open worlds. We obtain those worlds by narrowing down the class of all open worlds, based on certain principles we want to enforce on our analysis (p. 178).

> the 'anything goes' approach to content [...] doesn't give an appropriate analysis of same-saying. Some logical relations (including the one relating $A \wedge B$ to $B \wedge A$) preserve

13. Regarding a similar example, Båve (2019) considers at least three possible distinct analyses (syntactic decompositions), each of which would correspond to a different tuple, so to a different proposition.

14. A parallel argument, discussed also in Jago (2015), has to do with cases in which the semantics seems to be too *coarse*-grained. Consider David Bowie and Ziggy Stardust: they are identical, so the two names correspond to a unique individual. This means that they must have a unique Lagadonian worldmaking translation (the individual itself). As a consequence, the resulting IWS model cannot distinguish between "David Bowie is an alter-ego of David Bowie" and "Ziggy Stardust is an alter-ego of David Bowie", since the two sentences would share the same worldmaking translation, which in turn belongs to the same worlds. A possible solution might be considering predicates like 'being an alter-ego of David Bowie' as genuine properties. But this will just lead to yet another excess of granularity, since (Ziggy Stardust)* $\frown$ (being an alter ego of)* $\frown$ (David Bowie)* will then be distinct from (Ziggy Stardust)* $\frown$ (being an alter ego of David Bowie)*, which seems rather weird. Another solution suggested by Jago is to identify singular terms like proper names as clusters of properties, rather than individual objects. Although this may have better chances of working, it goes openly against what is considered to be the standard account of proper names at least since Kripke (1980).

same-saying. (p. 208).

In particular, they consider cases that strongly support same-saying preservation for at least the following operations: commutativity for 'and' and 'or', distributivity of 'or' over 'and', associativity for 'and' and 'or', idempotence of 'and' and 'or', and the De Morgan equivalences.

This leads them to propose a restriction of the class of open worlds such that only those obeying the above-mentioned principles should be employed to account for same-saying in natural language. Such a move generates the required De Morgan algebra of contents, allowing them to avoid the excess of granularity and to properly capture the linguistic data.

The obvious question elicited by the restriction – but not addressed by the authors – is whether such a move preserves the usefulness of the mechanism they developed to recapture compositionality. The short answer is no. The easiest way to realize this is by observing that any adequate (restricted) IWS model will lack many singletons. That is because, as long as we allow different object-language expressions A and B to have the same meaning, $[A]$ (=$[B]$) must contain only worlds with both A^* and B^* as members. Otherwise, $[A]$ and $[B]$ would be distinct, which amounts to taking A and B as two expressions with different semantic content, against our assumption. Therefore, $\bigcap[A] \neq \{A^*\}$, and equivalently, $\bigcap[B] \neq \{B^*\}$. But this means that one cannot move back and forth from the content of a sentence to its worldmaking translation anymore: the core of Berto and Jago's strategy to avoid the compositionality objection is compromised.

To better understand this point and see what happens with the functions for complex expressions, consider a simple toy model with only four object-language sentences:

(1) p

(2) q

(3) $p \wedge q$

(4) $q \wedge p$

To each of them will correspond a unique worldmaking translation (for the reasons discussed in the previous section). In particular, (3) will correspond to $p^* \frown \wedge^* \frown q^*$ and (4) to $q^* \frown \wedge^* \frown p^*$, so that $3^* \neq 4^*$ (worldmaking strings are sensitive to the order of their constituents). However, by commutativity of 'and', (3) and (4) say the same, so [3] = [4] according to any semantically adequate IWS model.

The second step to build an adequate model is to generate all the open worlds out of the worldmaking sentences, which gives us the following world space W:

w_1: {1*}	w_2: {2*}	w_3: {3*}	w_4: {4*}
w_5: {1*,2*}	w_6: {1*,3*}	w_7: {1*,4*}	w_8: {2*,3*}
w_9: {2*,4*}	w_{10}: {3*,4*}	w_{11}: {1*,2*,3*}	w_{12}: {1*,2*,4*}
w_{13}: {1*,3*,4*}	w_{14}: {2*,3*,4*}	w_{15}: {1*,2*,3*,4*}	w_{16}: $\varnothing$

Since (3) and (4) have the same meaning and, more generally, we want to restrict the worlds to satisfy the above-mentioned principles, we have to exclude from the model all those worlds that have only one between 3* and 4* as members. This gives us the following restricted world space:

$$W^- : \{w_1, w_2, w_5, w_{10}, w_{13}, w_{14}, w_{15}, w_{16}\}.$$

As desired, the model makes [3] = [4] = $\{w_{10}, w_{13}, w_{14}, w_{15}\}$. But clearly, it also makes $\bigcap$[3] = $\bigcap$[4] = {3*,4*}, which is, unsurprisingly, not a singleton.

Now, let us turn to the functions for the connectives. In order to have a compositional model in Berto and Jago's sense, $f(\wedge^*,[1],[2])$ must be identical to [3] (the content of $p \wedge q$ must depend on the contents of p and q). We have that [1]=$\{w_1, w_5, w_{13}, w_{15}\}$ and [2]=$\{w_2, w_5, w_{13}, w_{15}\}$, while $\bigcap$[1] = {p*} and $\bigcap$[2] = {q*}. To get $f(\wedge^*,[1],[2])$ – i.e. $\{w \in W^- \cup \{x \frown \wedge^* \frown y \mid x \in \bigcap[1], y \in \bigcap[2]\}\}$ – we just need to add $p^* \frown \wedge^* \frown q^*$ to each world in W^-. The resulting set is $\{w_6, w_8, w_{10}, w_{11}, w_{14}, w_{15}\}$. Not only this is not identical to [3], so we have lost compositionality, but w_6, w_8 and w_{11} are not even members of the restricted (and semantically adequate) world space W^-.

Summing up, any semantically adequate IWS model based on Berto and Jago's strain of linguistic ersatzism is not compositional, or at least not in the way in which (semantically inadequate) complete IWS models can be said to be.

3.5 Some More Issues

Leaving aside the more technical issues concerning semantic adequacy and compositionality, this specific version of IWS seems to face more general worries as well.

The first challenge concerns the choice of a Lagadonian worldmaking language. While Berto and Jago do not claim that this is the only option for an ersatzist, they stress how crucial the connection between the worldmaking language and reality should be for this kind of approach (p. 182):

> these [worldmaking sentences] need not be sentences of the object language. Indeed, it's best they're not, else we make little progress in connecting the object language to reality.

Clearly, we shall not choose the language for which we are trying to provide a semantics as our worldmaking language. But as long as our goal is to provide an objectual account of semantic content, the same applies to any other natural language. Instead, opting for a Lagadonian language seems to ensure the desired connection between language and reality.

A Lagadonian language seems to work smoothly as long as we consider only atomic sentences. Indeed, in setting up a standard PWS account, Lewis (1986) argues that the simplest way to do that only requires worldmaking translations for atomic formulas (which amounts to building worlds only out of possible individuals, properties and relations). In such a setting, the negations of atomic formulas of the object language do not need an explicit translation: a negated formula $\neg A$ will be true in a world w if and only if w does not contain the worldmaking translation of P (recall that possible worlds are maximally consistent). The same goes for the other connectives: for instance, $A \wedge B$ will be true at w if and only if w contains both A^* and B^*.

Things are obviously not that easy when we switch to ersatz IWS. Impossible worlds are unruly entities, and we cannot rely simply on what they do or do not contain to get a faithful semantics. More precisely, we have seen how, in order to get a compositional framework, Berto and Jago need explicit worldmaking translations for each of the connectives. If we choose to work with a Lagadonian language, this implies that we need to find real objects in the world to represent them. The most obvious option is then to identify Boolean connectives with functions from truth values to truth values, and quantifiers with higher-order functions (both standardly understood as sets). But, as rightly pointed out by Silva (2021), what about *variables*? It seems that we need worldmaking translation for them too in order to attain the minimal expressive power required to account for quantified sentences, but it is not clear what they should correspond to in the world.

The same applies to some of the most common sentential operators. For instance, suppose that we want the semantics to be able to deal with alethic modal ascriptions. This amounts to including the operators of possibility and necessity in the framework, so we will need worldmaking translations for them. If we understand them standardly as quantifiers over possible worlds, it seems that in order to include them in worldmaking sentences, we first need to have the objects over which they range (i.e. possible worlds). But since possible worlds are defined as sets of worldmaking sentences, the sentences out of which they are built must not themselves contain modal operators, on pain of circularity. A possible solution to this issue might be that of conceiving the construction of worldmaking sentences (and thereby worlds and propositions) as an iterative generation that happens in stages, giving rise to a cumulative hierarchy (I will say more about this idea in the next chapter). However, this would still not help to provide an answer to the question about worldmaking translations for variables, and more generally, it would put a further layer of technical complexity to an already quite convoluted account.

Another point that one may raise has to do, again, with Berto and Jago's suggestion of limiting the range of admissible open worlds in order to satisfy the principles

for synonymy listed above. Regardless of the technical problems with recapturing compositionality discussed in the previous section, one may reasonably suspect that such a move is ultimately equivalent to changing semantics altogether. The benefit of introducing open worlds is that they provide a very fine-grained account of content, allowing for all sorts of hyperintensional distinctions. But, as extensively discussed, this degree of granularity is not apt to model same-saying in natural language (not to mention that structured accounts with the same degree of granularity were already available). Therefore, imposing the required algebra of content to fit the linguistic data seems to defeat the whole purpose of open worlds: why should we bother with them in the first place, if drastic revisions will be needed anyway? If the basic domain of the framework needs to be replaced in order to satisfy certain principles, why don't we simply replace the whole semantics with another that naturally satisfies those principles "for free"?

Berto and Jago consider a similar objection (pp. 210-211):

> It seems that the notion of a world, in and of itself, is doing little theoretical work in this approach. All the work is done by selecting the right algebra. For given that algebra, it doesn't really matter what kind of entity the c_1s and c_2s are. Formally, they could be sets of root vegetables, and the approach would work just as well as if they were sets of worlds. The objection then continues: surely a better approach is to select some kind of conceptual tool which *generates* the appropriate algebra, rather than being generated by it? Jago (2022) presents an alternative approach, on which same-saying contents are *sets of truthmakers*. [...] The objection, however, is that understanding contents in terms of truthmakers generates the right results automatically, without having to impose further restrictions (in the form of closure conditions) on worlds.

They proceed to address this by highlighting the relative advantages of ersatz IWS over TMS. Notice, however, that the present objection is more general: I am not (yet) proposing TMS as a better alternative to IWS. Instead, I claim that there is no point in holding on to IWS (for semantic content) *regardless* of the alternative accounts on the market, as long as major revisions on the basic framework (which, recall, do not naturally preserve compositionality) are needed anyway.[15]

3.6 Conclusion

I this chapter I sketched and defended an externalist, non-representational picture of the notion of semantic content. This allowed me to discuss one of the most developed accounts of hyperintensional propositions, which derives from a specific semantic framework: ersatz IWS, as defended by Berto and Jago (2019).

I showed how the mechanism they adopt to recapture compositionality makes the account semantically inadequate with respect to same-saying in natural language, and

15. I am indebted to Matteo Plebani for this last point.

that the restriction on admissible open worlds suggested to avoid the problem makes their mechanism to recapture compositionality useless.

While this by itself does not entail that restricting open worlds makes the account incompatible with *any* form of compositionality, I believe that this problem, combined with some more general related worries, is more than enough to shift the burden of proof: the ersatzists need to find a new way to deal with both the compositionality objection and the semantic data in order to make IWS a viable account of semantic content.

Until then, if we grant that Berto and Jago's account is the most developed version of IWS on the market, IWS as a whole approach to formal semantics might be in serious trouble, at least as a candidate to account for same-saying in natural language.

Chapter 4

Hyperintensional Propositions II: Paradoxes

4.1 Background

The present chapter aims to explore how some hyperintensional accounts of propositions deal with two famous paradoxical results lurking in the shadows: the Russell-Myhill paradox (Russell (1902), Myhill (1958)) and Kaplan's paradox of propositions (Kaplan (1995)). Since both of these trade heavily on the possibility of quantifying over propositions (in a sense that needs to be clarified), a few introductory remarks about the debate over higher-order quantification are needed.

It is nowadays part of the received view that systems of higher-order logic are well-understood and useful pieces of formalism in the logician's toolkit, as testified by the wide range of applications they display and their long-standing tradition in the literature. For the current purposes, we will be especially interested in extensions of propositional languages with propositional quantifiers, which are standardly understood as quantifiers that bind variables in sentence position. Even if we will not need to work within a fully spelled-out formal system (formulas will be introduced just for the ease of exposition, when it is not explicitly stated otherwise), it is worthy to stress that, as shown by Fritz (2024) the extension of common models of propositional logic to models of propositional logic with propositional quantifiers is straightforward.[1]

A more controversial issue is the one concerning the informal interpretation of this kind of higher-order quantification. An opinion shared by some authors is that there is no such thing as the natural-language correspondents of higher-order formulas quantifying over predicates or propositions. In light of this, some authors claim that higher-order languages *augment* the expressive power of natural languages. Furthermore, this leads

1. Fritz (2024) presents also a model-theoretic interpretation of propositional quantification as quantification over arbitrary sets of possible worlds, which turns out to be exactly the sort of quantification needed for the current purposes. However, since I am considering the issues surrounding hyperintensional propositions independently of any specific formal setting other than the basic semantic assignments of the entities chosen to represent semantic contents to sentences, there is no need for a detailed exposition here.

many authors to the conclusion that there is a fundamental distinction between propositional quantification and standard first-order quantification. This gets reflected in some difficulties, clearly illustrated by Fritz (2024), that one encounters in interpreting certain formulas that quantify over propositional variables as if the quantification were first-order. Regarding this, he writes (p. 11):

> formal languages with propositional quantifiers may be poor models of natural language constructions. But this poverty in faithfulness need not be seen as a disadvantage. On the contrary, when using logics to improve on natural languages, such discrepancies are essential. In the present example, the formal language helps to distinguish individuals from the kinds of entities which can be believed.

This suggests a normative interpretation of the results of higher-order logics, not only with respect to what inferences can be performed in specific formal settings, but also with respect to what kinds of entity can be quantified over in natural language, in order to distinguish genuine individuals from constructions with a different ontological import.[2]

But, precisely in virtue of such discrepancies, it seems at the very least contentious whether certain features of formal systems can be projected onto natural language, especially if this is done in order to draw substantive metaphysical conclusions about propositions (as part of the higher-order metaphysics movement seems to suggest – see Skiba (2021) for an overview). In particular, the fact that propositional quantification *in a formal setting* may not be first order should not have any consequences on the type of quantification at stake when we utter perfectly well-formed and meaningful sentences in a natural language which *prima facie* quantify over entities like propositions. Notice that this becomes almost pleonastic if we are committed to an objectual account of propositions in the first place: if propositions are entities (let us say, some kind of set), they should already be part of our domain of objects. Therefore, propositional quantification (in natural language) becomes just ordinary quantification over a restriction of the domain.

But even if we do not want to assume an objectual picture of propositions, there seem to be independent reasons not to tie quantification over proposition in a natural language to higher-order quantification, as extensively argued by Hofweber (2022), pp. 38-39:

> in English we can almost always nominalize and put things into subject position and thereby make it available for interaction with quantifiers into subject or nominal position. The most famous case of this is Frege's concept of a horse problem: if predicates stand for concepts, while singular terms stand for objects, then 'the concept of a horse' seems to

2. A heavily theory-laden approach of this kind reminds the intensionalist stance towards a class of linguistic data about semantic distinctions, which according to the orthodoxy need to be explained away: this amounts to a normative interpretation of intensionalist semantics. A more data-driven approach should be reasonably suspicious of this idea, and the same applies to the idea of interpreting the expressive limitations of higher-order languages in a normative way.

> stand for an object, not a concept. And it seems to interact with just the same quantifiers that other singular terms interact with. Nominalizing thus seems to support a single domain over which all English quantifiers range. [...] Quantification over facts and propositions in English is not higher-order quantification.

If we take linguistic data seriously – just like we do when we try to develop an appropriate account of semantic content – we better not cherry-pick. The discrepancies between manifest natural language phenomena and features of higher-order systems may cast some serious doubts on the possibility of drawing consequences about the former in virtue of the latter (recall the motto of the first chapter: we shall not mistakenly attribute features of the formal frameworks to reality). To borrow Hofweber's words again (pp. 48-49):

> HOM [higher-order metaphysics] is ultimately based on a different conception of the place of formal languages in metaphysics. On this alternative approach, natural language is taken to be restrictive, flawed, and limited. This limitation can be overcome by endorsing HOL [higher-order logic] as an improved alternative to, or augmentation of, natural language. Once we have taken this step, so the idea, we can then articulate and achieve novel results that were elusive before. We can call this the *expressive improvement* conception of formal tools. They allow us to do and say things that so far were off limits. [...]. But the lower vs. higher-order distinction first and foremost applies to formal languages. It applies to natural languages only derivatively, where it can be seen either as a syntactic or a semantic distinction. [...] If it is a semantic one, then the issue could either be whether we can quantify over properties and propositions, for which the answer is clearly 'yes', or whether such quantifiers are a distinct kind of quantifier in natural language, for which the answer [...] is 'no'. But primarily the lower vs. higher-order distinction applies to formal languages, where there is a clear syntactic and semantic difference between lower and higher-order languages. This way moving from lower-order languages to higher-order gives rise to an expressive improvement in the technical sense discussed above, but that again only compares formal languages. It does not mean that adding a higher-order formal language to natural language leads to expressive improvement in the non-technical sense.

If we grant this, we can safely concede that natural language is, in general, a good guide to propositional quantification, and this vindicates the basic intuition behind objectual approaches to propositions.

Moving on to our main topic, any objectual approach to propositions – according to which propositions are real entities that can flank the identity sign and be quantified over – shall face some technical challenges arising when such entities are fine-grained enough to meet the hyperintensional needs discussed above (even if we set aside the more general worries about the interaction between natural and formal languages). Among these challenges, the Russell-Myhill paradox has gained a special status. The main reason for this is nicely captured by the words of Uzquiano (2015): «Russell's paradox

[of propositions] is best regarded as a limitative result on propositional granularity»(p. 1).[3]

The paradox affects the original Russellian account of propositions as structured entities, but it is easily generalized for more recent accounts of structured propositions. Following Uzquiano's reconstruction, the paradox relies on three principles commonly endorsed by structuralists:

(a) If m is a set of propositions, there is a proposition of the form 'every proposition in m is true' (let us call it the *logical product* of m).

(b) If m and n are different sets of propositions, then the proposition 'every proposition in m is true' is different from the proposition 'every proposition in n is true'.

(c) There is a set w of all and only propositions of the form 'every proposition in m is true, for some set of propositions m to which the proposition does not belong'. Formally, $w = \{p \mid \exists m(p = \forall q(q \in m \rightarrow q) \wedge \neg(p \in m))\}$.

With these assumptions in place, we can briefly illustrate the problem as follows: by (a), there is a proposition $r = \forall q(q \in w \rightarrow q)$, which says that every proposition in w is true. In other words, r is the logical product of w. Now, either $r \in w$ or $r \notin w$.

Suppose that $r \in w$. Then, r is a proposition that belongs to the set of which it is the logical product. But since w is the set of propositions that do not belong to the set of which they are the logical product, $r \notin w$.

Now suppose that $r \notin w$. Then, r is a proposition that does not belong to the set of which it is the logical product. Hence, $r \in w$.

A less common way of illustrating the paradox is through the set-theoretic implications of the idea that propositions are structured entities identified by their constituents. In particular, we can show that the structuralist principles lead to a violation of Cantor's Theorem, which says that the cardinality of the *powerset* of any set A – i.e. the set of all subsets of A – is strictly greater than the cardinality of A.

Let us call P the set of all propositions. Principle (a) says that each subset of P has a logical product. Propositions are identified with their constituents, so by principle (b) we know that the logical product of each subset of P cannot be the logical product of any other subset (each logical product contains a distinct subset of P as a constituent). Since logical products are themselves propositions (and thus members of P), this means that there must be at least as many unique logical products as sets of propositions. In other words, the cardinality of P must be at least as big as the cardinality of its powerset: this violates Cantor's Theorem.

One might be tempted to resist the paradoxical inference by rejecting the idea that there is a set of all propositions. However, the core of the problem seems independent from the adoption of a set-theoretic setting, as shown by McGee and Rayo (2000) and

3. It is worthy to notice that even if I will focus on a first-order interpretation of propositional quantifiers, Sbardolini (2021) shows how a variant of the Russell-Myhill paradox can be derived even with a higher-order interpretation.

Deutsch (2014), which illustrate variants of the paradox for a plural setting and a class-theoretic setting, respectively. Instead, the adoption of a structured (or an equally fine-grained) picture of propositions seems to be the real source of the issue.

Since in the present context we are interested in truth-conditional accounts of propositions, why should we bother about this result? As remarked in the previous chapter, the most developed truth-conditional semantics based on impossible worlds on the market has some huge analogies with structured accounts. Indeed, in the previous chapter I have shown how some of the problems concerning the excess of granularity that affects structured accounts also affect ersatz IWS. What about the Russell-Myhill paradox?

4.2 Russell-Myhill for Ersatz IWS

We can easily illustrate how a variant of the Russell-Myhill paradox haunts ersatz IWS. Recall that, according to this framework, propositions are sets of open worlds, and open worlds are sets of sentences of the worldmaking language. These can be identified with Russellian tuples. As long as the worldmaking translations of sets of propositions (the sets themselves, if the worldmaking language is Lagadonian) can figure as members of the Russellian tuples out of which open worlds are built, we will have semantic values of logical products: for each set of propositions A, its logical product will be the set of open worlds containing the worldmaking translation p^* of $p = \forall q(q \in A \rightarrow q)$, namely "all propositions in A are true". Principle (b) is then satisfied by the fact that each set of propositions is a distinct object, so there will be a distinct logical product for each set of propositions (moreover, recall that since $\{p^*\}$ is itself an open world, $\bigcap[p]=\{p^*\}$).

For a Cantorian diagnosis, let us call W the set of all open worlds. Since every set of open worlds is a proposition, the set of all propositions is $\wp(W)$, namely, the powerset of W. Logical products are propositions, so they must belong to $\wp(W)$. But as long as there is a distinct logical product for each set of propositions (in other words, as long as worldmaking sentences are structured) there must be at least as many logical products as sets of propositions, just like in the original setting. This means that the cardinality of $\wp(W)$ must be at least as big as the cardinality of $\wp(\wp(W))$, namely, the set of all sets of propositions. This violates Cantor's theorem.

At this point, one might ask: why should the ersatz IWS theorist concede that sets of propositions can themselves be members of worldmaking sentences? On the face of it, that seems to be a natural consequence of adopting an objectual approach to propositions: the account is committed to the existence of sets of open worlds to represent the semantic values of sentences, so, it seems natural to have worldmaking translations for sets of them (and thereby, it seems possible to build worldmaking sentences containing such sets). This becomes particularly noticeable when the worldmaking language is Lagadonian: in that case, we get the translations "for free". Moreover, we *can* talk about sets of propositions, and it is reasonable to expect an appropriate semantic framework to be able to track hyperintensional distinctions

among sentences about them.[4] If we imposed a restriction on what sort of entities can figure within a worldmaking sentence just to avoid the paradox, we would be making an *ad hoc* move which, at the same time, would drastically limit the expressive power of the semantics.

A different strategy to avoid the problem might be that of conceiving the generation of propositions as something that happens in stages, as for the generation of sets according to the iterative conception.[5] After all, one might object that in order to build logical products, we first need worldmaking translations for sets of propositions. But in order to have them, we first need an initial stock of propositions. These are identified with sets of open worlds, i.e. sets of worldmaking sentences. Therefore, the way to avoid impredicativity is by assuming that the worlds of the initial stock of propositions must not contain worldmaking sentences that already contain (worldmaking translations of) propositions (or sets of them). This represents a particular instance of the more general idea that quantification over propositions which are about some objects depends on the prior availability of those objects.

In order to get the worldmaking translations of logical products, we first need to generate all the sets of open worlds that do not contain them. In other words, at the first stage resulting from this cumulative generation we will only find sets of open worlds containing worldmaking translations of sentences that are not about propositions (or sets of them). By iterating the procedure we generate new stages, and at each of them new propositions that can only be about those found at the lower stages.[6]

Against the background of a cumulative hierarchy of this kind, there is no violation of Cantor's theorem and the paradox is neutralized. However, this would come at a huge cost: we would not be able to quantify over absolutely every proposition anymore. This seems particularly unfortunate if we hoped to employ ersatz IWS as an account of meaning for natural language. Furthermore, even if in this setting there would be no set of all propositions, the paradox can still be derived if we conceive the resulting totality of propositions as a proper class or a plurality. As stressed by Uzquiano (2015): «the thesis that there is no set of all propositions is ineffectual as a response to Russell's paradox of propositions» (p. 1). In other words, the problem will still show up, as long as the basic structuralist assumption survives: this is the real source of the problem.[7]

4. Notice that logical products do not seem to belong to the kind of artificial sentences that we get when we try to paraphrase into natural language formulas of second or higher-order logic: there is nothing linguistically weird about a sentence like "all propositions in A are true".

5. See Boolos (1971).

6. An approach of this kind has been proposed by Whittle (2022) precisely to provide a solution to the Russell-Myhill paradox, in light of how the iterative conception of sets provides a solution to the original Russell's paradox. However, he is well aware of the expressive limitations that come with adopting the hierarchy. It remains unclear whether the same strategy can actually be applied to ersatz IWS. The same goes for the modal iterative account proposed by Yu (2017).

7. A slightly different take on the issue is that of Sbardolini (2021), according to which paradoxes of the Russell-Myhill family are ultimately to be understood as *aboutness* paradoxes. If that is the case, satisfying hyperintensional principles even weaker than those endorsed by structuralists is enough to get the paradox. More precisely, any semantics that satisfies a minimal principle of hyperintensional synonymy such that if ψp is synonymous with ϕq , then $p = q$ (plus a reasonable comprehension schema for propositions),

A detailed discussion of this strategy would be outside the scope of the present work. However, what illustrated so far seems more than enough to declare that the Russell-Myhill paradox represents a serious challenge to ersatz IWS.

Maybe, the key to develop a better hyperintensional account of propositions is to abandon the structuralist assumptions incorporated by ersatzism, in favor of a different version of IWS. However, as we are about to see, there is a more general problem concerning hyperintensional propositions that seems to affect also the other strains of IWS.

4.3 The Intensional Kaplan's Paradox

Kaplan (1995) claimed that standard PWS for a propositional modal language extended with quantifiers, identity and propositional variables conflicts with a somewhat intuitive principle about propositions:

(K) $\forall p \Diamond \forall s(Qs \leftrightarrow p = s)$

Since p and s are propositional variables, (K) says that for every proposition p, there is at least one possible world where p alone has the property Q. An informal example of Q, proposed by Davies (1981) and Lewis (1986), is the property of 'being thought by someone at time t'. If Q is assigned this meaning, then (K) amounts to the claim that every proposition can be the only proposition that is thought by someone at a given time t.

The intuitive acceptability of (K) is challenged by the following argument:

(1) There is a set W of all possible worlds.

(2) p is a proposition := $p \in \wp(W)$, namely, each subset of W is a proposition.

(3) Modal operators are quantifiers that range over W.

(4) Propositional variables and quantifiers range over $\wp(W)$.

(5) For every p, there is at least one $w \in W$ where p alone has Q.

(6) $\therefore |W| \geq |\wp(W)|$.

Premise (5) comes from accepting (K), and together with (1)–(4) entails (6), which says that there are at least as many worlds as propositions. However, (6) violates Cantor's theorem. To quickly see this, suppose that the cardinality of W is $|W|$. Then,

is doomed. This means that even *prima facie* "unstructured" accounts such as those developed by Yablo (2014) and Fine (2017b) will be affected by some variant of Russell-Myhill.

the cardinality of $\wp(W)$ is $2^{|W|}$. But if we accept (1)–(5), we have to conclude that the cardinality of W is at least $2^{|W|}$, which contradicts our assumption.

According to Kaplan, the source of the problem is to be found in PWS itself. He stresses that (K) strikes us as an intuitive truth about propositions, so we should give up on the framework that falsifies it. In assessing Q, he writes: «I can think of no plausible reason why logic itself should refute the existence of such a property» (pp. 42-43).

In light of this, Kaplan proposes to drop the picture of propositions associated with possible worlds semantics, in favor of a *ramified* account. I will not dive into the technical details here, but the basic idea behind ramification is to impose *orders* on propositions, namely, some hierarchy that forbids to speak of 'all' propositions in the unrestricted sense – which amounts to denying at least premise (4) of the paradoxical argument. Such a move requires to complicate the framework quite drastically. Furthermore, as noticed by Anderson (2009), it faces the threat of being self-refuting: if we cannot speak about all propositions, then we should not be able to say that *all* propositions have orders.[8]

On the other hand, Lewis (1986) argues that we should not even bother with properties like Q. Sticking to its informal reading, he claims that it is simply not the case that any set of possible worlds is the content of some possible thought. So, there cannot be such a property, and (K) is to be rejected (p. 105):

> most sets of worlds, in fact all but an infinitesimal minority of them, are not eligible contents of thought. It is absolutely impossible that anybody should think a thought with content given by one of these ineligible sets of worlds.

One might object that such a reply is effective only in showing that Q cannot be a property about our thoughts, or some other propositional attitude, due to our limited cognitive faculties – obviously, we are not able to grasp the semantic content represented by every single set of worlds. But this fact alone does not rule out the existence of alternative readings of Q that do not involve propositional attitudes.

While rejecting (K) by forcing a specific reading on Q might be a questionable move, it is important to highlight that there is no real need for an informal interpretation of that property. Indeed, regardless of our *prima facie* intuitions towards the acceptability of (K), a result by Bueno, Menzel, and Zalta (2014) shows that (K) is actually a logical falsehood within the same framework adopted by Kaplan. By introducing a quantified version of the T-axiom $\forall p(\Box p \rightarrow p)$ and a simple comprehension schema for propositions $\exists r(r \leftrightarrow \varphi)$ for any formula φ of the language in which r does not occur free, the following turns out to be a theorem:

$$\text{(NK)} \qquad \exists p \Box \neg \forall s(Qs \leftrightarrow p = s)$$

We can read (NK) as follows: for some proposition p, it is necessary that p does not uniquely have Q.[9] This is obviously equivalent to the negation of (K), as we can easily

8. See Bacon, Hawthorne, and Uzquiano (2016) for a detailed discussion of the ramified approach in relation to Kaplan's paradox.

9. See Bueno, Menzel and Zalta (2014), p. 25, footnote 40, for the full proof.

show through the interdefinability between quantifiers and modal operators: (NK) is equivalent to $\exists p \neg \Diamond \forall s(Qs \leftrightarrow p = s)$, which is in turn equivalent to $\neg \forall p \Diamond \forall s(Qs \leftrightarrow p = s)$. A similar argument can be found in Anderson (2009), which presents a different proof of (NK).

While the debate has evolved to encompass various perspectives on the severity of the threat posed by Kaplan's paradox, for the current purposes it is sufficient to emphasize the availability of a purely logical solution to it. Indeed, the main strategies employed to avoid the paradox without dropping the possible worlds account of propositions tend to agree with Lewis in denying (K), i.e. step (5) of the argument presented above. This saves the standard account of modal quantification and the identification of propositions with sets of possible worlds – maybe at the expense of some *prima facie* intuition.

4.4 The Hyperintensional Kaplan's Paradox

Earlier in the chapter we introduced IWS as a way to account for hyperintensional propositions. However, we have seen how the most developed IWS on the market relies on a form of structuralism that ultimately leads to a paradoxical outcome. We may then drop Berto and Jago's strain of ersatzism in favor of a different version of IWS, in order to avoid the structuralist commitments while retaining the idea that propositions should be identified with sets of possible and impossible worlds.

Kaplan's paradox, on the other hand, affects an *unstructured* intensional account of propositions. Since the derivation of the paradox does not rely on any structuralist assumptions, it may be taken to represent a more general threat than the Russell-Myhill paradox.[10] Nevertheless, a purely logical solution to the paradox seems available for the standard PWS picture of propositions. Now we may ask: is it still the case if we extend this picture as conceived by proponents of IWS (even *without* assuming a structuralist background)?

Yagisawa (2010) is the first to openly address the issue.[11] Suppose that we extend the semantics of our modal propositional language with impossible worlds (we can still conceive them as open worlds, in order to get the most expressive models). This brings a new implicit commitment for the evaluation of impossible statements: $\neg \Diamond \varphi$ will still be true if and only if φ is false at every *possible* world, but now, in any model fine-grained enough to meet our hyperintensional needs, there must be also some *impossible* world at which φ is true. Moreover, even if it is still not the case that, for every proposition p, it is possible that p uniquely has Q – as established by proving (NK) – a weaker principle must remain true:[12]

10. At the same time, Kaplan's paradox may very well be considered *less* general than the Russell-Myhill paradox, since only the former relies on the interaction with modal operators.

11. Kripke (2011) also seems to be aware of this. In assessing the original Kaplan's paradox, he writes: «if someone has a more fine-grained notion of proposition than a set of possible worlds, this only makes the problem worse» (p. 374).

12. (Y) is my own formalization, Yagisawa (2010) only states the principle informally.

(Y) $\forall p(\Diamond\forall s(Qs \leftrightarrow p = s) \vee \neg\Diamond\forall s(Qs \leftrightarrow p = s))$

That is: for every proposition p, it is either possible or impossible that p uniquely has Q. In other words, for every proposition p, there is at least one world – either possible or impossible – where p alone has Q (notice that, while this particular reading is enabled only by the presence of impossible worlds in the semantics, (Y) should be valid even in a standard PWS framework).

Now, let W be the set of all possible worlds and I be the set of all impossible worlds. It seems that we can build a new paradoxical argument as follows:

(1_i) There is a set $W \cup I$ of all worlds.

(2_i) p is a proposition := $p \in \wp(W \cup I)$.

(3_i) Modal operators are quantifiers that range over $W \cup I$.

(4_i) Propositional variables and quantifiers range over $\wp(W \cup I)$.

(5_i) For every p, there is at least one $w \in W \cup I$ where p alone has Q.

(6_i) $\therefore |W \cup I| \geq |\wp(W \cup I)|$.

This time, the paradox-triggering premise (5_i) comes from accepting (Y). However, as stressed by Yagisawa, (Y) captures a truism, i.e. that a given fact about propositions is either possible or impossible: denying this on purely logical grounds does not seem a viable option. While the truth of (K) in the original setup was a contentious matter, (Y) seems relatively innocuous, so, if we want our framework to properly capture modal facts, we expect it to be valid.

Furthermore, in this case we cannot appeal to the alleged impossibility of Q as a way to sidestep the paradox. In order to see this, suppose with Davies and Lewis that Q is the property of 'being thought by someone at time t'. If it is not the case that any set of worlds gives the content of a *possible* thought, as Lewis claims, then any proposition that *cannot* be thought is thought at some impossible world. But then, for every proposition p, there must be at least a possible or an impossible world at which p is thought by someone at time t. In other words, here the alleged impossibility of Q would be captured precisely by the existence of impossible worlds at which Q is instantiated by some proposition. (Y) adds a uniqueness constraint on the instantiation of Q regardless of its modal status, so (Y) cannot be rejected simply by stressing that Q is impossible.

Let us now turn to a potential reaction on behalf of friends of impossible worlds. One might stipulate that, at each impossible world where Q is uniquely instantiated, Q is both uniquely instantiated *and not* uniquely instantiated (by some proposition p). This amounts to allowing for more than one proposition to uniquely have Q at the same

impossible world. After all, impossible worlds are not required to be closed under any logical rule, so we do not need to concede that there must be at least one world for each proposition – a single impossible world may account for the "unique" instantiation of Q by more than one proposition (or even all of them). This should suffice to block the paradox.[13]

But now, recall that as long as we take (NK) to be a metaphysical truth about propositions, it entails the existence of propositions that, necessarily, do not uniquely have Q. Suppose that p_1 is such a proposition and consider the following pair of sentences:

(a) p_1 uniquely has Q.

(b) p_1 uniquely has Q and p_1 does not uniquely have Q.

Both (a) and (b) are necessarily false, and they will be true at some impossible worlds. But if we adopt a model that satisfies the constraint described above (if a proposition p uniquely has Q at some impossible world w, then p both uniquely has Q and does not uniquely have Q at w), (a) and (b) turn out to be true at exactly the same impossible worlds. Thus, according to such a model, they would express the same content.

However, while (a) and (b) may be intensionally equivalent, they clearly express two different propositions, and impossible worlds were introduced precisely for tracking hyperintensional distinctions – including those between necessary falsehoods. In other words, if we want to keep the semantic distinction between sentences like (a) and (b), we want them to be identified with two different sets of worlds. But if we introduce the constraint to avoid the paradox, the model lacks the expressive power to account for such distinctions. Since we have no independent reason to weaken the framework in a way that goes explicitly against the main point for introducing impossible worlds, exploiting their bizarre nature to place such constraints looks like an *ad hoc* move.

At this point, the friend of impossible worlds might either bite the bullet and claim that sentences like (a) and (b) do in fact express the same proposition, or go against the received view and attempt to develop a framework that invalidates (NK), in order to allow sentences like (a) to be true at some *possible* worlds.

Summing up, no matter what model-theoretic preferences one might have, (1_i)–(6_i) seems to constitute a serious challenge to the analysis of propositions as sets of possible and impossible worlds associated with IWS. In what follows, we will explore some alternative strategies to avoid this new version of Kaplan's Paradox (HKP from now on), without giving up on a hyperintensional picture of propositions.

13. I am indebted to Matteo Plebani for the suggestion of this possible reply.

4.5 Giving up on Sets

The only explicit attempt to avoid HKP is presented by Yagisawa (2010). His proposed solution consists in denying that it is possible to speak of a set of all propositions (p. 238):

> I reject the absolutely unrestricted sense of a quantifying expression. [...] It is improper to speak of the collection of all propositions in the absolutely unrestricted sense of 'all'. We may not suppose the cardinality of such a collection to be anything at all.

In other words, Yagisawa's solution amounts to denying at least (4_i): if there is no set of all propositions, then propositional quantifiers do not range over $\wp(W \cup I)$ and the paradoxical inference is blocked (notice that, if Yagisawa rejects *any* use of 'all' in the unrestricted sense, then according to him it is also improper to speak of a set of all worlds, which would amount to denying (1_i) as well).

The prohibition on speaking of a set of all propositions seems to be primarily motivated by worries concerning the problem of absolute generality.[14] One might think that the domains at stake here – that of worlds and that of propositions – are *restricted* generalizations, namely, universal generalizations about absolutely every member of a comparatively limited kind.*Prima facie*, such generalizations do not seem to quantify over everything in the absolutely unrestricted sense, which is how quantification in the debate on absolute generality is typically framed.[15] But as we are about to see, this is not a real concern for the current purposes.

Instead of speaking of the set of all propositions, Yagisawa proposes to restrict the range of propositional quantification to specific collections of propositions, defined by the *type* of content that they express (e.g. a collection T_1 of propositions with non-modal intentional content, a collection T_2 of propositions with modal intentional content, and so on). Then, he claims that such collections of different types of propositions shall not be mixed together to form a single collection. Yagisawa does not provide independent reasons to justify the hierarchy, and it is easy to notice how his move resembles ramification. Indeed, an analogous objection might be moved against it: if we cannot speak of all propositions, then we cannot say that all of them have types.

One might nevertheless reply that we can speak of the properties of each member of a collection without speaking of the whole collection as a single entity (a set or a class). In order to achieve this, an obvious option is to employ plural quantification. Yagisawa does not explicitly take this path, but it is interesting for the current purposes to explore how one might use pluralities to solve HKP. So, let us try to sketch a plural approach to the problem and check whether it fares better than its set-theoretic counterpart.

Plural quantifiers do not bear a commitment to sets. Instead, they range over *pluralities* of entities. Assuming a plurality *ww* of all worlds, both possible and impossible, each sub-plurality of *ww* will represent a proposition. We can now

14. See Yagisawa (2010), p. 55.

15. Nevertheless, Williamson (2003) argues that even restricted generalizations ultimately need to quantify over absolutely everything.

replace our standard propositional quantifiers and modal operators with plural propositional quantifiers and plural modal operators. The latter will range over *ww*, while propositional quantifiers and variables will range over the sub-pluralities of *ww*, which taken together form the super-plurality *wwww*.[16]

Even with our new plural domains in play, we can stick to (Y) as an adequate formalization of the principle that, for every proposition *p*, there is at least one world where *p* alone has *Q*.[17] The new setup enables the following rephrasing of the paradoxical argument:

(1_p) There is a plurality *ww* of all worlds.

(2_p) *pp* is a proposition := $pp \preceq wwww$, namely, *pp* is a sub-plurality of *ww*.

(3_p) Modal operators range over *ww*.

(4_p) Propositional variables and quantifiers range over *wwww*.

(5_p) For every proposition *pp*, there is at least one world where *pp* alone has *Q*.

(6_p) $\therefore ww \geq wwww$.[18]

At first glance, (6_p) is just as paradoxical as (6) and (6_i). But now one might argue that, since there are no sets involved in (1_p)–(5_p), cardinality concerns are not in play anymore, therefore we shall simply concede that there are in fact at least as many worlds as propositions.

Such a conclusion would be too hasty. Florio and Linnebo (2021) present two proofs of a plural equivalent of Cantor's theorem, according to which for any plurality with two or more members, its sub-pluralities are strictly more numerous than its members. More precisely, given our plurality of worlds *ww* and our corresponding super-plurality *wwww* of propositions, there is no surjection from the former to the latter and no injection from the latter to the former. But this means that the only scenarios where (1_p)–(5_p) does not lead to a contradictory outcome are those where there are fewer than two worlds in the model. Obviously, such models would not bear sufficient expressive power for any semantic purpose, so any minimally expressive plural model would eventually run into paradox. It seems that, even if we gave up on sets in favor of plural domains, a plural

16. A super-plurality is a higher-order plurality, or a *plurality of pluralities*. Despite its being a controversial notion (see Florio and Linnebo (2021) for an in-depth discussion), for the current purposes it is fit to capture the analogy with the notion of powerset in order to sketch a plural approach to the problem.

17. For the sake of argument, I assume that such a smooth transition from the standard framework to the plural one is available, but this is not obvious. A possible example in this direction is the account of Hewitt (2012), which adds modal operators to PFO+, i.e. the same framework in which Florio and Linnebo (2021) prove a plural version of Cantor's theorem.

18. This is meant to be read as follows: the plurality of worlds is equally or more numerous than the plurality of propositions. In other words, there must be at least as many worlds as propositions.

variant of HKP would nonetheless survive.

An alternative option in the spirit of abandoning sets might be assuming that, instead of forming a set, the collection of all propositions forms a proper class. If that is the case, it cannot be identified with $\wp(W \cup I)$, as long as $W \cup I$ is itself considered a set. Hence, the collection of all worlds should not be a set as well.[19] There might be even some independent reasons to believe this, as Kripke (2011) notices (p. 378):

> it seems to me to be reasonable to suppose [...] that for every cardinality κ it is possible that there are exactly κ individuals. But then it would immediately follow that the possible worlds cannot form a set.

The point obviously extends to the collection of all worlds, since it includes the collection of all possible worlds. We might then conclude that $W \cup I$ should be a proper class as well. This would require to adopt an unorthodox framework where proper classes are employed as domains of quantification, as long as we want to formally capture the full range of hyperintensional distinctions among propositions. One might then argue that this would simply shift the problem to a higher level of mathematical abstraction, just like it happens with the Russell-Myhill paradox when we deny that there is a set of all propositions, as stressed by Uzquiano (2015). I will not explore this issue further here.

One way or another, giving up on sets requires drastic and complex revisions to the framework, that do not even seem to guarantee a straightforward solution to the paradox. In light of this, a safer way of approaching the issue may consist in revising the ontology of worlds instead, as we shall see in the next section.

4.6 Giving up on Worlds

An easier way to avoid HKP may not lie in the disposal of sets, but in the disposal of *worlds*, either from the background semantic framework or from the account of propositions (assuming that an account of propositions can be given independently from a background semantics). The former approach amounts to denying (1_i) and, consequently, (2_i), (3_i) and (4_i). The latter amounts to denying at least (2_i) and (4_i).

4.6.1 Modal Quantification without Worlds

Recent attempts to develop accounts of modalities that are not committed to possible worlds have been made primarily for meeting the ontological worries associated with the standard accounts, in particular to avoid the heavy inflationary ontology of Lewisian modal realism (see Chihara (1998), Divers (2006), Kahle (2011), Dunaway (2013) and Vetter (2013)). The relative advantages with respect to the ontology of such accounts are often out-weighted by their expressive limitations. Furthermore, they do not always provide full-fledged semantic clauses for modal operators to be used in a formal setting, in particular for natural language.

19. This particular view is defended in Pruss (2001).

To see an account of this kind in more detail, I will briefly focus on the one proposed by Dunaway (2013). According to Dunaway, by introducing primitive quantification into predicate position, paired with a primitive hyperintensional connective, we can have the benefits of modal quantification without committing to worlds: a lighter ontology at the cost of a more complex ideology.

Within this picture, modal expressions like $\Diamond p$ are analyzed as second-order equivalents of "there is some possible way for things to be W, and things being W entails p". *Ways* are maximal predicates that might have been instantiated, and the entailment relation is a hyperintensional relation $\Rightarrow$ employed to avoid the notion of truth-at-a-world. It roughly says that, for a maximal predicate W to be instantiated, p has to obtain. So, the definitions of modal operators will be the following:

$\Diamond p :=$ $\exists W$ things are $W \Rightarrow p$.

$\Box p :=$ $\forall W$ things are $W \Rightarrow p$.

Quantifiers are treated as primitive in order to avoid any kind of ontological commitment – according to this analysis, there is no domain over which they range. This means that the maximal predicates adopted here do not specify a corresponding set of properties to ground the truth of the predications. This may sound counterintuitive, but Dunaway claims that there might be independent reasons to accept primitive second-order quantification, in addition to the advantage of discharging our modal talk from the commitment to problematic entities.[20] As far as HKP is concerned, such a proposal amounts to successfully denying (1_i) and thus avoiding the paradox.

Nonetheless, the need for a *primitive* hyperintensional connective '$\Rightarrow$' seems problematic. As clarified by Dunaway: «by 'hyperintensional' I mean simply substituting cointensional arguments of the connective does not necessarily preserve truth value» (p. 168). The intension of a piece of language, at least in the context of formal semantics, is usually defined as a function from *worlds* to objects. If Dunaway's understanding of hyperintensionality involves an explicit reference to intensions, how are we supposed to understand them in a world-free setting like the one he is proposing? They cannot be functions from maximal predicates, since such predicates do not define a domain of objects, as stressed by Dunaway himself. But appealing to a primitive notion of intensionality, which is contextually employed as a technical term, does not seem to be a viable option. Since hyperintensional resources are required in his own account of modal quantification, Dunaway should not adopt a world-theoretic notion of hyperintensionality in the first place, on pain of contradiction. Moreover, his definition of hyperintensionality appeals also to necessity, which is the very phenomenon that is supposed to be analyzed in terms of modal quantification. In other words, it seems that he may also have a problem of circularity.

20. More specifically, Dunaway appeals to the existence of Ostrich Nominalism, the Quinean idea according to which simple predication does not need to be analyzed in terms of properties. If that is the case, then second-order generalizations likewise can be free of such an analysis.

A further limitation of this proposal is that it cannot give an account of propositions in terms of sets of worlds, as explicitly recognized by Dunaway: «if we give up on quantification over worlds, we cannot say that worlds are constituents of propositions» (p. 166). This may not be a problem in its own right, but it means that the account is not suited for delivering a hyperintensional picture of propositions: it does not provide the resources to account for distinctions among co-intensional propositions, since such distinctions are already assumed in the analysis of modal operators.

Despite these obvious limitations, it is important to stress again that any successful revisionary account of modal operators should indeed be able to avoid the paradox, as long as the entities employed to account for modal quantification are not the same employed to identify propositions (or as long as modalities are not understood as quantifiers at all, as in the agnostic accounts proposed by Divers (2006) and Kahle (2011)). The real challenge is that of finding an account that turns out to be as solid as the standard world-theoretic picture, without being implicitly committed to primitive notions of possibility, necessity, intensionality and so on.

Setting aside the reasonable ontological advantages, we can take the problems involved in abandoning the standard picture of modal quantification as a hint to try a different path to avoid HKP. Instead of giving up on worlds where they have proved themselves to be useful the most, we can try to switch to an alternative semantic background (and its corresponding account of propositions) in a way that allows us to recapture the worlds needed for modalities.

4.6.2 Propositions as Truthmaker Conditions

A risk with kicking the worlds out by the door in the context of modal quantification is that it may only invite them back through the window, in the form of some other proxy notion. The risk is even greater if such a move ultimately appeals to notions standardly understood in a world-theoretic fashion.

However, we can still try to deal with the paradox by focusing on the theory of propositions first. An alternative way to avoid HKP while keeping a hyperintensional picture of propositions may then consist in switching directly to an account that does not identify them with sets of worlds. But since we still need worlds to account for modal quantification, a background semantic framework which allows to uniformly recapture the notion of a world is arguably to be preferred.

We may then try to modify the framework in order to allow partial entities to represent the truth conditions of a sentence, following the tradition of relevance logics and situation semantics. In what follows, I will focus on a specific approach of this kind, which has been recently endorsed within the context of exact Truthmaker Semantics: the theory of propositions as *truthmaker conditions*.[21]

In the previous sections we identified a proposition with the set containing all and only the worlds at which it is true. This is defined by a characteristic function taking worlds as input and giving 'true' or 'false' as output. The truthmaker condition of a

21. See Fine (2017b), Jago (2022).

sentence is a characteristic function that takes also *partial* entities as input (rather than only complete worlds), and gives 'yes' or 'no' as output. A 'yes' answer means that the input entity, usually called a *state*, is a truthmaker for the sentence in question, and this amounts to identifying the proposition expressed by a sentence P with the set $|P|^+$ of all its truthmakers. [22] A fundamental feature of a state space of this kind is that it must be endowed with a mereological structure, which determines a partial order over it. In other words, states can be fused together and be parts of each other, according to a reflexive, transitive and anti-symmetric relation on the set of states. A nice feature of this account is that we can allow also for fusions of *incompatible* states, which can in turn be employed as truthmakers for necessary falsehoods. [23]

How do propositions as truthmaker conditions avoid HKP? Answering by merely pointing out that, according to this picture, (2_i) and (4_i) are false (since propositions are not sets of worlds anymore) would not be enough. As observed above, we may adopt a principle of ontological parsimony to the effect that, in order to account for modal quantification, the required space of worlds must be built out of the entities already employed for analyzing propositions, namely, states. As a number of authors have suggested, we can identify worlds with *maximal* truthmakers.[24] We can then characterize a maximal truthmaker as a state w such that, for every sentence P, either s verifies P or w verifies its negation.

The set of worlds as maximal truthmakers W will include impossible worlds as well, since the maximality constraint by itself does not rule out maximal *inconsistent* truthmakers.[25] Furthermore, a general consistency constraint on truthmakers seems to go against our hyperintensional needs, since inconsistent truthmakers are arguably required in order to provide a sufficiently fine-grained picture of semantic content.

But then, if *any* set of truthmakers qualifies as a proposition, propositional variables and quantifiers will range over the powerset $\wp(T)$ of the set of all truthmakers T. As a result, we would face a new, truthmaker-based variant of HKP:

(1_t) There is a set T of all truthmakers.

22. This kind of propositions may not be "fully" hyperintensional: even if they distinguish between necessary truths, they might not be able to distinguish between all necessary falsehoods. For instance, $A \wedge \neg A$ and $B \wedge \neg B$ might not differ with respect to their verifiers but might differ with respect to their falsifiers. In order to track such distinctions as well, we can pair the sets of verifiers of a sentence with the sets of its falsifiers, resulting in *bilateral* propositions. A bilateral proposition is then a pair $\{|P|^+; |P|^-\}$ such that $|P|^+$ is the set of all P's verifiers and $|P|^-$ is the set of all P's falsifiers. For the sake of simplicity, in what follows I will focus on *unilateral* propositions only. I will say more on TMS and its conception of propositions in chap. 6.

23. Two states s and t are said to be incompatible if and only if their fusion $s \sqcup t$ is an impossible state. Of course, this kind of characterization requires to employ a primitive notion of possibility in the background metaphysics of truthmakers. For an alternative approach that employs a primitive notion of incompatibility between states and defines impossible states in terms of it, see Plebani, Rosella, and Saitta (2022).

24. See for instance Plantinga (1978) and Restall (1996), as well as the idea of a *modalized* state space in Fine (2017b).

25. A maximal inconsistent truthmaker can be characterized as a maximal truthmaker that contains as parts some incompatible truthmakers, or equivalently, some impossible state.

(2_t) There is a set $W \subset T$ of all worlds (the set of all maximal truthmakers).

(3_t) p is a proposition := $p \in \wp(T)$, namely, p is a subset of T.

(4_t) Modal operators are quantifiers that range over W.

(5_t) Propositional variables and quantifiers range over $\wp(T)$.

(6_t) For every p, there is at least one $w \in W$ where p alone has Q.

(7_t) $\therefore |W| \geq |\wp(T)|$.

The problem here would be that, since W is a proper subset of T, its cardinality must be strictly smaller than the cardinality of $\wp(T)$.

One might think that this is not a serious worry: we should simply keep the task of finding a suitable semantics for modal logic separate from the task of providing a theory of propositions in terms of truthmakers – although the adoption of a semantic framework often motivates the adoption of a corresponding theory of propositions. But even if we take a less liberal stance towards the interdependence of those two tasks, the just sketched account may still have access to the resources for avoiding the paradox.

Given its mereological structure, in order to extract an account of propositions from subsets of T, there are some reasonable conditions that we may want to place on them. Where s, t and u are arbitrary truthmakers in T, in order for a subset of T to count as a proposition, we may require it to be (i) upwards closed with respect to fusion: if $s, t \in |p|^+$, then their fusion $s \sqcup t \in |p|^+$; and (ii) *convex*: if $s, u \in |p|^+$ and some part t of u has s as a part, then $t \in |p|^+$. Since it is not the case that every subset of T satisfies such conditions, we can deny (3_t) and (5_t): even if we take worlds to be maximal truthmakers, HKP is apparently rejected.

There seems to be independent justification for placing such conditions for propositionhood.[26] However, employing them weakens the expressive power of the account. If our goal is to provide a picture of semantic content that captures the full range of hyperintensional distinctions, restricting the domain of propositions in this setting might backfire similarly to how placing constraints on impossible worlds backfired in sec. 4.4. Moreover, regardless of what criteria for propositionhood we might place, we still need to be careful in employing the set of worlds *qua* maximal truthmakers as the domain for modal quantification. With some constraints into play, propositions will represent only a proper subset of $\wp(T)$ – for instance, those which conform to the criteria discussed above. Let us call the resulting set of all propositions R.

Since the singleton of each world counts as a proposition (world-singletons are closed under fusion and convex), the set of world-singletons S must be a proper subset of R. Obviously, S and W have the same cardinality. Now, either S has the same cardinality of R, or the cardinality of R is strictly greater than the cardinality of S. The former case

26. See Fine (2017a, 2017b) and Jago (2022).

does not lead to problems with Cantor's theorem, but it may still appear undesirable since it amounts to conceding that there are exactly as many worlds as propositions. On the other hand, if we want to resist that conclusion and assume that $|R|$ is strictly greater than $|S|$, we run again into troubles:

(1_{ti}) There is a set $W \subset T$ of all worlds (the set of all maximal truthmakers).

(2_{ti}) There is a set $R \subset \wp(T)$ of all propositions.

(3_{ti}) There is a set $S \subset R$ of all world-singletons.

(4_{ti}) $|R| > |S|$.

(5_{ti}) $|W| = |S|$.

(6_{ti}) Modal operators are quantifiers that range over W.

(7_{ti}) Propositional variables and quantifiers range over R.

(8_{ti}) For every p, there is at least one $w \in W$ where p alone has Q.

(9_{ti}) $|W| \geq |R|$.

(10_{ti}) $\therefore |R| \leq |S|$.

The conclusion contradicts our assumption (4_{ti}). It seems that, as long as we build worlds out of truthmakers and (8_{ti}) is a correct reading of (Y), we run into some variant of the paradox, unless we concede that there must be exactly as many worlds as propositions. At this point, one might suggest to place some further conditions for propositionhood in order to exclude world-singletons from the domain of propositions – a blatant *ad hoc* move that would further reduce the expressive power of the framework.

Luckily, there is an easier way out: all we need to do is to exclude *impossible* worlds from the domain of modal quantification. We can do so by placing a restricted consistency constraint on maximal truthmakers: in order for a truthmaker to count as a (possible) world, it must be maximal and it must not contain incompatible truthmakers. If our domain of modal quantification contains only *possible* worlds, then (6_t) and (8_{ti}) will not be acceptable readings of (Y) anymore. It would still be true that, for every proposition p, it is either possible or impossible that p uniquely instantiates Q. But this would not amount to claiming that, for every p, there is at least one world at which p uniquely has Q. The reason is pretty obvious: there will not be impossible worlds in the account, so modal quantifiers will behave just like they do in the standard setting. More precisely, $\neg\Diamond\varphi$ will be true if and if only if there is no possible world at which φ is true. As a result, in the new framework the validity of (Y) is guaranteed by the fact that for

every p, at no possible world (i.e. maximally consistent truthmaker) p uniquely has Q.

Notice that, unlike the restriction on impossible worlds discussed in sec. 4.4, the just sketched solution would not constitute an *ad hoc* move: we simply *do not need* impossible worlds anymore. Recall that, in sec. 4.4, impossible worlds were introduced precisely to provide a hyperintensional account of propositions. But here we already did that in terms of truthmaker conditions. Worlds have been subsequently defined in order to account for modal quantification, but they were not required in the first place. Within this setting, we can effectively provide the truth conditions for modal claims in terms of standard quantification over possible worlds, identified with maximally consistent truthmakers, without losing expressive power.

I close this section by highlighting two further advantages of the solution just sketched: first, it does not force to concede that there are exactly as many worlds as propositions. And, most importantly, it is not committed to any particular constraint for propositionhood. This allows to attain the full expressive power guaranteed by the identification of semantic contents with sets of truthmakers, at no additional cost.[27]

4.7 Conclusion

In this chapter I discussed how two of the most popular accounts of hyperintensional propositions deal with two famous paradoxes. In particular, I presented a variant of the Russell-Myhill paradox that specifically affects the ersatz IWS presented in the previous chapter. This provided further reasons to drop the ersatz picture in favor of an alternative conception of IWS.

However, even non-ersatz versions of IWS seem to be affected by a more general threat: a variant of Kaplan's paradox of propositions. After discussing why the main solutions proposed for the original Kaplan's paradox fall short when applied to its hyperintensional variant, I explored two general paths to a possible solution: giving up on sets and giving up on worlds.

My diagnosis is that abandoning sets altogether in favor of plural quantification or quantification over classes does not seem to be a viable option, and the same applies to the idea of abandoning worlds in the analysis of modal operators. On the other hand, giving up on worlds in the account of propositions in favor of truthmakers allows for a simple solution to the paradox. Most importantly, this is the case even if we opt to keep worlds in the account of modal quantification.

The discussion of Kaplan's paradox has been particularly helpful for investigating the interaction between the standard semantics of alethic modalities and different conceptions of propositions. In order to fulfill the need for both a hyperintensional account of propositions and a worldly picture of modal quantification (which represent the two key ingredients for the paradox), as an alternative to IWS I focused on the theory

27. A different approach to modalities in TMS has been recently developed by Kim (2024). While exploring the details of this account is out of the scope of the present work, it is reasonable to think that it should be able to avoid the paradox as well, since it does not rely on a notion of world at all, but instead provides exact truthmaker clauses for modal statements.

of propositions associated with TMS, since it allows for a straightforward recapturing of the standard notion of possible world. However, it is important to stress that my case study does not rule out that other approaches to propositions rooted in alternative semantic frameworks may also effectively avoid the paradox.

Chapter 5

Counterfactuals, Non-vacuism and Strong Emergence

This second part of the work will be devoted to a specific instance of worldly hyperintensionality: counterfactual conditionals. A fairly obvious observation that might be made at this point is that, as long as we subscribe to a hyperintensional notion of semantic content like those sketched in the previous chapters, we would clearly end up with hyperintensional counterfactual propositions. However, developing a full-fledged semantics for this kind of constructions – i.e. a satisfactory way to assign them hyperintensional content – is far from easy.

As we shall see in more detail, the debate on the hyperintensionality of counterfactuals has been developed primarily along two distinct but related axes: the *non-vacuism* axis and the *formal invalidities* axis. The former will be explored in the present chapter, where new evidence against vacuism will be provided; the latter will be explored in chap. 6, where a full hyperintensional semantics for first-degree counterfactuals will be developed. Finally, the last part of the chapter will be devoted to the discussion of a general way to deal with both of the axes, by further extending the basic semantic framework.

5.1 Background

Counterfactuals can be roughly identified as subjunctive conditionals of the form 'if it were the case that p, then it would be the case that q'. More generally, we may say that counterfactuals are conditionals that describe the consequences (in the broadest sense of the notion) of something that did not happen, typically in virtue of some dependence relation holding between the antecedent and the consequent. Attempts to provide a formal semantics for this kind of constructions have been focusing primarily on counterfactuals with *contingently* false antecedents. There are some fairly obvious historical and technical reasons for this, most notably the fact that late 20th century has been dominated by the intensional paradigm in logic and philosophy of

language, as discussed in chap. 2. Not surprisingly, the power and versatility of PWS led to the development of what still nowadays represents the standard approach to counterfactuals, which is best exemplified by the frameworks of Stalnaker (1968) and Lewis (1973).

In a nutshell, according to such accounts a counterfactual $A \boxrightarrow C$ is true (at a world w) if and only if C is true in (all of) the *closest-to-w* possible world(s) where A is true (I will discuss a specific version of Lewis' semantics in more detail in the next chapter). The notion of closeness is typically understood in terms of comparative similarity: a way of ordering possible worlds that gets embedded in the semantics by means of some relation or selection function. Usually, the relevant similarity metric to be picked is largely determined by the context of utterance, and it is generally a tricky notion to characterize in a rigorous way.

Regardless of the technicalities concerning similarity, standard accounts deal pretty well with counterfactuals with contingently false antecedents, for which they usually deliver intuitive results. For instance, to evaluate a counterfactual like "if I had missed the train this morning, I would have been late for work", we simply check what happens at the closest world(s) where I missed the train this morning. If in all those worlds I am also late for work, the counterfactual is true at the actual world.

Troubles begin when we try to evaluate counterfactuals with *necessarily* false antecedents, namely, *counterpossibles*. In a standard framework, there is simply no world at which the antecedent of a counterpossible is true (recall that in PWS necessary truths are identified with the set of all worlds, and impossibilities with the empty set). Therefore, there will be no world (and *a fortiori*, no closest world) at which both the antecedent and the consequent are true. But this just means that any counterpossible will be identified with the same set of worlds, no matter what the consequent is. An obvious consequence is that all counterpossibles will *vacuously* have the same truth value. The view according to which this is the correct way to evaluate counterpossibles is known as *vacuism*. The standard form of vacuism (the one on which I will focus here) has that all counterpossibles are vacuously *true*. The main reason for this is logical in nature: it seems desirable that $A \boxrightarrow A$ is always true, even when A is impossible (notice also that a material conditional $A \supset A$ is standardly assumed to imply its counterfactual counterpart $A \boxrightarrow A$).[1]

Many philosophers find vacuism extremely counterintuitive, and it is very easy to see why. Consider this classic example from Nolan (1997):

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

(1) has an impossible antecedent and seems intuitively false, but it is true according to the standard intensional account.

1. We get an alternative route to this type of vacuism by simply recognizing that a standard clause for counterfactuals requires to quantify over an empty subset of the set of worlds. Since universal quantification over an empty domain is always true, a counterfactual with an impossible antecedent will always be *vacuously* true, no matter what the consequent is.

Defenders of the orthodoxy – most notably, Williamson (2018, 2024) – try to explain away this kind of intuitions by stressing that alleged examples of false counterpossibles depend entirely on representational phenomena, cognitive limitations and heuristics that should not affect our judgments about the correct semantics for counterfactuals. However, the dissatisfaction with vacuism eventually led a number of philosophers to build stronger cases for *non-vacuism* than mere appeals to intuition. This has been pursued both by developing direct arguments against the alleged cognitive mistakes involved in evaluating certain counterpossibles as false (see Berto et al. (2018), Berto (2024)); and by providing independent evidence for the *need* of non-vacuous counterpossibles in several theoretical contexts.[2]

For instance, non-trivial counterpossibles have been invoked for explaining a number of important metaphysical notions such as causation, essences, dispositions and omissions.[3] Most notably, Schaffer (2016) and Wilson (2018) provide an application of a popular semantic model for counterfactuals based on structural equations (see Briggs (2012)) to deal with metaphysical grounding, and argued that the resulting *interventionist* analysis commits to a non-vacuist picture of counterpossibles.[4] The mutual epistemic support between causation, grounding and counterpossibles hints to the conclusion that, in a large number of cases, such conditionals track worldly dependence relations. As Hicks (2022) nicely sums up (p. 27):

> Counterpossible reasoning is useful to us primarily as a guide to real-world dependence relations, whether those are causal – as when we use metaphysically impossible idealizations to model some of a system's causal relations in isolation from others – or metaphysical – as when we engage in counterpossible suppositions to see whether they conflict with what we know about the world. In this way, the counterpossibilities stand in for or represent the actual dependence relations among real-world things.

Indeed, other philosophers highlighted how counterpossible reasoning seems to figure in a variety of non-philosophical disciplines, including mathematics and even natural sciences.[5] If counterpossibles are consistently employed in some of our best scientific practices, it seems that we can devise a very powerful argument against vacuism. Consider the following example discussed by Tan (2019):

---

2. See Kocurec (2021) for an overview.

3. See Jenkins and Nolan (2012), Brogaard and Salerno (2013), and Bernstein (2016).

4. As observed by Wilson (2018), the relation between counterpossibles and grounding is one of mutual epistemic support, not a full reduction of one concept to the other. He argues that grounding entails non-vacuous counterpossibles, not that grounding should be analyzed in terms of counterpossibles, even though it certainly remains an open option (just like causation has been influentially analyzed in terms of counterfactuals): «the interventionist framework can be used to reduce causation to counterfactuals, or to specify truth-conditions for an interesting class of counterfactuals in causal-theoretic terms, or to articulate a two-way interdependence. [...] By taking this line, interventionists can exploit the connection between causation and counterfactuals without having to settle on which is ultimately to be reduced to which» (p. 3).

5. See Jenny (2018), Tan (2019), McLoone (2021), Wilson (2021), McLoone, Grützner, and Stuart (2022), Baker (2024). For a critical assessment of Jenny's influential paper, see Plebani, San Mauro, and Lenta (2024).

(2) If diamond was not covalently bonded, it would be a better electrical conductor.

This counterfactual is meant to capture the dependence of a macroscopic property of diamond on the structural properties of its atoms, and partly accounts for its poor electric conductivity. [6] Being covalently bonded is an essential property of diamond, therefore it is covalently bonded as a matter of metaphysical necessity. But this means that the antecedent of (2) expresses a metaphysical impossibility: we have a (non-vacuously) true counterpossible that, according to Tan, figures in a scientific explanation. Hence, it seems that we have independent reasons for expecting our best semantics to capture the distinction between true and false counterpossibles.

In what follows, I will try to show how a similar strategy can be adopted by exploiting some insights from a debate that is currently going on in the metaphysics of chemistry. In a series of papers, Robin Hendry characterizes *strong emergence* (SE) in chemistry in terms of *downward causation*. In order to make this idea precise, Hendry (2009) employs a *counternomic* criterion: claiming that a system exhibits downward causation amounts to claiming «that its behaviour would be different were it determined only by the more basic laws governing the stuff of which it is made» (p. 185). In other words, downward causation – and thereby strong emergence – is spelled out in counterfactual terms.

The satisfaction of the counternomic criterion is nicely exemplified by the case of isomers: distinct molecules sharing exactly the same atoms in the same number, but arranged in different structures. For instance, ethanol and methoxymethane, while being composed of the same atoms in the same number, display very different properties. This fact is accounted for by their different microstructures, that according to Hendry are (or may well be) strongly emergent on the physical entities that constitute the molecules: at the physical scale, we are not able to retrieve any relevant structural feature. In counterfactual terms (p. 189):

> since a molecule's causal powers depend on its structure, its behaviour would be different were it determined by more basic (quantum-mechanical) laws governing the particles of which the molecule is made.

In other words, molecular structure satisfies the counternomic criterion for downward causation, and this allows us to consider it a strongly emergent property. Taking this whole story about molecular structure seriously provides justification for the truth of counterfactuals like the following:

(3) If the properties of a molecule were determined at the physical scale only, ethanol and methoxymethane would be the same molecule.

However, statements of this kind interact with another popular thesis in the metaphysics of chemistry, endorsed by Hendry (2023) himself: microstructural essentialism (ME), the view that microstructure is an *essential* property of chemical substances (*qua* individual

6. Tan's view – borrowed from Woodward (2003) – is that any explanation of some fact is incomplete until we specify counterfactually how that fact would have been different under various circumstances.

molecules). If molecular structure is strongly emergent, supposing that a molecule's properties were determined only at the physical scale amounts to considering a scenario in which molecules have no structure. But, if ME is true, this implies that (3) is a counterpossible.

How does vacuism affect someone endorsing both SE and ME? The aim of this chapter is to explore this question and defend the idea that the combination of SE and ME (SE+ME) entails that some counterpossibles are false.

5.2 Strong Emergence in Chemistry

To borrow Seifert (2022)'s words: «emergence is the idea that there is something more to things than what we can say about them by looking only at their constituents» (p. 1). In the philosophy of chemistry, emergentism holds that some chemical properties or behaviors are emergent. A simple way to illustrate emergentism is by considering its main rival first, namely, *reductionism*. This is the view according to which, at least in principle, the whole field of chemistry is reducible to physics. From an epistemic standpoint, reductionism amounts to the view that every description or explanation of some chemical phenomena could in principle be translated into the language of physics; from an ontological standpoint, it amounts to the view that every chemical property or behavior is "nothing more" than a physical property or behavior, so that interactions between physical particles exhaust the whole range of phenomena that we observe at the scale of chemistry.

Emergentism is one way of rejecting reductionism. Such a rejection comes in different degrees, corresponding to the distinction between *weak* and *strong* emergence. As a general sketch, we can say that a property or behavior is weakly emergent if it is entirely determined at the physical scale, but the current theoretical resources of physics are not sufficient for predicting or explaining the fact that a system displays it (so that we need to defeasibly posit its higher-lever reality).

A property or behavior is strongly emergent if such a predictive or explanatory failure is a necessary outcome: in other words, if it would be impossible to explain or predict the existence of the observed property or behavior by means of physics alone, even with an ideal perfect theory. This is usually taken to be a consequence of the fact that a strongly emergent system has some higher degree of autonomy from the basic entities that constitute it, for instance by possessing entirely novel causal powers.

One of the main advocates for the possibility of SE in chemistry is Robin Hendry, who devoted a series of papers to defend the compatibility of the actual scientific practice with the existence of strongly emergent chemical properties. In particular, Hendry (2017) adopts a causal criterion for the reality of a property, according to which existing requires the possession of causal powers. Following this principle, one can identify the presence of a strongly emergent chemical property by identifying the presence of causal powers in addition to those explained by physics. This should not conflict with the laws of physics themselves, since (p. 2):

> the possession of novel causal powers does not require the violation of more fundamental laws. Strong emergence requires not that these laws be broken, but only that they fail to determine what happens.

So, what sort of special causal powers are displayed by strongly emergent systems? Hendry's view is that their distinctive feature is *downward causation*: emergent systems exhibit higher-level properties that determine some difference at the lower level from which they arise.[7] In Hendry's own words: «the subsystems of an emergent supersystem sometimes do something different to what they would do if the causal structure of the world were as imagined by the reductionist» (p. 2). This is a counterfactual conditional – which Hendry labels *counternomic* due to its reference to non-actual laws (those endorsed by the reductionist) in the antecedent – and it represents the general kind of statements that are entailed by ascriptions of downward causation. According to Hendry, a property or behavior is strongly emergent when it satisfies this counternomic criterion for downward causation.

One of the best purported examples of a strongly emergent property is molecular structure. On this matter, the reductionist stance is that molecules are nothing more than systems of charged particles, interacting according to the laws of quantum mechanics. This is formally captured by the Schrödinger equation: a mathematical expression that describes an isolated molecule purely in terms of its electrons and nuclei and that, once solved, should account for its structure. However, in order to solve the Schrödinger equations for atoms more complex than hydrogen, or for entire molecules, quantum chemists need to calibrate the physical models with a set of assumptions drawn from chemistry. In other words, quantum chemists need to introduce well-understood approximations at the fundamental level – known falsehoods – that will affect the calculations. As remarked by Hendry (2009, p. 183),

> quantum chemistry turns out not to meet the strict demands of classical reductionism, because its models bear only a loose relationship to exact atomic and molecular Schrödinger equations, and its explanations seem to rely on just the sort of chemical information that, in a classical reduction, ought to be derived.

For isomers, this kind of move is most evident. Isomers are distinct molecules with the same molecular formula, namely, distinct molecules composed of exactly the same atoms in the same number, that differ only with respect to their structures. For instance, in ethanol (CH_3CH_2OH) and methoxymethane (CH_3OCH_3), the same basic entities (six atoms of hydrogen, two atoms of carbon and one atom of oxygen) are arranged in different structures, which give rise to very different macroscopic properties.

In order to build the complete Schrödinger equation of a molecule, one needs its *resultant Hamiltonian*, an operator corresponding to the total energy of the system which

7. This is a common way to characterize SE in the relevant literature. Consider the definition of a strongly emergent property provided by Gillett (2016): «a property is strongly emergent just in case it is a property of a composed individual that is realized and that (in addition to having same-level effects) non-productively determines the individual's parts to have powers that they would not have given only the laws/principles of composition manifested in simpler collectives».

takes into account all the intra-molecular interactions, using only fundamental physical interactions as input. But since isomers differ only in structure, they will share the same resultant Hamiltonian: it is impossible to distinguish them at the purely physical scale.[8] To distinguish between isomers, quantum chemists employ the *configurational Hamiltonian*, an operator that allows to specify the nuclear positions within a molecule to account for its structure. But this structural insight comes from chemistry's empirical observations: it is not something that could be deduced, even in principle, from our best physical descriptions of the molecules. In other words, configurational Hamiltonians assume higher-level information that "impose" on the Schrödinger equation the structure that is supposed to be derived. According to Hendry (2009), this satisfies the counternomic criterion for downward causation, making molecular structure a strongly emergent property (p. 185):

> In an emergent complex system, the behaviour of the basic stuff of which it is made is governed by a configurational Hamiltonian, which is different from what it would be were its behaviour governed by the resultant Hamiltonian. Since the Hamiltonian of a system determines the precise nature of the physical law that governs its behaviour, to say that some system exhibits downward causation is to make a counternomic claim about it – that its behaviour would be different were it determined only by the more basic laws governing the stuff of which it is made.

The need for configurational Hamiltonians suggests that it may be impossible to specify all the chemical properties of a molecule at the physical scale, and this gets reflected in the actual scientific practice. However, this may not be merely because the mathematics required to compute certain properties would be too complex, but because *such properties are simply not there*: they may not exist at the scale of quantum physics. A resultant Hamiltonian may not be able to specify the structure of a molecule at the quantum scale precisely because *there is no structure at that scale*. If SE about molecular structure is true, such a property is exhibited at the chemical scale only, and it causally determines part of the behavior of the entities constituting the molecule at the physical scale.

5.3 Microstructural Essentialism

Microstructural essentialism (ME) is the thesis that structure is an essential property of substances, *qua* individual molecules. It is a far less controversial thesis than SE, and far more widespread among philosophers of chemistry. Hendry (2023) defines it as follows (p. 278):

> Microstructural essentialism about chemical substances is the thesis that the structure of a chemical substance at the molecular scale is what makes it the substance that it is, and not another one.

8. Besides their inadequacy for distinguishing molecular structure, resultant Hamiltonians are never used primarily for practical reasons: they would be too complex to be computed (see Hendry (2009)).

Embracing ME obviously amounts to maintaining that molecular structure fixes the identity conditions of chemical substances.

This naturally connects to the fact that the possession of an essential property is a metaphysically necessary matter.[9] Generally speaking, if a property of a given natural kind is essential, then members of that kind necessarily have that property. Hence, if molecular structure is an essential property, in order for a molecule to belong to a natural kind (namely, to *be* a certain chemical substance), it must display the structural features that identify that kind. This represents a theoretical virtue of essentialism, as highlighted by Hendry (p. 278):

> essentialism explains the possibility of reliable inductive inferences because possession of the essential property explains why members of the kind have many other properties associated with the kind.

In arguing for the microstructural part of the thesis, Hendry emphasizes the centrality of molecular structure for chemical classification, the role of microstructure in explaining and predicting the behaviour of substances, and the fact that no other systematic basis for individuating substances is consistent with the actual chemical practice. For the essentialist part of the thesis, Hendry (2023) stresses that microstructure is *causally upstream* from other candidates for substance identification (see van Brakel (2000), Needham (2011)), such as thermodynamic and spectroscopic behaviour, or ground-state electronic structure (p. 283):

> Water's molecular structure is part of the reason why it has the thermodynamic behaviour it has, but the converse is not true. Similarly, gold's atomic number being 79 is part of the reason why its ground-state electronic structure and spectroscopic behaviour are what they are, and the converse is not true. The microstructure is what grounds the behaviour, but not vice versa.

For the current purposes, there is no need to discuss Hendry's arguments for ME in detail. However, an important aspect of essentialism that should be highlighted is its modal profile. As already remarked, an essential property is one that is possessed as a matter of metaphysical necessity. The standard way to analyze this kind of modal talk is by means of PWS, according to which the ascription of a property to an object is necessarily true if and only if it is true in all (metaphysically) possible worlds, i.e., if and only if the object instantiates that property in all possible worlds. In light of this, an essential property of a chemical substance is one that the substance instantiates in all possible worlds, so that – assuming ME – every chemical substance has the same molecular structure in every world.

9. Being instantiated as a matter of metaphysical necessity is a necessary (yet arguably not sufficient) condition for being an essential property.

5.4 SE+ME Entails Non-vacuism

I am now in the position to argue for the following claim: the combination of SE and ME entails that some counterpossibles are false, which amounts to denying vacuism. As a corollary, I will argue that someone who endorses both SE and ME should, in fact, be unhappy with vacuism. Let us begin by considering the following counterfactual:

(3) if the properties of a molecule were determined at the physical scale only, ethanol and methoxymethane would be the same molecule.

As previously discussed, SE about molecular structure (at least in Hendry's formulation) seems to provide justification for the truth of (3). Molecular structure determines the causal powers of chemical substances and fixes their identity conditions. Furthermore, if it is a strongly emergent property, then it is absent at the physical scale. Therefore, counternomically, if the properties of a molecule were determined only at the physical scale (at which there is no structure), two isomers like ethanol and methoxymethane could not be distinguished in terms of their structural features. But then, since isomers are distinguished only in terms of their structural features, in every closest world that verifies the antecedent, ethanol and methoxymethane are indiscernible. By the identity of indiscernibles, in those worlds ethanol and methoxymethane are simply the same molecule.[10]

How does this interact with ME? Notice that, if ME is true, the truth value of counterfactuals like (3) can be correctly delivered only by considering what happens at some impossible worlds. This can be illustrated with a simple argument:

(a) If there is no structure at the physical scale, then if the properties of a molecule were determined at the physical scale only, molecules would lack an essential property.

(b) There is no structure at the physical scale.

(c) Therefore, if the properties of a molecule were determined at the physical scale only, molecules would lack an essential property.

The premises (a) and (b) are direct consequences of maintaining that molecular structure is a strongly emergent essential property. SE entails that, at least in the actual world, there is no structure at the physical scale; ME entails that absence of structure is absence of an essential property. This leads to the truth of (c): the closest worlds verifying the antecedent (those in which we keep everything fixed but the fact that the properties of a molecule are not determined only at the physical scale, let us call them *A*-worlds), must be worlds in which molecules lack an essential property.

10. French (1989) discusses some issues for identity of indiscernibles in the context of quantum physics. However, the arguments against it have small significance for the present purposes, unless we subscribe to the full reducibility of chemical entities to their quantum components, which is obviously incompatible with SE.

But then, it seems that the closest A-worlds (according to the just sketched similarity metric) must clearly be scenarios involving the violation of a metaphysical law, since objects possess their essential properties as a matter of metaphysical necessity. As long as we work with complete worlds, the metaphysical impossibility of the consequent "bubbles up": if (c) is true, any closest A-world should be a world verifying the consequent (let us call them C-worlds), hence, all such worlds represent *impossible* scenarios.

However, if we restrict ourselves to only the *possible* worlds verifying the antecedent of A – here I am assuming that SE might be merely contingent, so that there will be possible worlds at which the properties of a molecule are determined at the physical scale only –, all of them *falsify* the consequent. The reason for this is straightforward: there is no *possible* world at which the consequent is true. Suppose that there were both C-possible worlds and $\neg C$-possible worlds (no matter which are closer to the actual world). Then, the truth of C – whether or not isomers lack an essential property – would be a contingent matter. But we know that this cannot be the case, unless we subscribe to the implausible view that the lack (or, equivalently, the possession) of certain essential properties is itself a contingent matter.

Within a standard setting, this forces us to evaluate (c) as *false.* But this goes against what has been established by our argument: the truth of SE+ME provides good reasons to believe that (c) is in fact true.[11] In light of this, it turns out that in order to deliver the correct evaluation of (c), we would need to check what happens at the closest *impossible* A-worlds. Obviously, this cannot be done within a standard framework, for the reasons discussed above.

If we grant that impossible worlds are required for a correct evaluation of (c), the same obviously goes for (3) since they share the same antecedent (at least as long as we stick to the same similarity metric). Furthermore, notice that if natural kinds are rigid designators, as it is standard to hold, identity claims about them should be either necessarily true or impossible. Since in the actual world ethanol and methoxymethane are distinct substances (distinct natural kinds), they are necessarily distinct. But then, worlds that verify the consequent of (3) – worlds in which ethanol and methoxymethane

11. This represents an instance of an odd phenomenon that we might label as *reverse vacuism*. It happens when the impossibility of the consequent in a counterfactual is sufficient to make it trivially false, even if the antecedent is possible and the counterfactual looks *prima facie* true. Cases of this kind have been discussed by Nolan (1997, 2017) and Van der Laan (2004), and have been used as counterarguments to a principle that is sometimes assumed in extended semantics that employ impossible worlds: Strangeness of Impossibility Condition (SIC), according to which any possible world is closer to the actual world than any impossible world. We may strengthen this point further by considering the strict conditional version of (c):

(cs) necessarily, if the properties of a molecule are determined at the physical scale only, molecules lack an essential property.

(cs) asserts an intuitive modal connection (given SE+ME) between the worlds that falsify SE and the worlds where molecules lack an essential property. However, if we take (cs) to be true, (c) must be true as well, since strict conditionals entail their counterfactual counterpart. But this cannot be the case, as long as we have only possible worlds in our semantics.

are the same molecule – must be, once again, impossible worlds. Hence, if we want (3) to be true, all of the closest *A*-worlds must be impossible.

Is this enough to claim that SE+ME entails non-vacuism? Impossible scenarios may be required for a correct evaluation of counterfactuals like (3) and (c), but this does not allow us to conclude that they are counterpossibles. So far, we have simply shown that all of *the closest* (under the most reasonable interpretation) *A*-worlds are impossible, but in order to claim that (c) is a counterpossible we need to show that *all* of such worlds are impossible. Furthermore, we need show that the same arguments for the non-vacuous truth of such counterpossibles can be employed for justifying that other counterpossibles are false. It seems that in order to achieve this, we need to either assume or establish some further facts about the totality of our modal space.

First and foremost, we should say something about the modal strength of SE. While the necessity of ME (if it is true at all) does not seem to be a matter of dispute, since it is explicitly invoked to account for a molecule's identity, the same cannot be assumed about SE. Indeed, one may well take SE to be merely contingent: there might be metaphysically possible worlds at which molecular structure is a property entirely determined at the physical scale. If this is the case, even if we agree on the need for impossible worlds to properly evaluate counterfactuals like (3) and (c), they would still not qualify as counterpossibles. Hendry does not take an explicit stance about this in his papers, but he stresses that his arguments are only intended to show that all of the available evidence from the chemical practice is *compatible* with the truth of SE in the actual world. Furthermore, his appeal to violations of merely *physical* laws in the counternomics employed to analyze downward causation suggests that we should in fact take SE as merely contingent.

Even though this is not the main aim of this chapter, I want to raise some doubts towards this view. In particular, I think that it is far more reasonable to believe that if SE is true at all, it must be necessarily true (even if we set aside the trivial remark that whatever is presented as a metaphysical truth, is usually presented as a metaphysically *necessary* truth).

Let us begin with some questions. If we grant that SE is merely contingent, can we still consider it a thesis about the *nature* of molecular structure? SE is presented as a metaphysical claim about chemical substances. In particular, it characterizes how an essential property of molecules is determined by a certain interaction happening between the chemical and the physical scale. Do we really want to say that this interaction is merely contingent? To borrow a popular jargon from the metaphysical literature, can something merely contingent be *joint-carving*? To answer positively amounts to holding a very weak conception of SE.

Now, compare this to SE's main rival: reductionism. If reductionism is true, then molecules are literally *identical* with the sum of their physical constituents, at least in every physically possible world. In other words, reductionism sets some identity facts. But since (numerical) identity is the metaphysically necessary relation *par excellance*, it seems reasonable to assume that if reductionism is true, it is true in every *metaphysically* possible world. On the other hand, SE denies reductionism by allowing that in some

metaphysically possible worlds (such as the actual one), molecules are *not* identical with the sum of their physical constituents. Clearly, 'some' does not entail 'some *and not all*'. In light of this, suppose for *reductio* that there were both metaphysically possible worlds where reductionism is true and metaphysically possible worlds where SE is true. Then, since reductionism fixes identity facts about molecules, such identity facts would be merely contingent (after all, there would be worlds where SE is true, hence where molecules are *not* identical with the sum of their physical constituents). All of this sounds rather weird, and goes against what is standardly assumed to be the modal profile of identity.

This may not represent a definitive argument against the contingency of SE, but I think that it is more than enough to shift the burden of proof. At the current state, the more reasonable option is to defeasibly assume the necessity of SE (if true at all). If that is the case, (3) and (c) are obviously counterpossibles, since their antecedent asserts something that is implied to be false by SE. Furthermore, notice that in this case we do not even require ME to be true: the antecedent would be impossible even if molecular structure was not essential.

Summing up, given ME and SE, it is metaphysically impossible that the properties of molecules are determined at the physical scale only, so, in order to correctly evaluate counterfactuals like (3) we need to consider what happens at impossible worlds, which requires to extend the standard semantics accordingly. Moreover, under further reasonable assumptions about the modal strength of the notions involved, counterfactuals like (3) are (true) counterpossibles, regardless of ME.

Why should this be a problem if vacuism is true (so that statements like (3) are true anyway)? First, notice that under the vacuist picture even Hendry's counternomic criterion for downward causation turns out to be vacuous. Recall his general scheme: «[the behaviour of a molecule] would be different were it determined by more basic (quantum-mechanical) laws governing the particles of which the molecule is made». Here, the antecedent 'the behaviour of a molecule is determined by quantum-mechanical laws' is very close to the antecedent of (3) and (c): as long as we take the behaviour of a molecule to be fully grounded on its properties, the counternomic criterion requires to consider the same impossible scenarios discussed above! Since the counternomic criterion is a fundamental part of Hendry's analysis, embracing vacuism would then undermine the epistemic fruitfulness and acceptability of SE.

But there is an even more straightforward reason why someone endorsing both SE and ME should be unhappy with vacuism. After all, there is nothing genuinely *vacuous* about the truth of (3): one could justify it precisely by appealing to the arguments for SE and ME illustrated above.[12] Moreover, those very same arguments should force one to reject a counterfactual like:

(4) if the properties of a molecule were determined at the physical scale only, ethanol and methoxymethane would *not* be the same molecule.

12. Compare this to a sentence like "if 2+2 were equal to 5, then pigs would fly", for which the only way to "justify" its alleged truth would be by appealing to its vacuity.

But since (3) and (4) have the same antecedent, vacuism forces us to consider (4) as a true sentence, which is an unpalatable result. In sum, either the combination of SE and ME is true and some counterpossibles are false, contrary to the orthodoxy; or vacuism is true and either SE or ME (or both) are false.

5.5 Conclusion

In the present chapter I established a conceptual link between non-vacuism and the combination of strong emergence and microstructural essentialism as defended by Robin Hendry, showing how the former is entailed by the latter under few reasonable assumptions.

In light of this, one might appeal to vacuism to reject SE or ME. If vacuism is true and SE+ME entails that it is false, then either SE or ME (or both) must be rejected. However, I do not believe that this would constitute a compelling argumentative strategy. Indeed, such a move would commit to the idea that a certain result predicted by a specific abstract model (according to the standard semantics for counterfactuals, all counterpossibles are true) should guide us about a non-representational metaphysical issue (whether a certain property is strongly emergent/essential).

But it seems way more reasonable to change the semantics in order to model the language required to speak about some real-world phenomena, rather than the other way around (namely, revising our metaphysical theories to accommodate a byproduct of our semantics). Since the arguments for the truth of SE and ME are independent from the adoption of any specific semantics for counterfactuals, I think that we should be even more suspicious of any attempt of explaining away the deviant data (i.e. that some counterpossibles are false) in order to save vacuism: readapting a popular dictum, *if the facts don't fit in the semantics, so much worse for the semantics.*

Chapter 6

A Combined Approach to Hyperintensional Counterfactuals

6.1 Background

As introduced at the beginning of chap. 5, the second axis towards hyperintensionality in counterfactuals concerns a widely discussed family of formal invalidities. Ever since the very first attempts to develop a logic for counterfactuals, one of the main challenges has been that of invalidating certain inferences about them, while preserving other fundamental logical principles. In what follows I will delve into a significant portion of this debate, in order to discuss one of the most promising hyperintensional semantics for counterfactuals, its shortcomings, and how they can be fixed.

For the ease of exposition, I will begin by considering one of the most debated principles in counterfactual logic, *Simplification of Disjunctive Antecedents* (SDA):

$$A \vee B \boxrightarrow C \vDash A \boxrightarrow C \qquad\qquad A \vee B \boxrightarrow C \vDash B \boxrightarrow C \qquad \text{(SDA)}$$

While (SDA) is not valid on standard approaches to counterfactual logic, such as the ones of Stalnaker (1968) and Lewis (1971, 1973b), authors like Nute (1975), Fine (1975), and Chellas (1975) point out that (SDA) enjoys strong pre-theoretic plausibility. In fact, the principle gets *prima facie* linguistic support from examples like the following:

(1) If it were rainy or windy, then I would be cold.

(2) a. If it were rainy, then I would be cold.

b. If it were windy, then I would be cold.

It is plausible to think that if (1) is true, so are both (2a) and (2b).

At the same time, there is a strong case *against* the general validity of (SDA), in the form of well-known linguistic counter-evidence to the principle. The standard counterexample is due to McKay and Van Inwagen (1977):

(3) If Spain had fought with the Allies or the Axis, then Spain would have fought with the Axis.

(4) a. If Spain had fought with the Allies, then Spain would have fought with the Axis.

b. If Spain had fought with the Axis, then Spain would have fought with the Axis.

According to (SDA), we have that (3) implies both (4a) and (4b). But while (3) seems true (given Franco's explicit offer to join the Axis in exchange for support of his colonial ambitions) and (4b) is harmless, (4a) seems clearly false. More recently, Lassiter (2018) presents additional counterexamples involving complex sentential operators like "usually" and "probably" in the consequent.

In light of this, it seems reasonable to want a counterfactual logic where (SDA) is valid in some contexts, but invalid in others. More formally speaking, one of the basic theoretical *desiderata* that will guide us in the present chapter is the development of a semantics where (SDA) is valid over *some* models, but invalid over all models. In this way, we can accommodate both the seemingly valid instances of (SDA), like the inference from (1) to (2), *and* the apparent failure of the principle in cases like the inference from (3) to (4a). This project, however, faces technical difficulties.

In addition to the linguistic counterexamples, Ellis, Jackson, and Pargetter (1977) point out that there are non-trivial *logical* issues associated with (SDA). The main problem involves the following congruentiality principle:[1]

$$A \dashv\vDash B \quad \Rightarrow \quad A \boxright C \dashv\vDash B \boxright C \qquad \text{(Left-Congruentiality)}$$

In words, (Left-Congruentiality) states that counterfactuals respect logical equivalence in their antecedents. This principle is validated by many approaches to counterfactual logic, among which the canonical Stalnaker-Lewis approach.

The problem is that, under rather minimal logical assumptions, we can derive the problematic principle (Strengthening) from (SDA) and (Left-Congruentiality):

$$A \boxright C \vDash (A \wedge B) \boxright C \qquad \text{(Strengthening)}$$

This principle is in conflict with the defeasible nature of counterfactual inference. To see this, suppose that I forgot my umbrella at home and thus the following counterfactual is true:

(5) If it were rainy, then I would get wet.

By (Strengthening), we can infer:

(6) If it were rainy and I had my umbrella with me, then I would get wet.

1. Generally, an n-ary operator O is i-congruential iff $A_i \dashv\vDash B_i$ entails $O(A_1, \ldots, A_i, \ldots, A_n) \dashv\vDash O(A_1, \ldots, B_i, \ldots, A_n)$, where $\dashv\vDash$ expresses logical equivalence. An n-ary operator is congruential iff it is i-congruential for all $i \leq n$.

But (6) is clearly false, illustrating the pre-theoretic implausibility of (Strengthening).

To derive (Strengthening) from (SDA) using (Left-Congruentiality), consider the following equivalence:

$$A \vee (A \wedge B) \dashv\vDash A \qquad \text{(Absorption)}$$

This law holds in any logic that treats $\vee, \wedge$ as lattice operators – which includes classical logic, intuitionistic logic, FDE, and many others.[2] Using the law, we can reason as follows:

1. $A \boxdot\!\!\rightarrow C$ (Assumption)
2. $A \vee (A \wedge B) \boxdot\!\!\rightarrow C$ (Absorption), (Left-Congruentiality)
3. $\therefore A \wedge B \boxdot\!\!\rightarrow C$ (SDA)

In short, whenever we have (SDA), (Left-Congruentiality), and (Absorption), we also have (Strengthening) – which is unacceptable.

So, if we want to validate (SDA) – even just over a restricted class of models – something has got to give. In this chapter I shall not explore the option of changing the background logic to some non-classical logic in order to get rid of (Absorption) and similar principles – such a logic would have to be too weak to be of interest as a general-purpose background logic. Instead, I will focus on an approach that rejects (Left-Congruentiality) in order to be able to validate (SDA) (restricted to a class of models) without validating (Strengthening). In particular, we can take the lesson from Ellis, Jackson, and Pargetter (1977) to be that – on pain of (Strengthening) – endorsing (SDA) requires us to take counterfactual antecedents as creating hyperintensional contexts. In recent years, Fine (2012a, 2012b) and Ciardelli, Zhang, and Champollion (2018) provided further arguments for the very same conclusion.

To capture this feature, several authors proposed full-fledged hyperintensional theories of counterfactuals. [3] These proposals differ among each other, but they all have in common the rejection of the standard PWS account of propositions, which underlies the Stalnaker-Lewis approach. As previously illustrated, according to this model the proposition $[A]$ expressed by a statement A is identified with the set $\{w : w \vDash A\}$ of worlds w where A is true.

In fact, rejecting this model is semantically required for rejecting (Left-Congruentiality). To see this, first note that if two statements A and B are logically equivalent, $A \dashv\vDash B$, then according to PWS, $[A] = [B]$. By compositionality, the proposition $[O(P_1, \ldots, P_n)]$ expressed by a sentence $O(P_1, \ldots, P_n)$, which results from applying an n-ary operator

2. Ellis, Jackson, and Pargetter (1977) use the following law, which we may call the law of *Complete Possibilities* in analogy to the probabilistic law of *Total Probability*:

$$(A \wedge B) \vee (A \wedge \neg B) \dashv\vDash A \qquad \text{(Complete Possibilities)}$$

However, I will use (Absorption) since it does not involve negation and is thus a simpler assumption. In fact, (Absorption) holds in all lattice-based logics, while (Complete Possibilities) is decisively classical, failing in intuitionistic logic, FDE, and many others.

3. See Nute (1980), Fine (2012b), Ciardelli, Zhang, and Champollion (2018), and Santorio (2018).

to sentences $P_1, \ldots, P_n$, is a function of the propositions $[P_i]$, which is determined by the operator O. In the case of counterfactuals, this means that $[A \boxright C]$ functionally depends on $[A]$ and $[C]$ via some function $f_{\boxright}$, i.e.:

$$[A \boxright C] = f_{\boxright}([A], [C]).$$

This observation allows us to derive (Left-Congruentiality) from a PWS model. Suppose that $A \dashv\vDash B$. It follows that $[A] = [B]$, as noted before. But then, by compositionality:

$$[A \boxright C] = f_{\boxright}(\underbrace{[A]}_{=[B]}, [C]) = [B \boxright C].$$

That is, if A and B are equivalent then $A \boxright C$ and $B \boxright C$ are *synonymous*: they express the same proposition. So, *a fortiori*, the two sentences are logically equivalent. In other words, any PWS model entails (Left-Congruentiality). Thus, in order to obtain a hyperintensional framework for counterfactuals, we need an alternative model of propositions.

In light of this, my starting point will be the theory of counterfactuals developed by Fine (2012b), which builds on the TMS account of propositions introduced in the previous chapters. While in PWS $[A]$ is identified with the set $\{w : w \vDash A\}$, in TMS we identify the proposition with the set $\{s : s \Vdash A\}$ of states s that exactly verify A. Using exact truthmaking, the framework allows us to distinguish the propositions expressed by equivalent statements, blocking the derivation of (Left-Congruentiality), and in this way setting the stage for a hyperintensional theory of counterfactuals.

The TMS for counterfactuals developed by Fine is based on the idea of *possible outcomes* of a state. Think, for example, of the window breaking being a possible outcome of throwing a rock against it. Fine postulates the following two principles governing the relation between counterfactuals and possible outcomes: (p. 236)

- *Universal Realizability of the Antecedent* (URA): the truth of a counterfactual requires that the truth of the consequent is an outcome of *every way for the antecedent to be true*.
- *Universal Verifiability of the Consequent* (UVC): the truth of a counterfactual requires the truth of the consequent under *every possible outcome* of the antecedent being true.

Together, these two principles lead to Fine's semantic clause, according to which $A \boxright C$ is true if and only if every outcome of every exact truthmaker of A contains a truthmaker of C. The resulting theory has proven to be a fruitful, hyperintensional framework for counterfactual theorizing; see e.g. Briggs (2012) and Rosella and Sprenger (2023).

The crucial aspect of Fine's theory for the current purposes is that it validates (SDA) without validating (Strengthening). In TMS, a truthmaker of $A \vee B$ is a truthmaker of A or a truthmaker of B:[4]

$$s \Vdash A \vee B \Leftrightarrow s \Vdash A \text{ or } s \Vdash B$$

4. There are variants of the semantics, but the main point here remains unchanged. I will discuss more below.

From this, it immediately follows that if every outcome of every truthmaker for $A \vee B$ contains a truthmaker for C, then both every outcome of every truthmaker for A and every outcome of every truthmaker for B contain truthmakers for C as a part. In other words, if $A \vee B \boxright C$ is true on Fine's semantics, then so are $A \boxright C$ and $B \boxright C$.

At the same time, in TMS, even though A and $A \vee (A \wedge B)$ are logically equivalent (they have the same truth-conditions, i.e. they are true at the same worlds), the truth*makers* of the two can differ. In fact, the *possible outcomes* of making A true and of making $A \vee (A \wedge B)$ true can differ, which gives Fine a natural way of blocking the derivation of (Strengthening) from (SDA) as per Ellis, Jackson, and Pargetter (1977). While the semantics does validate (SDA) and (Absorption), it does not validate (Left-Congruentiality) and (Strengthening).

This makes Fine's truthmaker semantics a good starting point for an SDA-friendly counterfactual logic. But unfortunately, the theory has a hard time accommodating the *prima facie* counterexamples to (SDA), like the paradigmatic Van Inwagen-McKay inference from (3) to (4). Fine (2012b, p. 231–32) formulates a pragmatic analysis of the situation, which explains the apparently invalid inference away in terms of what he calls "Suppositional Accommodation". But, as I shall argue, this analysis is ultimately unsatisfactory. Furthermore, Embry (2014) presents a series of additional counterexamples to Fine's semantics. These are geared specifically towards his theory, but as I shall argue, they are due to the same root issue as the counterexamples of McKay and Van Inwagen (1977) and Lassiter (2018): using principles like (URA) to validate (SDA) ultimately leads to "too many" validities.

Consequently, my proposal is to *restrict* (URA): it is enough for the outcomes of *some* truthmakers of A to contain truthmakers of C for $A \boxright C$ to be true. I will implement this idea by means of a Stalnaker-Lewis-style *selection function*, which instead of acting on worlds, acts on the states of TMS. For now, we can roughly think of the f-selected truthmakers of A as the "most likely" truthmakers for A given the current situation. Roughly, then, my proposal is that $A \boxright C$ is true if and only if the outcome of every f-selected truthmaker of A contains a truthmaker for C as a part. The resulting semantics does not validate (SDA) across all models. It does, however, allow for a *restricted* validity of (SDA) over an interesting class of models, *without* validating (Strengthening) over that same class of models.[5] A major advantage of (re-)introducing selection functions into the TMS frameworks is that it allows us to accommodate the evidence about (SDA) in the aforementioned ways, while at the same time allowing us to accommodate the independent evidence for the hyperintensionality of counterfactuals presented by Fine (2012a, 2012b).

Here is an overview of the rest of the chapter: in Section 6.2, I will briefly discuss matters of syntax to the extent that it is required for the more formal aspects of this chapter. In Section 6.3, I will review the Stalnaker-Lewis semantics and how it deals with (SDA). In particular, I will discuss the often overlooked fact that Lewis (1971) has a way of analyzing a class of seemingly valid instances of (SDA) as *enthymemes*. However, since I will argue that the approach ultimately fails to properly account for *all*

5. A similar result is achieved by Santorio (2018).

valid instances of (SDA), I will move on to the (SDA)-friendly truthmaker approach of Fine in Section 6.4. I will then discuss the problems displayed by the account involving determinable antecedents and (SDA) in Section 6.5, and some unsatisfactory solutions to them suggested by Embry (2014) in Section 6.6. I will move on to discuss my own proposal in detail in Section 6.7. Finally, in Section 6.8 I will sketch some possible refinements of the accounts, required to deal with counterpossibles and nested counterfactuals.

6.2 Syntax

For most of this chapter, following Fine (2012b), I restrict my attention to *first-degree* counterfactuals, namely, counterfactuals of the form $A \boxright B$, where neither A nor B contains the counterfactual operator $\boxright$.[6] To be a bit more explicit, for whenever I will be working with formal models, I will assume a fixed background set $\mathcal{P} = \{p, q, r, \ldots\}$ of propositional variables. The propositional language $\mathcal{L}_{\neg,\wedge,\vee}$ is defined over $\mathcal{P}$ adopting the propositional connectives $\neg, \wedge, \vee$. The restriction to $\neg, \wedge, \vee$ is common in TMS, mainly because there is no standard TMS clause for the conditional $A \supset B$ and biconditional $A \equiv B$ other than treating them as defined in the usual way as $\neg A \vee B$ and $(A \supset B) \wedge (B \supset A)$, respectively. For the present purpose, however, we can follow suit and just treat the conditional operators as defined. Since (SDA) only concerns the interaction of $\boxright$ and $\vee$, nothing much is going to depend on the treatment of other conditional operators.

The *Backus-Naur Form (BNF)* of $\mathcal{L}_{\neg,\wedge,\vee}$ is then:

$$A ::= p \mid \neg A \mid (A \wedge A) \mid (A \vee A)$$

In general, I will use $A, B, C, \ldots$ as variables ranging over $\mathcal{L}_{\neg,\wedge,\vee}$.

A first-degree counterfactual, then, is any expression of the form:

$$A \boxright B,$$

where $A, B \in \mathcal{L}_{\neg,\wedge,\vee}$.

Occasionally, I shall be working with $\mathcal{L}_{\neg,\wedge,\vee}$-formulas and first-degree counterfactuals in the same context. This happens, for example, when considering:

$$A, A \boxright C \vDash C \qquad \text{(Counterfactual Modus Ponens)}$$

6. I need to stress that in more recent works, in particular Fine (2023b), a full-fledged *truthmaker* clause for counterfactuals is provided, which allows to deal also with nested counterfactuals. While it would be interesting to explore an analogous exactification of my own proposal from the outset, the issues dealt in the present chapter are largely independent from this choice. Furthermore, the principles upon which Fine relies for his exact clause that are relevant for the present discussion are very close to those that I will criticize in the next sections. So, for the sake of simplicity, I will stick to Fine (2012b)'s proposal for most of the chapter. I will say something about full truthmaker clauses for counterfactuals in sec. 6.8.

Similarly, I shall sometimes consider truth-functional combinations of $\mathcal{L}_{\neg,\wedge,\vee}$-formulas and first-degree counterfactuals, as in the (Nesting) axiom that will be discussed in the next section:

$$(A \vee B \mathrel{\Box\!\!\rightarrow} A) \vee (A \vee B \mathrel{\Box\!\!\rightarrow} B) \vee [A \vee B \mathrel{\Box\!\!\rightarrow} C \equiv (A \mathrel{\Box\!\!\rightarrow} C) \wedge (B \mathrel{\Box\!\!\rightarrow} C)]$$

Technically, then, I will be working in the language $\mathcal{L}_{\neg,\wedge,\vee,\Box\rightarrow}$ with BNF:

$$\phi ::= p \mid \neg\phi \mid (\phi \wedge \phi) \mid (\phi \vee \phi) \mid (A \mathrel{\Box\!\!\rightarrow} A)$$

I will adopt the usual conventions regarding parentheses, where $\neg$ binds stronger than $\wedge$, which binds as strongly as $\vee$, and the weakest connective is $\Box\!\!\rightarrow$. Most of the time, however, it will be clear from the context which language I am working with.

6.3 SDA in Stalnaker-Lewis Semantics

In this section, I will review how the Stalnaker-Lewis semantics deals with (SDA). As we will see, the system has *something* interesting to say about (SDA), but, I shall argue, it is ultimately unable to deal with the positive evidence for (SDA) in a satisfactory fashion.

Different implementations of the semantic ideas underlying the Stalnaker-Lewis approach exist, and not all systems agree with each other. The *loci classici* are Stalnaker (1968), Lewis (1973b), and Nute (1980). In this chapter, we will work with a selection function-based implementation following Lewis (1971). Not much depends on this concrete implementation choice, but Lewis' system allows for a straightforward discussion of the relevant issues.

The selection function semantics builds on the general PWS framework, which starts from a (non-empty) set $W = \{w, v, u, \ldots\}$ of worlds. In line with PWS, in this setting we think of sets of worlds as propositions, that is P is a proposition over W if and only if $P \subseteq W$.

A *frame* $\mathcal{F} = \langle W, f \rangle$ is the result of equipping the set W with a *selection function* $f : W \times \wp(W) \rightarrow \wp(W)$. Intuitively, $f(w, P)$ is the set of the most-similar-to-w P-worlds, the idea being that the worlds in $f(w, P)$ are "just like" w except for changes necessary to accommodate the truth of P.[7] The role of the selection function is to interpret counterfactuals following the Stalnaker-Lewis idea that $A \mathrel{\Box\!\!\rightarrow} C$ is true at w if and only if C is true in all the most-similar-to-w A-worlds. The decision of using selection functions for this purpose rather than similarity spheres in the style of Lewis (1973b), for example, is one of those implementation details mentioned above which are largely irrelevant to the issues at hand.

7. Notice that since f takes a pair consisting of a world and a *proposition* as input, the correct notation would be $f(w, [P])$, but I will stick to the simplified notation $f(w, P)$ for the rest of the chapter.

In line with the intended informal reading, standard practice is to postulate a series of conditions on f. The postulates of Lewis (1971) are:

$$f(w, P) \subseteq P \qquad \text{(Success)}$$
$$f(w, P) = \{w\}, \text{ if } w \in P \qquad \text{(Strong Centering)}$$
$$f(w, P) \subseteq Q \text{ and } f(w, Q) \subseteq P \Rightarrow f(w, P) = f(w, Q) \qquad \text{(Uniformity)}$$
$$f(w, P \cup Q) \subseteq P,\ f(w, P \cup Q) \subseteq Q, \text{ or } f(w, P \cup Q) = f(w, P) \cup f(w, Q) \qquad \text{(Nesting)}$$

Each of these conditions corresponds to logical laws for counterfactuals. For example, (Success) is a minimal condition that guarantees the validity of

$$\vDash A \mathrel{\Box\!\!\rightarrow} A \qquad \text{(Identity)}$$

(Strong Centering), instead, corresponds to the validity of (Counterfactual Modus Ponens). (Nesting), instead, will be crucial in the analysis of (SDA) below.

A *model* $\mathcal{M} = \langle W, f, v \rangle$ results from taking a frame $\langle W, f \rangle$ and equipping it with a valuation function $v : \mathcal{P} \to \wp(W)$, which assigns a proposition $v(p) \subseteq W$ to each propositional variable $p \in \mathcal{P}$ (i.e. the set of worlds where p is true according to the model). We extend this assignment to the $\mathcal{L}_{\neg,\wedge,\vee}$-formulas by recursively defining the proposition $[A]_W$ expressed by A in the model by:

$$[p]_W = v(p) \qquad (W\text{-propositional Atom})$$
$$[\neg A]_W = W \setminus [A]_W \qquad (W\text{-propositional Negation})$$
$$[A \wedge B]_W = [A]_W \cap [B]_W \qquad (W\text{-propositional Conjunction})$$
$$[A \vee B]_W = [A]_W \cup [B]_W \qquad (W\text{-propositional Disjunction})$$

Following the idea underlying the PWS model, i.e. $[A]_W = \{w \in W : w \vDash A\}$, we define truth at a world for $\mathcal{L}_{\neg,\wedge,\vee}$ formulas by setting:

$$w \vDash A \Leftrightarrow w \in [A]_W \qquad (W\text{-propositional Truth})$$

Moving to counterfactuals, the semantic clause for counterfactual propositions is:

$$[A \mathrel{\Box\!\!\rightarrow} C]_W = \{w \in W : f(w, [A]_W) \subseteq [C]_W\} \qquad (W\text{-propositional Counterfactual})$$

Extending (W-propositional Truth) to counterfactuals, we get:

$$w \vDash A \mathrel{\Box\!\!\rightarrow} B \Leftrightarrow f(w, [A]_W) \subseteq [C]_W \qquad (W\text{-Counterfactual Truth})$$

In a similar vein, truth at a world is defined for all $\mathcal{L}_{\neg,\wedge,\vee,\Box\rightarrow}$-formulas ϕ by adding the following clauses:

$$w \vDash \neg\phi \Leftrightarrow w \nvDash \phi \qquad (W\text{-Negation Truth})$$
$$w \vDash \phi \wedge \psi \Leftrightarrow w \vDash \phi \text{ and } w \vDash \psi \qquad (W\text{-Conjunction Truth})$$
$$w \vDash \phi \vee \psi \Leftrightarrow w \vDash \phi \text{ or } w \vDash \psi \qquad (W\text{-Disjunction Truth})$$

Under the common view of consequence as necessary truth-preservation, this semantics defines a counterfactual logic by saying that for all $\phi \in \mathcal{L}_{\neg,\wedge,\vee,\boxrightarrow}$ and $\Gamma \subseteq \mathcal{L}_{\neg,\wedge,\vee,\boxrightarrow}$:

$$\Gamma \vDash \phi \Leftrightarrow \text{in all } \mathcal{M}, \text{ we have } w \vDash \phi, \text{ if } w \vDash \gamma, \text{ for all } \gamma \in \Gamma \qquad (W\text{-Consequence})$$

That is, a premise set Γ implies a conclusion ϕ if and only if for all worlds $w \in W$ in all models, if all Γs are true at w, then so is ϕ.

Lewis (1971) determines the resulting logic in the current setting (modulo our restriction to first-degree counterfactuals). Obviously, changing the conditions on f affects the logic. [8]

As mentioned above, (SDA) is invalid under the Lewis-Stalnaker semantics. An instructive example to discuss is provided by a formal analysis of (3) and (4a) and (4b):

(3) If Spain had fought with the Allies or the Axis, then Spain would have fought with the Axis.

(4) a. If Spain had fought with the Allies, then Spain would have fought with the Axis.
b. If Spain had fought with the Axis, then Spain would have fought with the Axis.

If we let p stand for "Spain fights with the Axis" and q for "Spain fights with the Allies',' the inference becomes:

(3) $p \vee q \boxrightarrow p$

(4) a. $q \boxrightarrow p$
b. $p \boxrightarrow p$

The problematic inference is from (3) to (4a). Here is a simple Stalnaker-Lewis countermodel for this inference:

- $W = \{w_1, w_2, w_3\}$
- $f(w_1, \{w_2, w_3\}) = \{w_2\}$
- $f(w_1, \{w_2\}) = \{w_2\}$
- $f(w_1, \{w_3\}) = \{w_3\}$
- $v(p) = \{w_2\}$ and $v(q) = \{w_3\}$

It is easily checked that this is indeed a model (i.e. the conditions on the selection function are satisfied). Moreover, it straightforwardly captures the intuitive understanding of the situation: Since $w_2 \vDash p$ and p means that Spain fights with the Axis, and $w_3 \vDash q$ and q means that Spain fights with the Allies, $f(w_1, \{w_2, w_3\}) = \{w_2\}$ simply means that the unique closest world to ours where Spain fights with the Axis *or* Allies (i.e. where $\{w_2, w_3\}$ is true) is the world where Spain fights with the Axis (i.e. w_2). In fact, $p \vee q \boxrightarrow p$ is true at w_1, while $q \boxrightarrow p$ is not. To see this:

8. For a systematic overview of the various systems resulting from different conditions on f, see, for example, the classic Nute (1980). Burgess (1981) provides a particularly pleasant proof theory.

- Note that $[p \vee q]_W = \{w_2, w_3\}$ and so, $f(w_1, [p \vee q]_W) = f(w_1, \{w_2, w_3\}) = \{w_2\} = [p]_W$. This means that $f(w_1, [p \vee q]_W) \subseteq [p]_W$ and so $w_1 \vDash p \vee q \boxrightarrow p$.
- At the same time, $f(w_1, [q]_W) = f(w_1, \{w_3\}) = \{w_3\}$. And so, since $[p]_W = \{w_2\}$, $f(w_1, [q]_W) \nsubseteq [p]_W$, which just means that $w_1 \nvDash q \boxrightarrow p$.

In this way, the Stalnaker-Lewis semantics quite naturally handles the counterexamples to (SDA).

But what about the seemingly *valid* instances of (SDA), like the inference from (1) to (2a) and (2b)? The first thing to notice is that the logic satisfies both:

$$A \eqvDash B \quad \Rightarrow \quad A \boxrightarrow C \eqvDash B \boxrightarrow C \qquad \text{(Left-Congruentiality)}$$

$$C \eqvDash D \quad \Rightarrow \quad A \boxrightarrow C \eqvDash A \boxrightarrow D \qquad \text{(Right-Congruentiality)}$$

The argument is essentially the one sketched above: from (W-propositional Truth) the following principle directly follows:

$$A \eqvDash B \Rightarrow [A]_W = [B]_W \qquad \text{(Intensionality)}$$

So, assuming $A \eqvDash B$ and thus $[A]_W = [B]_W$, we get:

$$[A \boxrightarrow C] = \{w \in W : f(w, \underbrace{[A]_W}_{=[B]_W}) \subseteq [C]_W\} = [B \boxrightarrow C]_W$$

This immediately gives us (Left-Congruentiality) by applying (W-propositional Truth). The argument for (Right-Congruentiality) is completely analogous. Note further that the background logic for $\neg, \wedge, \vee$ is classical. In particular, we get the absorption law in its two relevant forms:

$$A \eqvDash A \vee (A \wedge B)$$

$$[A]_W = [A \vee (A \wedge B)]_W \qquad \text{(Absorption)}$$

Correspondingly, the Stalnaker-Lewis semantics is subject to the Ellis, Jackson, and Pargetter (1977) problem: we cannot have (SDA) be valid over any class of models of the logic without at the same time validating the problematic principle (Strengthening) over the same class of models:

$$A \boxrightarrow C \vDash (A \wedge B) \boxrightarrow C \qquad \text{(Strengthening)}$$

In light of this, is there a way for the Stalnaker-Lewis semantics to account for the apparently valid inferences? It turns out that it is possible using the (Nesting) condition,

$$f(w, P \cup Q) \subseteq P, \text{ or } f(w, P \cup Q) \subseteq Q, \text{ or } f(w, P \cup Q) = f(w, P) \cup f(w, Q),$$

which is often overlooked in the literature. Logically, this condition postulated by Lewis (1971) corresponds to the following principle:

$$(A \vee B \boxrightarrow A) \vee (A \vee B \boxrightarrow B) \vee [A \vee B \boxrightarrow C \equiv (A \boxrightarrow C) \wedge (B \boxrightarrow C)] \qquad \text{(Nesting Axiom)}$$

Since the background logic is classical in the Stalnaker-Lewis semantics, this axiom entails the validity of the following inference:

$$\begin{array}{l}(A \vee B \boxright C), \neg(A \vee B \boxright A), \neg(A \vee B \boxright B) \vDash A \boxright C \\ (A \vee B \boxright C), \neg(A \vee B \boxright A), \neg(A \vee B \boxright B) \vDash B \boxright C\end{array} \qquad \text{(Enthym. SDA)}$$

This principle can be called *enthymematic* SDA, since it allows for an analysis of the positive (SDA) instances, like the inference from (1) to (2), as *enthymemes*: inferences with hidden premises. Recall the example:

(1) If it were rainy or windy, then I would be cold.

(2) a. If it were rainy, then I would be cold.
    b. If it were windy, then I would be cold.

Letting p stand for "it is rainy", q for "it is windy", and r for "I am cold", the inference becomes:

(1) $p \vee q \boxright r$

(2) a. $p \boxright r$
    b. $q \boxright r$

First of all, it is easy to see that the inference from (1) to both (2a) and (2b) is invalid in the Stalnaker-Lewis semantics for essentially the same reasons why the inference from (3) to (4a) and (4b) is invalid. In fact, we can take the same frame as before:

- $W = \{w_1, w_2, w_3\}$
- $f(w_1, \{w_2, w_3\}) = \{w_2\}$
- $f(w_1, \{w_2\}) = \{w_2\}$
- $f(w_1, \{w_3\}) = \{w_3\}$

We equip this frame with the valuation v:

- $v(p) = \{w_2\}, v(q) = \{w_3\}$, and $v(r) = \{w_2\}$.

And we get:

- Since $[p \vee q]_W = \{w_2, w_3\}$, $f(w_1, [p \vee q]_W) = f(w_1, \{w_2, w_3\}) = \{w_2\} = [r]_W$. Hence $f(w_1, [p \vee q]_W) \subseteq [r]_W$ and so $w_1 \vDash p \vee q \boxright r$.
- At the same time, $f(w_1, [q]_W) = f(w_1, \{w_3\}) = \{w_3\}$. And so, since $[r]_W = \{w_2\}$, $f(w_1, [q]_W) \nsubseteq [r]_W$, which just means that $w_1 \nvDash q \boxright r$.

So, while (1) is true, (2b) is not. However, in contexts where the inference from (1) to (2) seems valid, it may seem at least *prima facie* plausible that the following are true:

(7) a. It is not the case that if it were rainy or windy, then it would be rainy.
$\neg(p \vee q \boxrightarrow p)$

b. It is not the case that if it were rainy or windy, then it would be windy.
$\neg(p \vee q \boxrightarrow q)$

One way to make this plausible is to think about under which conditions we would normally affirm (1), i.e. when we would say that we would be cold if it were rainy or windy. It may be argued that a condition under which to affirm (1) is precisely the kind of *uncertainty* about the weather expressed by (7). If we are uncertain about what non-actual weather condition is the most likely, we should not be able to assert neither that if it were rainy or windy then it would be rainy, nor that if it were rainy or windy then it would be windy, since both options amount to asserting that a certain weather condition is the one that would in fact obtain. Nonetheless, we would be able to assert both (1) and (2) regardless of our uncertainty about the weather: it seems perfectly plausible to claim that we would be cold in either case.

Notice that our previous model no longer works to reject (Enthym. SDA), since it fails to make the "hidden" premise $\neg(p \vee q \boxrightarrow p)$ true at w_1. Since $f(w_1, [p \vee q]_W) \subseteq \{w_2\} = [p]_W$, we have $w_1 \vDash p \vee q \boxrightarrow p$.

Furthermore, (Enthym. SDA) is *not* threatened by the argument of Ellis, Jackson, and Pargetter (1977). If we assume $A \boxrightarrow C$, we can, by (Left-Congruentiality) and (Absorption), infer that $A \vee (A \wedge B) \boxrightarrow C$. But from this we can only infer $A \wedge B \boxrightarrow C$ via (Enthym. SDA) if we can derive the required "hidden premise" case $\neg(A \vee (A \wedge B) \boxrightarrow A)$. By again applying (Absorption) and (Left-Congruentiality), this is easily seen to be equivalent to $\neg(A \boxrightarrow A)$, which directly contradicts (Identity). This means that in the current setting, the argument of Ellis, Jackson, and Pargetter (1977) only goes through using (Enthym. SDA) if we are dealing with inconsistent background assumptions.

So far, so good. This seems to account for at least *some* seemingly valid instances of (SDA). Unfortunately, however, even if (Enthym. SDA) may work in some contexts, the Stalnaker-Lewis approach ultimately fails to adequately account for *all* seemingly valid instances of (SDA). Suppose that neither the Labors nor the Conservatives won the election, the Conservatives got way more votes than the Labors, and they both planned to raise taxes. Now consider the following inference:

(8) If the Labors or the Conservatives won the election, then taxes would go up.

(9) a. If the Labors won the election, then taxes would go up.

b. If the Conservatives won the election, then taxes would go up.

This looks like a straightforwardly valid (SDA) application. Note, however, that in this case an accurate model where (8) and (9) are true should arguably *not* make (10) false:

(10) If the Labors or the Conservatives won the election, then Conservatives would have won the election.

Indeed, an accurate model for the just assumed setup should make (10) *true*, since according to the actual scenario it was significantly more likely that the Conservatives would have won. But this just means that (Enthym. SDA) does not account for all the cases in which (SDA) seems valid.

In light of this, setting aside the worries related to vacuism discussed above, we can summarize the main reasons why the Stalnaker-Lewis approach alone is unsatisfactory as follows:

1. Fine (2012a, 2012b) presents powerful arguments against (Left-Congruentiality) that are *independent of* considerations concerning (SDA). The corresponding worries are dealt within Fine's semantics but *cannot* be dealt within Stalnaker-Lewis semantics (for the reasons discussed in the introduction). At the same time, Fine's semantics universally validates (SDA), which in light of the various counterexamples that will be discussed in this chapter, is not tenable.

2. There is something that the Stalnaker-Lewis semantics *cannot* achieve. The argument of Ellis, Jackson, and Pargetter (1977) shows that we cannot have any form of validity, even over a restricted class of models, of (SDA), on pain of validating (Strengthening) over those models. Moreover, the attempt of accounting for the seemingly valid instances of (SDA) through (Enthym. SDA) does not ultimately comply with our semantic intuitions.

In the next sections, I will transfer the Stalnaker-Lewis analysis of (SDA) to the hyperintensional setting of TMS while, at the same time, allowing for the possibility of having (SDA) valid over certain classes of models.

6.4 Counterfactuals in Truthmaker Semantics

I briefly discussed the general idea behind propositions in TMS in chap. 4. In this section, we will see in more detail how a semantics of this kind works, in order to analyze Fine's (2012b) application to counterfactuals.

The basic frame underlying any TMS model is a *state space* $\langle S, \sqsubseteq \rangle$, where S is a set of states and $\sqsubseteq$ is a binary *parthood* relation on S. A state s is any actual or possible entity that can verify (or falsify) a sentence, or a *way* for a sentence to be true (or false).

The relation $\sqsubseteq$ induces a partial order over S by endowing the states with a mereological structure: states in S can be part of each other and fused together. Given any two states s and t, their fusion $s \sqcup t$ will be their *least upper bound*, i.e. the smallest state that contains s and t as parts, and it will belong to S.

We have seen that a fundamental difference with possible worlds is that states can be partial, namely, they are not supposed to settle everything that is the case according to a given complete scenario. However, S may also contain *world-states*, namely possible states that have as a part any other state compatible with them. To make this precise, we can define a W-space as a triple $\langle S, S^\diamond, \sqsubseteq \rangle$ where $S^\diamond \subseteq S$ is a non-empty set of *possible* states. Furthermore, $S^\diamond$ will be *downward closed*: if $s \in S^\diamond$ and $t \sqsubseteq s$, then $t \in S^\diamond$. Within

this setup, we can define compatibility between states as follows: a set of states $T \subseteq S$ is compatible when $\bigsqcup T \in S^\diamond$, and incompatible otherwise (where $\bigsqcup T$ is the fusion of all states in T). In a W-space, all possible states are part of a world-state. We will call W the set of world-states.

In order to develop a semantics for counterfactuals, two notions of truthmaking are required: *exact* $\Vdash$ and *inexact* $\Vdash\!\!\!\!\Vdash$ verification. A state s exactly verifies a sentence P when s is wholly relevant for the truth of P. For instance, if s is the state of the snow being white, s will be an exact verifier of "snow is white". A state s inexactly verifies a sentence P if s has an exact verifier of P as a part. For instance, if s is the state of the snow being white and the grass being green, s will be an inexact verifier of "snow is white".

As observed earlier in chap. 4, the notion of exact verification can be paired with a notion of exact *falsification* $\dashv$ to provide a richer picture of semantic content. The kind of TMS model $\mathcal{M} = \langle S, S^\diamond, \sqsubseteq, v^+, v^- \rangle$ with which we will be working results from taking a W-space and equipping it with two valuation functions $v^+, v^- : \mathcal{P} \to \wp(S)$ that assign a positive content $v^+(p) \subseteq S$ and a negative content $v^-(p) \subseteq S$ to each atom $p \in \mathcal{P}$. We will denote $[A]^+$ the set of A's exact verifiers, and $[A]^-$ the set of A's exact falsifiers, and we will require them to be non-empty and non-overlapping.

We will be working with an *inclusive* semantics, so we can spell out the clauses for the basic connectives as follows:

$$s \Vdash \neg A \Leftrightarrow s \dashv A \qquad \text{(Exact } \neg^+\text{)}$$

$$s \dashv \neg A \Leftrightarrow s \Vdash A \qquad \text{(Exact } \neg^-\text{)}$$

$$s \Vdash A \wedge B \Leftrightarrow s = t \sqcup u, t \Vdash A \text{ and } u \Vdash B \qquad \text{(Exact } \wedge^+\text{)}$$

$$s \dashv A \wedge B \Leftrightarrow s \dashv A, \text{ or } s \dashv B, \text{ or } s \dashv A \vee B \qquad \text{(Exact } \wedge^-\text{)}$$

$$s \Vdash A \vee B \Leftrightarrow s \Vdash A, \text{ or } s \Vdash B, \text{ or } s \Vdash A \wedge B \qquad \text{(Exact } \vee^+\text{)}$$

$$s \dashv A \vee B \Leftrightarrow s = t \sqcup u, t \dashv A \text{ and } u \dashv B \qquad \text{(Exact } \vee^-\text{)}$$

This allows us to extend the assignment of propositions to the formulas of $\mathcal{L}_{\neg,\wedge,\vee}$ recursively, corresponding to:

$$[p]^+ = v^+(p) \qquad [p]^- = v^-(p) \qquad \text{(S-Propositional Atom)}$$

$$[\neg A]^+ = [A]^- \qquad [\neg A]^- = [A]^+ \qquad \text{(S-Propositional Negation)}$$

$$[A \wedge B]^+ = \{s \mid s \Vdash A \wedge B\} \qquad [A \wedge B]^- = \{s \mid s \dashv A \wedge B\} \qquad \text{(S-Propositional Conjunction)}$$

$$[A \vee B]^+ = \{s \mid s \Vdash A \vee B\} \qquad [A \vee B]^- = \{s \mid s \dashv A \vee B\} \qquad \text{(S-Propositional Disjunction)}$$

As remarked above, it is easy to show how such a framework has the resources to track hyperintensional distinctions relevant for the present discussion: consider A and $A \vee (A \wedge B)$. While they are intensionally equivalent (i.e., they are true at the same worlds according to any PWS-model), they might have different verifiers (there might be verifiers for B only among the verifiers of $A \vee (A \wedge B)$, but not among those of A). This is enough to invalidate (Left-Congruentiality) and the related worries concerning (Strengthening).

We obtain a counterfactual model $\mathcal{CM} = \langle S, S^\diamond, \sqsubseteq, \mapsto, v^+, v^- \rangle$ by equipping a W-space with valuation functions and a *transition* relation $\mapsto \subseteq W \times S \times S$. This is a ternary relation on worlds, states and states. Within the current setup, $s \mapsto_w t$ can be read as: t is an outcome of imposing the change s on the world w. [9] An intuitive way of understanding this is by considering the state of throwing a rock against a window, which leads to (or *transitions* into) worlds (i.e. possible outcomes) containing the state of the window shattering as a part.

Fine (2012b) places some reasonable conditions on the transition relation:

$$\text{If } s \mapsto_w t \text{ then } s \sqsubseteq t \qquad \text{(Inclusion)}$$

$$\text{If } s \sqsubseteq w \text{ then } s \mapsto_w t \text{ for some } t \sqsubseteq w \qquad \text{(Actuality)}$$

$$\text{If } s \mapsto_w t \text{ and } t' \sqsubseteq t \text{ then } s \sqcup t' \mapsto_w t \qquad \text{(Incorporation)}$$

$$s \mapsto_w t \text{ only if } t \text{ is a world-state} \qquad \text{(Completeness)}$$

Each of them guarantees the validity of desirable logical principles.[10]

Now that we have the whole background machinery in place, we can finally move to Fine's clause for counterfactuals, which relies on two basic assumptions: Universal Realizability of the Antecedent (URA) and Universal Verifiability of the Consequent (UVC). The former is a claim about which states in the antecedent should be considered for the evaluation of a counterfactual: in Fine's opinion, all of them. In other words, every verifier of the antecedent will be relevant for the truth conditions of the counterfactual. The latter principle says that a counterfactual is verified by any *outcome* of some exact verifier of the antecedent. Since an outcome of a given state might contain also states that are irrelevant for the truth of the consequent, Fine proposes to capture (UVC) in terms of inexact verification. This leads us to the truth-condition for counterfactuals:

$$w \vDash A \boxdot\!\!\rightarrow C \Leftrightarrow \forall s, t \text{ such that } s \Vdash A \text{ and } s \mapsto_w t, t \nVdash C \qquad \text{(TC)}$$

This can be read as follows: a world-state w verifies $A \boxdot\!\!\rightarrow C$ if and only if every outcome at w of each exact verifier of A is an inexact verifier of C. Or, equivalently: a world-state w verifies $A \boxdot\!\!\rightarrow C$ if and only if every exact verifier of A *transitions* at w into an inexact verifier of C.

Notice that (TC) tells us only when a counterfactual is verified by a world, but remains silent about what an *exact* verifier for it should look like. As said above, this is not a substantial issue for the present purposes, but I will nevertheless discuss a possible way to exactify the clause at the end of the chapter.

Moving back to the main topic, it is easy to notice that within the just sketched framework, (URA) guarantees the validity of (SDA). Suppose that $w \vDash A \vee B \boxdot\!\!\rightarrow C$.

9. The notion of 'outcome' employed here is not thoroughly clarified, but Fine (2012b) claims that it should be understood broadly enough to cover cases beyond the causal ones.

10. Specifically, (Inclusion) validates (Identity); (Actuality) validates (Counterfactual Modus Ponens); (Incorporation) validates (Restricted Transitivity); and (Completeness) validates (Classical Weakening) in the consequent, allowing for the evaluation of counterfactuals with embedded counterfactuals in the consequent.

Then, all exact verifiers of $A \vee B$ lead to some inexact verifier of C. Since $[A]^+ \subseteq [A \vee B]^+$ and $[B]^+ \subseteq [A \vee B]^+$, all exact verifiers of A and all exact verifiers of B must lead to some inexact verifier of C. But this just means that $w \vDash A \boxdot\!\!\rightarrow C$ and $w \vDash B \boxdot\!\!\rightarrow C$.

I will now turn to the critical assessment of (SDA) and the related problems for Fine's semantics.

6.5 (More) Problems with SDA in Truthmaker Semantics

As seen earlier, (SDA) stands as one of the most contentious principles within the debate on counterfactual logic. Lewis' strategy for its rejection relied on (Left-Congruentiality), which together with (SDA) allows to derive (Strengthening), a principle that appears incompatible with the nature of counterfactual reasoning.

In TMS there is no need to worry about (Left-Congruentiality) anymore. Indeed, by assuming (URA), Fine's original proposal allows him to retain (SDA) – which he believes to be a desirable principle – without validating (Strengthening). Nevertheless, it is widely acknowledged that there exist *independent* grounds to challenge (SDA). Recall the example by McKay and Van Inwagen (1977):

(3) If Spain had fought with the Allies or the Axis, then Spain would have fought with the Axis.

While (3) seems true, it does not seem to entail:

(4) a. If Spain had fought with the Allies, then Spain would have fought with the Axis.

I will soon discuss the standard reply to this kind of more general objections to (SDA), but let us first consider a related class of counterexamples that are specifically directed at Fine's account: counterfactuals involving *determinable* antecedents. Embry (2014) asks us to consider a scenario in which the following is true:

(11) If Sue were to take some of these pills, she would get better ($A \boxdot\!\!\rightarrow C$).

Suppose that $s \Vdash$ "Sue takes 25 of these pills". It seems plausible to assume that $s \in A$. Fine's clause relies on (URA), so s should be taken into account for the evaluation of (11). But since it is not the case that, had Sue taken 25 pills, she would still get better, any inexact verifier of C will not be an outcome of s. So, it is not the case that every exact verifier of A leads to some inexact verifier of C. Hence, according to Fine's clause (11) turns out as false, contrary to our assumption.

Even though Embry does not address the problem from this perspective, it is easy to notice how, within Fine's TMS, determinable propositions can be naturally thought to be equivalent to the disjunctions of their corresponding *determinate* propositions. Indeed, Fine (2017a) distinguishes between *definite* propositions – those of the form $P = \{p\}$, for some state p – and *disjunctive* propositions, i.e. those that are verified by multiple states.

In other words, determinable antecedents are equivalent to disjunctive antecedents in a very natural version of Fine's TMS. [11]

For instance, it seems reasonable to expect an appropriate model to make the antecedent of (11) exactly equivalent to a long (possibly infinite) disjunction of definite propositions about the precise amount of pills that Sue could take. Let A_n be "Sue takes n pills" and s_n be the state of Sue taking n pills. Then $A = \bigvee_{n>0} A_n$, where $A_n = \{s_n\}$. One can easily see how applying (SDA) to (11) in such a model allows one to infer false counterfactuals such as "if Sue were to take 25 of these pills, she would get better".

To better see this point, consider a simpler case involving only three disjuncts. Suppose that John has various shades of blue paint, including azure and cyan, and no other colors. He paints a canvas in azure. Now consider the following (keeping in mind that cyan is the only primary color among those currently available to John):

(12) If John had painted the canvas in a primary color, he would have painted it in cyan.

The antecedent of (12) is determinable, and it is equivalent to the following disjunction: "either John paints the canvas in cyan, or in magenta, or in yellow". If (12) is true, by (SDA) we get:

(13) a. If John had painted the canvas in magenta, he would have painted it in cyan.

b. If John had painted the canvas in yellow, he would have painted it in cyan.

Both (13a) and (13b) are clearly absurd. But (12) strikes us as a fairly innocuous truth: given the actual setup, the most plausible way for John to paint the canvas in a primary color is obviously that in which he paints it in cyan.

So, what is the standard reply from defenders of (SDA)? Usually, it appeals to some pragmatic phenomena allegedly involved in asserting this kind of counterfactuals. According to Fine (2012a), by asserting (3) we exclude certain scenarios from the domain of relevant possibilities: those in which Spain fought with the Allies. But when we subsequently try to evaluate (4a), we tend to consider precisely those scenarios, as a result of a process of "Suppositional Accomodation". Taking the antecedent of (4a) at face value, i.e. as non-empty, requires to admit possibilities that had been excluded in the first place by asserting (3). This leads to the apparent failure of (SDA), which is nothing more than the result of applying a fallible heuristics. [12]

A similar point has been raised by Willer (2015):

11. More precisely, in any TMS-model that requires propositions to be at least *closed* (a proposition P is closed if and only if $\bigsqcup Q \in P$ for every non-empty $Q \subseteq P$), as those described in Fine (2017a), a determinable proposition *just is* a disjunction of its corresponding definite propositions, since they will always have the same exact verifiers. This identity breaks down only within a *full* domain of proposition, where no constraint for propositionhood is placed. Notice also that, regardless of any specific semantic setting, the disjunctive reading is a common strategy for reducing determinable properties ascriptions to determinate ones even in the context of metaphysics. See Dosanjh (2021) for a recent account of this kind.

12. In Fine's view, (SDA) will obviously still be valid, since asserting (3) "excludes" all verifiers of the antecedent of (4a), making it empty. This amounts to considering (4a) as *vacuously* true.

> there is a principled pragmatic explanation for why certain counterfactuals resist simplification [(SDA)]. The reason, in brief, is that certain counterfactuals imply information in addition to asserting a strict material conditional, and that this additional information may in fact eliminate possibilities that have been brought into view by presupposed content (p. 435).

In light of this, an important question that needs to be addressed before further developing the case against (SDA) is the following: is it really the case that every *prima facie* failure of (SDA) can be explained away with pragmatics?

Before developing an answer, there is an important fact to be stressed. Just like any other framework in which the De Morgan laws hold, Fine's TMS generally makes $\neg(A \wedge B)$ *exactly* equivalent to $\neg A \vee \neg B$. As a result, within this framework (SDA) not only concerns disjunctive antecedents, but also *negated conjunctive* antecedents. [13] However, Willer observes that:

> not all counterfactuals with negated conjunctions simplify: "If John had not had that terrible accident last week and died, he would have been here today" does not license "If John had not died, he would have been here today" since he still might have had that accident. What underlies this observation, I suggest, is that negated conjunctions sometimes give rise to a 'neither' rather than a 'not both' reading and that, unsurprisingly, only the latter licenses simplification (p. 429, footnote 1).

Here, an alternative explanation is offered for alleged failures of simplification with negated conjunctions: the correct logical form of antecedents that *do not* simplify is $\neg A \wedge \neg B$, rather than $\neg(A \wedge B)$. Only the latter allows for the 'not both' reading, for which simplification is valid.

Now, suppose that John is a bachelor and that he would rather get married than change his gender. The following seems true:

(14) If John were not a bachelor, then he would be married.

As a matter of analyticity, the antecedent of (14) is equivalent to the negation of "John is unmarried and John is a man". So, in any model validating De Morgan, it will be equivalent to "either John is married or John is not a man". If (14) is true, then by (SDA) we get:

(15) If John were not a man, then he would be married.

Not only (15) seems false in this specific context, but one's being married does not seem to be a possible outcome of one's not being a man under any reasonable interpretation of the transition relation.

13. Some authors have labeled this phenomenon as Simplification of Negated Conjunctive Antecedents (SNCA), and used putative counterexamples to it as an argument against the equivalence between $\neg(A \wedge B)$ and $\neg A \vee \neg B$; see for instance Ciardelli, Zhang, and Champollion (2018). In the present work I grant the equivalence and consider counterexamples to (SDA) and (SNCA) as two faces of one and the same phenomenon. See Alonso-Ovalle (2009) and Starr (2014) for alternative approaches where (SDA) is valid and (SNCA) fails.

What sort of reading is conveyed by the antecedent of (14)? Surely, it does not force the stronger 'neither' reading (though it seems to allow for it), for which even according to Willer (SDA) can fail. So, suppose that the context of utterance forces only the weaker 'not both' reading:

(16) If John were *not both* unmarried and a man, then he would be married.

Assuming that (16) is the correct reading of (14) and Simplification is valid *at least* for counterfactuals with this reading, it seems that we can still infer (15) from (16), as long as we take (14) to be true. But (15) is false. Hence, Willer's idea that 'not both' readings are those which license Simplification seems ungrounded.

What about the phenomenon of Suppositional Accomodation? Clearly, in asserting (14) we are not excluding verifiers of "John is not a man" – the antecedent of (15) – from the domain of relevant possibilities. We are simply claiming that *the most plausible way* (according to the actual setup) for John's not being a bachelor is John's being married. There is no contextual shift in moving from (14) to (15): the former does not "rule out" any scenario that is relevant for a non-vacuous evaluation of the latter.

It is far from obvious how this kind of counterexamples to (SDA) could be resisted by appealing to sneaky pragmatic phenomena. The same applies to counterexamples involving determinable antecedents, such as those proposed by Embry. For instance, asserting "if Sue were to take some of these pills, she would get better" does not seem to rule out verifiers for 'Sue takes 25 of these pills' from the relevant possibilities: it is simply *not* the case that any appropriate model that verifies the former must necessarily be assigning the empty set to the latter. Therefore, in Fine's framework, either the counterfactual is false (*contra* intuition) or, if it is true, by (URA) it must entail all its (SDA)-consequences *non-vacuously*. Against this, the evidence suggests that in a large class of cases the failure of (SDA) is a genuine *semantic* feature, that should in fact be captured by the formal framework.[14]

In light of this, I offer an alternative and simpler explanation for why (SDA) does not seem to be generally valid. Notice that, in all of the problematic cases discussed, the intuitive mismatch in truth value between a disjunctive counterfactual and some of its (SDA)-consequences can be easily accounted for by appealing to a certain notion of *plausibility* or *similarity*. As popularized by the classic Stalnaker-Lewis approaches, we typically judge a counterfactual as true if and only if its consequent is true in the closest scenarios at which its antecedent holds. When it comes to counterfactuals with disjunctive antecedents, the most plausible scenarios may often verify only *some* of the disjuncts, but not all of them. This is why we typically judge many disjunctive

14. Ironically, the view that semantic explanations are to be preferred to pragmatic ones is shared by Fine (2023a) himself. In his reply to Bacon (2023), he writes: «There is, of course, no substitute for a detailed consideration of the arguments for or against a pragmatic explanation of a given phenomenon. But let me mention a couple of general considerations that bear on the issue. In the first place, pragmatic, as opposed to semantic, explanations, are cheap, especially when the explanations are not especially systematic in character [...]. I can even imagine a view (call it "semantic monism") according to which all sentences had the very same meaning and the differences between them were entirely attributable to pragmatic factors! For this reason, semantic explanations are, as a general rule, to be preferred» (p. 403).

counterfactuals as true, even though we may not be disposed to accept all of their (SDA)-consequences. Obviously, this undermines the acceptability of (URA), since that assumption forces us to consider *every* verifier of an antecedent in order to evaluate a counterfactual.

Before moving on, I want to conclude this section by considering a final, less general worry concerning (SDA) within Fine's semantics, that has been raised by Bacon (2023). It involves two other principles:

$$A \boxright C, B \boxright C \vDash A \vee B \boxright C \qquad \text{(Disjunction)}$$

$$A \boxright C, B \boxright C \vDash A \wedge B \boxright C \qquad \text{(Weak Strengthening)}$$

While (Disjunction) is generally not considered problematic, (Weak Strengthening) conflicts with some of our intuitive judgements about counterfactuals. Indeed, while the following can definitely be both true:

(17) If I were to drink this hot beverage, I would have a pleasant time.

(18) If I were to ride this roller-coaster, I would have a pleasant time.

It is arguably false that:

(19) If I were to drink this hot beverage and ride this roller-coaster, I would have a pleasant time.

Unfortunately for Fine, it is possible to derive (Weak Strengthening) in many of his counterfactual models, as long as they are of the inclusive kind discussed above:[15]

1. $A \boxright C$ (Assumption)
2. $B \boxright C$ (Assumption)
3. $A \vee B \boxright C$ (Disjunction)
4. $(A \vee B) \wedge (A \vee B) \boxright C$ (Idempotence)
5. $((A \vee B) \wedge A) \vee ((A \vee B) \wedge B) \boxright C$ (Distributivity)
6. $(A \wedge (A \vee B)) \vee (B \wedge (A \vee B)) \boxright C$ (Commutativity)
7. $((A \wedge A) \vee (A \wedge B)) \vee ((B \wedge A) \vee (B \wedge B)) \boxright C$ (Distributivity)
8. $(A \vee (A \wedge B)) \vee ((B \wedge A) \vee B) \boxright C$ (Idempotence)
9. $(A \vee B) \vee ((A \wedge B) \vee (A \wedge B))) \boxright C$ (Associativity), (Commutativity)

15. This is because $A \vee B \dashv\vDash (A \vee B) \wedge (A \vee B)$ (Idempotence) is valid only on the inclusive semantics.

10. $(A \vee B) \vee (A \wedge B) \; \Box\!\!\rightarrow C$ (Idempotence)

11. $\therefore A \wedge B \; \Box\!\!\rightarrow C$ (SDA)

In light of this, Bacon concludes that an appropriate semantics for counterfactuals should invalidate (Disjunction) – even if that amounts to invalidating several other intuitive principles – , while Fine (2023a) is more inclined to reject (Idempotence). But since we have independent reasons to reject (SDA), we can try to resist the inference by embracing a framework that does not assume (URA) and in which both (Disjunction) and (Idempotence) are valid, just like in the Stalnaker-Lewis semantics.

However, as clarified in the previous sections, Fine's semantics has independent advantages over classic similarity-based frameworks. Therefore, what we discussed in the present section represents sufficient justification to motivate a combined approach. So, how should we capture the idea of selecting only a proper subset of an antecedent's content within a truthmaker setting? As we shall see in more detail, Embry (2014) argues that there is no satisfactory way to do so.

6.6 "Failed" Solutions

In the previous section we observed how the validity of (SDA) and the counterintuitive results with determinable antecedents in Fine's framework rely on a crucial assumption: (URA). So, as already pointed out by Embry (2014), the most straightforward way to deal with the problematic cases is simply by rejecting (URA).

Embry tries to achieve this by combining Fine's semantics with the classic similarity-based approaches. He considers two options:

$$w \vDash A \; \Box\!\!\rightarrow C \Leftrightarrow \text{for all closest-to-}w \text{ world } v \text{ in which } t \Vdash A \text{ and } t \mapsto_v u, u \Vdash C \qquad \text{(TC*)}$$

This just means that a world verifies a counterfactual $A \; \Box\!\!\rightarrow C$ if and only if in all closest-to-w worlds v, we have that $A \; \Box\!\!\rightarrow C$ is true at v according to Fine's clause (TC).[16]

Alternatively, one could go for a more Lewis-style clause:

$$\begin{aligned} w \vDash A \; \Box\!\!\rightarrow C \Leftrightarrow \; & \text{there is a world } v \text{ in which } u \Vdash C, t \Vdash A, t \mapsto_v u, \qquad \text{(TC**)} \\ & \text{and there is no world } v^* \text{ closer than } v \text{ to } w \text{ in which } A \wedge \neg C \text{ is true} \end{aligned}$$

Both (TC*) and (TC**) have their own advantages and shortcomings. Closeness among worlds for (TC*) must be captured in a model by means of some world-selection function. In other words, any combined model of this kind will contain some function f such that, for any pair containing a world w and a proposition $[A]^+$, $f(w, A)$ will be the set of the closest-to-w A-world.

16. Embry's original clause refers to a single *closest world*, implicitly assuming a Stalnakerian uniqueness constraint according to which $f(w, A)$ contains *at most* one world. This detail makes no difference for the current purposes.

This should provide a solution to the determinable/disjunctive antecedents problem: f might select worlds containing as parts verifiers only for some (and not all) of the disjuncts of a disjunctive antecedent, or, equivalently, only a *proper* subset of the verifiers for a determinable antecedent. Hence, such an account can deliver the correct evaluations for cases like (11), (12) and (14) without thereby validating (SDA), just like it would happen within the original Stalnaker-Lewis frameworks.

However, Embry is rather quick in rejecting (TC*), by pointing out that it is committed to the controversial Limit Assumption. This says that for all worlds w and all propositions A, there must always be a closest A-world (or a set of closest A-worlds) to w. The standard argument against the Limit Assumption is the one presented by Lewis (1973), which invites us to consider Linus, an actual line that is exactly 1 inch in length:

> In a nearby world Linus is 1 1/2" in length. In a still closer world Linus is 1 1/4" in length. In a still closer world Linus is 1 1/8" in length. Obviously, this can go on *ad infinitum* (Embry 2014, p. 285).

In light of this, Embry concludes that (TC*) is not a viable option, and moves on to the assessment of (TC**).

This second clause represents a rephrasing in truthmaker fashion of another clause due to Lewis (1973), which was developed precisely to avoid the commitment to the Limit Assumption. Unfortunately, Lewis' clause has its own problems, that get reflected upon (TC**). In particular, it invalidates another intuitive principle:

$$A \boxdot\!\!\rightarrow C_1, A \boxdot\!\!\rightarrow C_2, A \boxdot\!\!\rightarrow C_3, ... \vDash A \boxdot\!\!\rightarrow (C_1 \wedge C_2 \wedge C_3 \wedge ...) \quad \text{(Infinitary Conjunction)}$$

In order to avoid this limitation, Fine (2012a) proposed to modify Lewis' original clause for *stranded* worlds only. A set of worlds is said to be stranded with respect to a base world i if and only if the set contains no world that is closest to i. The members of a stranded set are called stranded worlds. We will find stranded sets of worlds in any model that does not obey the Limit Assumption. To avoid the failure of (Infinitary Conjunction), Fine proposed the following clause: $w \vDash A \boxdot\!\!\rightarrow C$ if and only if all of the *stranded* A-worlds are C-worlds.

Embry argues that Fine's move can be adapted to the truthmaker setting by splitting the clause for stranded and non-stranded worlds, as follows:

$$w \vDash A \boxdot\!\!\rightarrow C \Leftrightarrow \text{there is a closest world } v \text{ where } u \Vdash C, t \Vdash A \text{ and } t \mapsto_v u; \quad \text{(N-Stranded)}$$

$$\text{or}$$

$$\text{the } A\text{-worlds are stranded with respect to } w \text{ and all of them contain a state } u \Vdash C \text{ and a state } t \Vdash A, \text{ where } t \mapsto_v u. \quad \text{(Stranded)}$$

However, this leads to yet another problem, since both the original reformulation of Lewis' clause and the split clause proposed by Embry invalidate (Disjunction). The argument within the standard setting runs as follows: consider a set of worlds w_1, w_2, w_3,... getting progressively closer to a base world i. Suppose that A is true only in w_1 and w_2, while B and C are true only in the infinitely many worlds w_2, w_3,... leading up to i.

Since the A-worlds are not stranded, and the closest A-world (w_2) is a C-world, $A \boxdot\!\!\rightarrow C$ is true at i. The B-worlds are stranded, but since all of the B-worlds are C-worlds, $B \boxdot\!\!\rightarrow C$ is also true at i. Now, the $A \vee B$-worlds are also stranded, but not all of them are C-worlds (w_1 is an $A \vee B$-world, but not a C-world). So, $A \vee B \boxdot\!\!\rightarrow C$ will be false at i, invalidating (Disjunction).

In light of this, Embry concludes that there is no satisfactory way to combine Fine's account with similarity-based approaches. However, notice that Embry considers only counterfactual clauses in which the content selection – required to avoid problems with determinable/disjunctive antecedents – is performed at the world-level, either by means of some world-selection function or by employing a Lewis-style system. But adopting a TMS framework allows for an alternative way of performing the kind of selection required. Remember that TMS provides a more fine-grained account of semantic content than the one employed in standard similarity-based approaches. Within a TMS model, nothing forbids to perform the required selection *directly at the state-level*, rather than at the world-level. In other words, it is possible to exploit the resources of truthmaker semantics for picking out the closest *exact verifiers* of a given antecedent, instead of selecting the closest worlds containing such verifiers. Here is where my alternative combined approach comes into play.

6.7 State-Selection Functions to the Rescue

To develop an alternative solution to the problems encountered in the previous sections, I will start from a basic W-space of the kind seen above: $\langle S, S^{\diamond}, \sqsubseteq \rangle$.

However, the models will differ from both those proposed by Fine and those suggested by Embry. A combined counterfactual model $\mathcal{CCM}$: $\langle S, S^{\diamond}, \sqsubseteq, \mapsto, f, v^{+}, v^{-} \rangle$ is the result of equipping a W-space with the usual valuation functions, the transition relation, and a *state*-selection function f.

The valuation functions and the transition relation $\mapsto$ will behave exactly as they did in the basic framework discussed earlier. The only novel implementation is the selection function $f : W \times \wp(S) \rightarrow \wp(S)$, which takes a pair consisting of a world-state and a set of states as input and gives a set of states as output.

I will follow the main intuition of the standard Stalnaker-Lewis approaches, namely, the idea of considering only the closest scenarios for evaluating a counterfactual. However, instead of the closest worlds, I want to consider only the closest exact ways for an antecedent to be true – in other words, the most plausible exact verifiers of the antecedent with respect to a given world. The state-selection function will behave accordingly: given a world-state w and a proposition $[A]^{+}$ (corresponding to a set of states in the model), $f(w, A)$ will be the set of the most plausible states with respect to the world w.[17] What counts as a most plausible or closest state to a given w will be largely determined by contextual factors, just like it happened with the choice of the

17. I will omit the square brackets to denote propositions in the function's notation for an easier reading, but it will be clear from the context that $f(w, A)$ now corresponds to $f(w, [A]^{+})$.

relevant similarity metric across worlds within the original Stalnaker-Lewis accounts. Obviously, introducing the selection function amounts to denying (URA), since $f(w, A)$ may be only a *proper* subset of $[A]^+$.

An obvious condition that will be placed on the selection function, in line with the standard approaches, is the following:

$$f(w, A) \subseteq A \qquad \text{(Success)}$$

Intuitively, given an antecedent A, the states picked by the selection function must belong to the exact verifiers of A. As introduced earlier, I will follow Lewis (1971) in placing two further conditions:

$$f(w, A) \subseteq B \text{ and } f(w, B) \subseteq A \Rightarrow f(w, A) = f(w, B) \qquad \text{(Uniformity)}$$

$$f(w, A \vee B) \subseteq A,\ f(w, A \vee B) \subseteq B, \text{ or } f(w, A \vee B) = f(w, A) \cup f(w, B) \qquad \text{(Nesting)}$$

As we shall see, these will turn out useful to validate all the desired principles. In particular, (Nesting) guarantees the possibility of validating (Disjunction) and invalidating (SDA), but allowing for the restricted validity of (SDA) over some class of models.

The clauses for the basic connectives are the usual ones, while the new clause for counterfactuals will be the following:

$$w \vDash A \mathrel{\square\!\!\rightarrow} C \Leftrightarrow \forall s, t \text{ such that } s \in f(w, A) \text{ and } s \mapsto_w t, t \Vdash C \qquad \text{(CTC)}$$

In words, a world will verify a counterfactual if and only if all outcomes of the most plausible exact verifiers of the antecedent are inexact verifiers of the consequent.[18]

We can illustrate the advantages of this approach by returning briefly to some of the problematic cases discussed earlier:

(7) If Sue were to take some of these pills, she would get better.

(10) If John were not a bachelor, then he would be married.

(11) If John were not a man, then he would be married.

18. An alternative, more elegant way to spell out (CTC) is by making f act directly on the *outcomes* of the verifiers of A (and changing the conditions accordingly), so that $f(A)$ will be the set of the outcomes of the most plausible verifiers of A. This would give us the following clause:

$$w \vDash A \mathrel{\square\!\!\rightarrow} C \Leftrightarrow \forall s(s \in f(w, A) \Rightarrow s \Vdash C)$$

or, even more succinctly:

$$w \vDash A \mathrel{\square\!\!\rightarrow} C \Leftrightarrow f(w, A) \subseteq [C]^+_i$$

where $[C]^+_i$ is the set of the inexact verifiers of C.

First, consider the case of counterfactuals with determinable antecedents. Since we do not require that every verifier of an antecedent A must be considered for the evaluation, we can deliver the expected truth values in all the intended models where the function selects only the most plausible verifiers of A (with respect to the context of evaluation). For instance, in order to deal with Embry's pills counterexample, we may set a minimal and a maximal threshold for the amount of pills relative to (11)'s context of utterance, and make our function select only the states in A that lie between such thresholds. [19] Unlike Fine's original setup, this allows to evaluate (11) correctly. Furthermore, it will not allow to infer any implicit unpalatable (SDA)-consequence, such as "if Sue were to take 25 of these pills, she would get better".

The strategy for cases involving disjunctive antecedents like (14) is similar. Here we can employ the selection function to "isolate" the relevant disjunct in the antecedent with respect to the context of utterance (in this case, the verifiers for "John is married"), and then check what are the outcomes of those states only.

Consider a toy model in which s is the only exact verifier of "John is married" and t is the only exact verifier of "John is not a man". Then, $A = \{s, t, s \sqcup t\}$. Given the context of evaluation, we obviously have that $f(w, A)=\{s\}$. In other words, since John would rather get married than change gender, the most plausible verifier of (14)'s antecedent is the state that verifies "John is married", namely s. By (Inclusion), s is part of any of its outcomes. Then, since s is an exact verifier of C, *a fortiori* any outcome of s will be among the inexact verifiers of C. This amounts to saying that all states in $f(w, A)$ lead to some inexact verifier of C. Therefore, according to (CTC), (14) is true, as expected. However, this time the truth of (14) does not suffice to establish the truth of (15), since $f(w, A)$ does not even contain verifiers of "John is not a man": it is perfectly fine to assume that, in the same intended model, there are some outcomes of t that are not inexact verifiers of C, thereby making (15) false. In other words, by employing the selection function, we are not forced to accept the unwanted consequence of (SDA).

Notice that, in all of the problematic cases, we have a background setup that provides us with information of which counterfactuals must be true and which false. However, some purely formal feature of the original account prevents us from evaluating the counterfactuals in the correct way. What about cases in which even intuition suggests that (SDA) *should* be valid? It is easy to show how the new combined approach is flexible enough to deliver the correct evaluations for them as well. Recall our very first example:

(1) If it were rainy or windy, then I would be cold.

(2) a. If it were rainy, then I would be cold.

b. If it were windy, then I would be cold.

In this case, it seems reasonable to want our function to select all the verifiers of the antecedent, which amounts to considering all those states as equally plausible. Then, in all intended models, $f(w, A) = A$. If (1) is true, then by (CTC) all exact verifiers of "it is

19. A similar strategy can be found in Stalnaker (1980) for dealing with the aforementioned problem with the Limit Assumption. See also Bennett (2003).

windy or rainy" lead to some inexact verifier of "I am cold". Hence, all exact verifiers of "it is windy" must lead to some inexact verifier of "I am cold" (and the same applies *a fortiori* to any selection from those exact verifiers). But this means that all intended models that make (1) true will also make (2a) true. The same obviously holds for the other disjunct, "it is rainy". Therefore, we have that all models that make (1) true will also make (2b) true. In other words, employing the selection function allows us to easily capture the intuitive instances of (SDA) that we want to be valid.

More generally, we will have (SDA) valid in every model where, for any w and any A, $f(w, A) = A$. However, unlike the standard Stalnaker-Lewis framework, such models will not also validate (Strengthening), since (Left-Congruentiality) fails there.[20]

All of this is achieved without sacrificing other intuitive logical principles. Indeed, we can easily show how this approach validates both (Disjunction) and (Infinitary Conjunction). First of all, we can prove:[21]

$$f(w, A \vee B) \subseteq f(w, A) \cup f(w, B) \qquad \text{(Classical } f\text{-Disjunction)}$$

To prove (Disjunction), suppose that $w \vDash A \boxrightarrow C$ and $w \vDash B \boxrightarrow C$. Then, for all $s \in f(w, A)$ and all $s \in f(w, B)$, if $s \mapsto_w t$, $t \Vdash C$. Now suppose that $s \in f(w, A \vee B)$. By (Classical f-Disjunction), we know that $f(w, A \vee B) \subseteq f(w, A) \cup f(w, B)$. So, $s \in f(w, A) \cup f(w, B)$. This means that either $s \in f(w, A)$ or $s \in f(w, B)$. But in either case, we know that if t is an outcome of s at w, then $t \Vdash C$. So, we can conclude that $A \vee B \boxrightarrow C$, as desired.

To prove (Infinitary Conjunction), suppose that $w \vDash A \boxrightarrow C_1$, $w \vDash A \boxrightarrow C_2$, $w \vDash A \boxrightarrow C_3$, ... By (CTC), this means that for all $s \in f(A)$ and t such that $s \mapsto_w t$, $t \Vdash C_1$, $t \Vdash C_2$, $t \Vdash C_3$ .... We want to show that $A \boxrightarrow (C_1 \wedge C_2 \wedge C_3 \wedge \ldots)$ is true, i.e. we have to show that if $s \in f(w, A)$ and $s \mapsto_w t$, then $t \Vdash C_1 \wedge C_2 \wedge C_3 \wedge \ldots$. So, pick an arbitrary $s \in f(w, A)$ and t such that $s \mapsto_w t$. Our previous observation yields that $t \Vdash C_1, t \Vdash C_2, t \Vdash C_3 \ldots$. So, there will be a $u_1 \sqsubseteq t$ such that $u_1 \Vdash C_1$, a $u_2 \sqsubseteq t$ such that $u_2 \Vdash C_2$, and so on. Now consider $u = u_1 \sqcup u_2 \sqcup \ldots$. First, observe that $u \sqsubseteq t$ since fusions of parts are parts. And second, observe that $u \Vdash C_1 \wedge C_2 \wedge C_3 \wedge \ldots$ by the clause for conjunction. It follows that $t \Vdash C_1 \wedge C_2 \wedge C_3 \wedge \ldots$. Hence, $w \vDash A \boxrightarrow (C_1 \wedge C_2 \wedge C_3 \wedge \ldots)$.

What about the infamous Limit Assumption? The first thing to notice here is that, since the state-selection function does not operate on worlds, strictly speaking there is no commitment to what is usually meant under the "Limit Assumption" label. Nevertheless, one might retort that the account is still committed to a state-theoretic version of the Limit Assumption: after all, we are assuming that for any pair $\langle w, A \rangle$ there is always a set of most plausible states. This means that one might rephrase Lewis' line argument in a state-theoretic fashion in order to reject any semantics relying on a state-selection function.

20. Notice, however, that in all inclusive models where (SDA) is generally valid, (Weak Strengthening) can still be derived.

21. Given (Nesting), either $f(w, A \vee B) \subseteq A$, or $f(w, A \vee B) \subseteq B$, or $f(w, A \vee B) = f(w, A) \cup f(w, B)$. *Case 1*: Assume that $f(w, A \vee B) = f(w, A) \cup f(w, B)$. Then, clearly, $f(w, A \vee B) \subseteq f(w, A) \cup f(w, B)$. *Case 2*: Assume that $f(w, A \vee B) \subseteq A$. By (Success), $f(w, A) \subseteq A$. Since $A \subseteq A \vee B$, $f(w, A) \subseteq A \vee B$. Then, by (Uniformity), $f(w, A \vee B) = f(w, A)$. Therefore, $f(w, A \vee B) \subseteq f(w, A) \cup f(w, B)$. *Case 3*: Assume that $f(w, A \vee B) \subseteq B$. The proof is analogous to that of Case 2. Therefore, $f(w, A \vee B) \subseteq f(w, A) \cup f(w, B)$.

However, on a more careful scrutiny, this may not constitute a serious worry for the present account. Notice that there is nothing specific about this proposal that prevents us from deflecting worries related to the Limit Assumption in ways similar to those already extensively discussed in the relevant literature (Pollock (1976), Stalnaker (1980), Warmbrōd (1982), Bennett (2003) just to mention a few examples). [22] Indeed, state-theoretic adaptations of the existing replies to Lewis' original argument could be easily devised to resist Embry's point (just like Lewis' original argument can be adapted in state-theoretic jargon). Furthermore, I want to stress that the current proposal is primarily concerned with the interplay between our intuitive semantic judgments and counterfactual *logic*: the Limit Assumption (or any state-theoretic variant of it) does not make any difference with respect to the axioms upon which a system for counterfactual logic relies. While the debate on the Limit Assumption and its consequences is ongoing (see Kaufman (2017) for some recent developments), the current proposal is not meant to take a stance about its philosophical implications.

6.8 Exact Truthmakers for Counterfactuals and Counterpossibles

I am now in the position to briefly address two last issues of special interest for the present work. Both of them concern the possibility of refining the TMS for counterfactuals discussed in the previous sections even further, in two distinct yet surprisingly intertwined ways: (i) the development of an *exact* clause for counterfactuals, i.e. a fully explicit *truthmaker* condition that allows us to assign an hyperintensional content (in the form of a set of verifiers) to them; and (ii) the adjustments required to provide a non-vacuist treatment of *counterpossibles* within this same setting. Since dealing with the former issue will turn out to be the very last piece needed to complete the puzzle, I will begin from the latter.

The most detailed discussion of how TMS may deal with counterpossibles to date is due to Sendłak (2024), who highlights the relevant features of Fine's semantics that make it vacuist (yet hyperintensional). Since such features are shared by the refined accounts discussed earlier, for this first part of the section I will stick to Sendłak's discussion of the original clause (TC), but all of the following obviously applies to (CTC) as well.

22. Embry (2014) does not mention the existence of these solutions, so he is very quick in accepting Lewis' counterexample as a fatal flaw for any semantics relying on selection functions. However, in light of the existing replies to Lewis, there should be plenty of room for defending Embry's first combined proposal (TC*) as well on the same grounds, but I will not pursue this task here. I will instead suggest in sec. 6.8 how the state-theoretic variant of the selection function approach underlying (CTC) might have some relative advantages over (TC*), since it does not have to rely on worlds at all: this shall make it easier to exactify the clause. On the other hand, (TC*) might have an advantage if we want a more robust formal account of closeness, since this may be captured in terms of overlap among worlds. Here is a rough sketch: a world v^* is closer to w than a world v if and only if $v^* \sqcap w$ is a "bigger" state than $v \sqcap w$ (where $v \sqcap w = \bigsqcup\{s \mid s \sqsubseteq v \wedge s \sqsubseteq w\}$).

First, recall two of the conditions placed on the transition relation:

$$\text{If } s \mapsto_w t \text{ then } s \sqsubseteq t \qquad \text{(Inclusion)}$$

$$s \mapsto_w t \text{ only if } t \text{ is a world-state} \qquad \text{(Completeness)}$$

A state w is a world-state if and only if $w \in S^\diamond$ and, for every s, either $s \sqsubseteq w$ or s is incompatible with w (i.e., $s \sqcup w \notin S^\diamond$). A direct consequence of this definition, (Inclusion), and (Completeness) is:

$$\text{If } s \mapsto_w t \text{ then } s \sqcup t \in S^\diamond \qquad \text{(Consistency)}$$

To see this, we just need to notice that every world-state belongs to $S^\diamond$ by definition, so since by (Completeness) all outcomes are world-states, and by (Inclusion) any state is part of any of its outcomes, the fusion of a state with one of its outcomes must belong to $S^\diamond$. In other words, there are no impossible outcomes (and no impossible world-states).

Now, consider again the (false) counterpossible:

(1) If diamond was not covalently bonded, it would be a better electrical conductor.

Suppose that a_1 is an arbitrary exact verifier of the (impossible) antecedent A of (1). Clearly, $a_1 \notin S^\diamond$. So, for any state s, $a_1 \sqcup s \notin S^\diamond$. Now suppose that, for some state t, $a_1 \mapsto_w t$. By (Consistency), $a_1 \sqcup t \in S^\diamond$. Contradiction.

In light of this, for no state s, $a_1 \mapsto_w s$. In other words, a_1 has no outcomes. But this means that (TC) is vacuously satisfied by (1): since there are no outcomes of a_1, it is never the case that $a_1 \Vdash A$ *and* $a_1 \mapsto_w t$, for some state t. This makes the antecedent of the conditional in the right-half of (TC) false, and thereby the whole conditional true. In other words, according to (TC) a counterfactual is true whenever its antecedent is impossible, which means that Fine's TMS account of counterfactuals is still vacuist.

The same obviously applies to the refined semantics underlying (CTC): if a_1 has no outcomes, then it is never the case that $a_1 \in f(w, A)$ *and* $a_1 \mapsto_w t$, for some state t. Equivalently, if we use f to select the *outcomes* of the most plausible exact verifiers of A, then $f(A)$ will always be the empty set, whenever A is impossible. In short, even the refined semantics presented in chap. 6 turns out to be vacuist.

As pointed out by Sendłak (2024), if we hope to provide non-vacuous truth-conditions for counterpossibles within this setting, we will need impossible outcomes. This amounts to rejecting at least (Consistency), and since this condition ultimately depends on (Inclusion) and (Completeness), that means that we should drop at least one of them.[23] As seen earlier, the former guarantees the validity of (Identity), while the latter enables to provide truth-conditions for counterfactuals with embedded counterfactuals in the consequent. Clearly, we cannot give up on (Identity), so (Completeness) has to go. Interestingly, there seems to be a way to do this without any real trade-off with respect to the expressive power of the semantics and the resulting logic: as already noticed in Fine (2012b), we can simply change the counterfactual clause

23. We may also choose to keep both (Inclusion) and (Completeness) and change the definition of world-states in order to allow for them to be *impossible* states, but I will not explore this option here.

(TC) by using *loose* verification in place of inexact verification. This kind of verification is defined as follows:

$$s \vDash A \Leftrightarrow \forall t(t \dashv\Vdash A \Rightarrow s \sqcup t \notin S^{\diamond}) \qquad \text{(Loose Verification)}$$

A state s is a loose verifier of A if and only if it is incompatible with all exact falsifiers of A. Obviously, loose verification is equivalent to classic truth-at-a-world. (TC) can then be modified to obtain:

$$w \vDash A \boxdot\!\!\rightarrow C \Leftrightarrow \forall s, t \text{ such that } s \Vdash A \text{ and } s \mapsto_w t, t \vDash C \qquad \text{(TCL)}$$

This simple change allows us to evaluate counterfactuals with embedded counterfactuals in the consequent *without* assuming that outcomes must be complete world-states. That is because, in order to evaluate a counterfactual at a world, we need to check whether its consequent is loosely verified by the outcomes of all the exact verifiers of the antecedent; in other words, we just need to check which worlds containing those outcomes verify the consequent. Thus, to evaluate a sentence of the form $A \boxdot\!\!\rightarrow (B \boxdot\!\!\rightarrow C)$ we just need to know the worlds at which $B \boxdot\!\!\rightarrow C$ is true: this means that once the semantics settles which first-degree counterfactuals are true at which worlds, it settles also what their loose verifiers are, and thereby the truth-conditions for all counterfactuals in which they are embedded as consequents.

Is this enough to make the semantics non-vacuist? Unfortunately, as Sendłak highlights, the answer is still no. We can illustrate this by considering that, by the definition of loose verification:

$$s \nvDash A \Leftrightarrow \exists t(t \dashv\Vdash A \wedge s \sqcup t \in S^{\diamond}) \qquad \text{(LV Failure)}$$

That is to say, a state s *fails* to loosely verify A if and only if there is some exact falsifier of A that is compatible with s.

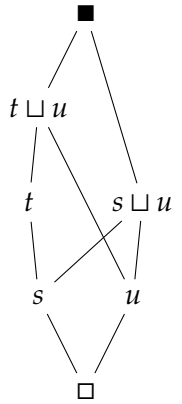
Now, take our counterpossible (1) again. Under the new clause (TCL), (1) will be false if an outcome of a_1 fails to loosely verify its consequent. Again, since $a_1 \notin S^{\diamond}$, if $a_1 \mapsto_w t$, by (Inclusion) $t \notin S^{\diamond}$. Hence, for any s, $s \sqcup t \notin S^{\diamond}$. But this means that no outcome of a_1 satisfies (LV Failure). In other words, an outcome of a_1 cannot fail to loosely verify anything: an impossible state will loosely verify every sentence.

Once again, the resulting semantics is vacuist: whenever the antecedent is impossible, a counterfactual will always turn out as true. And once again, this would be the case even if we operated the same modification to the refined clause (CTC).

Summing up, dropping (Completeness) does not bring us far if we also change the counterfactual clauses. What if, instead, we stick to the original (TC) with *inexact* verification? To see what happens with counterpossibles, in this case we need to consider what it takes for a state to fail to inexactly verify a sentence:

$$s \nVdash A \Leftrightarrow \neg\exists t(t \sqsubseteq s \wedge t \Vdash A) \qquad \text{(IV Failure)}$$

This says that a state s fails to inexactly verify a sentence A if and only if there is no part of s that is an exact verifier of A. Now, the question arises: can impossible states satisfy (IV Failure)? The answer is clearly yes. To see this, consider this simple model:



- $S = \{\Box, s, t, u, s \sqcup u, t \sqcup u, \blacksquare\}$[24]
- $S^{\diamond} = \{\Box, u\}$
- $W = \{u\}$
- $[A]^{+}_{e} = \{s\}$
- $[C]^{+}_{e} = \{t \sqcup u\}$; $[C]^{+}_{i} = \{t \sqcup u, \blacksquare\}$
- $s \mapsto_{u} t$ (t is the only outcome of s).

We can easily observe that, given that $s \sqsubseteq t$, the model satisfies (Inclusion), and no parts of t are exact verifiers of C. Furthermore, $t \notin S^{\diamond}$. This means that an impossible state satisfies (IV Failure), and that the counterpossible $A \mathrel{\Box\!\!\rightarrow} C$ turns out false according to (TC).

Apparently, the easiest route to non-vacuism is via the standard clause without (Completeness). However, as noticed above, dropping (Completeness) while sticking to inexact verification in the counterfactual clause amounts to a serious loss of expressive power: we cannot evaluate counterfactuals with embedded counterfactuals anymore. That is because, in order to evaluate $A \mathrel{\Box\!\!\rightarrow} (B \mathrel{\Box\!\!\rightarrow} C)$ according to (TC), we need to know what counts as an inexact verifier of $B \mathrel{\Box\!\!\rightarrow} C$. But (TC) tells us only when an entire world verifies a counterfactual, and since outcomes are not assumed to be world-states anymore, knowing which world-states verify $B \mathrel{\Box\!\!\rightarrow} C$ is not enough to evaluate $A \mathrel{\Box\!\!\rightarrow} (B \mathrel{\Box\!\!\rightarrow} C)$: among the outcomes of A there might be states that are not worlds. Obviously, things would not look any better if we adopted (CTC) instead, since just like (TC) it only provides worldly truth-conditions.

It seems that, in order to recapture the expressive power of the original semantics, we need to identify inexact verifiers for counterfactuals. A straightforward way to do so is by identifying *exact* verifiers for counterfactuals, and defining inexact verifiers in terms of them, as it is standard to do in TMS. Notice that this would increase the

24. $\Box$ and $\blacksquare$ are respectively the *null* and the *full* state: for any state $s \in S$, $\Box \sqsubseteq s$ and $s \sqsubseteq \blacksquare$.

expressive power of the semantics even further, since it would allow us to evaluate all sorts of nested counterfactuals, including those with embeddings in antecedent position. Luckily, a proposal by Fine (2020), subsequently refined in Fine (2023b), can be adapted to our needs in a fairly straightforward way.

We can start by considering transitions between states as states in their own right: we will follow Fine in calling them *paths*. Instead of centering the transitions on worlds, Fine proposes to center them on arbitrary states. So, $t \mapsto_s u$ will be used as a shorthand for 's is a path from t to u', meaning that s is an underlying condition in the world that enables u to be a possible outcome of t. Since a sentence is usually verified by multiple states, there will usually be many such paths for each pair of sentences. We can then think of an exact verifier for a counterfactual $A \boxright C$ as a fusion of all paths from exact verifiers of A to exact verifiers of C.

The way to make this precise is by introducing a function $g : S \to S$ that pairs A-states with C-states, such that $g(t) \Vdash C$ whenever $t \Vdash A$. Then, $t \mapsto_s g(t)$ says that s is a path from a specific A-state to a specific C-state. An exact verifier of $A \boxright C$ will then be the fusion of all such path-states:[25]

$$s \Vdash A \boxright C \Leftrightarrow \text{ for some } g \text{ such that } \forall t \in [A]^+, g(t) \Vdash C \quad \text{and } s = \bigsqcup \{u \mid t \mapsto_u g(t)\} \qquad (\text{Exact } \boxright^+)$$

Since there may be many such g-functions, s will not in general be unique. Implementing this idea in the original setup allows us to evaluate (non-vacuously) all sorts of counterfactual constructions, since we can now define an inexact verifier for $A \boxright C$ as a state t such that $s \Vdash A \boxright C$ and $s \sqsubseteq t$. Likewise, a world w will verify a counterfactual whenever $s \sqsubseteq w$.

What about the combined approach discussed in the previous section? Clearly, while ($\boxright^+$) allows for evaluating more counterfactuals, it does not encode the kind of information that we need when we consider failures of the exact counterpart of (SDA):

$$A \vee B \boxright C \vDash_e A \boxright C, B \boxright C \qquad (\text{Exact SDA})$$

To find a solution, it seems reasonable to restrict our focus on how the exact clause should be modified, regardless of how it relates to (CTC). This is because, as just seen, an exact clause allows us to evaluate embedded counterfactuals even if we choose to drop the world-talk altogether. This brings us to the last question that I shall briefly address: how can we capture the idea of state-selection functions within an *exact* clause for counterfactuals?

Recall that the original idea to avoid problems with determinable antecedents and (SDA) was to introduce a notion of closest states with respect to worlds. In particular, (CTC) appealed to the A-states closest to *the world of evaluation*, which was also the state

25. Obviously, here we can replace the variables A and C of $\mathcal{L}_{\neg,\wedge,\vee,\boxright}$ with variables ranging over all well-formed formulas (including counterfactual formulas with embedded counterfactuals), since an exact clause is not restricted to first-degree counterfactuals only. However, I will stick to the notation employed in the rest of the chapter for the ease of exposition.

around which the transition relation was centered. However, it is not obvious how we may appeal to worlds in a combined exact clause: counterfactuals are not evaluated (only) at worlds anymore, and the transition relation is now centered around specific states, which may be part of different worlds. Furthermore, merely relativizing the selection to the state of evaluation itself would be philosophically poor: it is not clear what the most plausible verifiers of A with respect to *a fusion of paths from verifiers of A to verifiers of C* are supposed to be, since paths alone (and fusions thereof) do not provide a proxy context against which we may form non-arbitrary plausibility or similarity judgments.

Apparently, the easiest solution from a purely formal standpoint is simply to drop the idea that the most plausible verifiers of a given sentence should be relativized to some world or state parameter. After all, we may think that the only thing that should affect what counts as a closest A-state, for some A, according to a model, is *the model itself*. If we take this idea seriously, we can think of a model $CC\mathcal{M}^+$ as a usual $CC\mathcal{M}$-structure equipped with a simpler selection function $f_1 : \wp(S) \to \wp(S)$. We may then say that, in each model $CC\mathcal{M}^+$, $f_1(A)$ will simply be the set of closest exact verifiers of A according to the model (which would here represent the context against which we assess similarity and plausibility facts).

This allows for a straightforward rephrasing of Fine's exact clause as follows:

$$s \Vdash A \boxright C \Leftrightarrow \text{ for some } g \text{ such that } \forall t \in f_1(A), g(t) \Vdash C \qquad (\text{Exact } \boxright^+_{f_1})$$
$$\text{and } s = \bigsqcup \{u \mid t \mapsto_u g(t)\}$$

This clause combines all the desired features discussed throughout the chapter: a non-vacuist, hyperintensional semantics that delivers the intuitive verdicts when the antecedents are determinable, does not generally validate (SDA), and allows for evaluating all sorts of nested counterfactuals.[26] Furthermore, notice that a frame for a model in which ($\boxright^+_{f_1}$) can be employed does not even need to be a W-space: we can dispense with world-states (and even possible states), and use any ordinary state space as a frame. Clearly, the details of the resulting logic will ultimately depend on the conditions placed on f and on paths, just like in the original setting, which may be independently motivated.

In conclusion, I would like to suggest two alternative routes which may be worth exploring for those who take relativization to a world-parameter essential for making sense of the notion of closest states.

The first is the following: instead of using a W-space as a frame, we can enrich it by turning it into an A-space: $\langle S, S^\diamond, @, \sqsubseteq \rangle$. An A-space is just a W-space where we single out a world-state $@ \in S^\diamond$ to represent the *actual* world, which can be thought of as the fusion of all the actual obtaining states.

26. Notice that ($\boxright^+_{f_1}$) invalidates (Exact SDA). This is because, for instance, s may verify $A \vee B \boxright C$ without verifying $B \boxright C$: it is sufficient that $f(B)$ contains some state that does not have a verifier of C as an outcome.

Now, we can define a model $\mathcal{CCM}^{@} : \langle S, S^{\diamond}, @, \sqsubseteq, \mapsto, f_2, v^{+}, v^{-} \rangle$, and relativize the selection function to the actual world by taking $f_2 : \{@\} \times \wp(S) \to \wp(S)$. This would amount to having $f_2(@, A)$ as the set of the closest A-states with respect to how things are in the actual world. The corresponding clause for counterfactuals would then be:

$$s \Vdash A \boxdot\!\!\rightarrow C \Leftrightarrow \text{ for some } g \text{ such that } \forall t \in f_2(@, A), g(t) \Vdash C \qquad (\text{Exact } \boxdot\!\!\rightarrow^{+}_{f_2})$$
$$\text{and } s = \bigsqcup \{u \mid t \mapsto_u g(t)\}$$

Under this clause, what counts as a verifier for a counterfactual would always partly depend on how things are at the actual world. This may seem natural if we believe that the actual world must always be the privileged standpoint from which we ascertain such plausibility facts. However, one might object that this does not fully comply with our intuitions: after all, what states count as more or less plausible may vary across worlds. The problem can perhaps be mitigated by allowing @ to change across models, in order to allow them to represent different ways in which plausibility facts are settled.

A more natural response to this issue is to take the second alternative route, which is the closest in spirit to (CTC). Specifically, we may choose to relativize to worlds not only *selections* of states, but also states themselves. More precisely, we can reintroduce the original $f : W \times \wp(S) \to \wp(S)$ and implement it in the following world-relativized clause:

$$s \Vdash_w A \boxdot\!\!\rightarrow C \Leftrightarrow \text{ for some } g \text{ such that } \forall t \in f(w, A), g(t) \Vdash C \qquad (\text{Exact } \boxdot\!\!\rightarrow^{+}_{f,w})$$
$$\text{and } s = \bigsqcup \{u \mid t \mapsto_u g(t)\}$$

This says that a state s verifies a counterfactual at a world w whenever s is the fusion of all path-states from the most plausible verifiers of A-states according to w to C-states.

A similar kind of world-relativization of truthmakers has already been explored by Korbmacher (2016) and Rosella (2019), who employed it to developed a truthmaker semantics for alethic modal operators. Hence, there seems to be independent reasons for wanting a notion of truthmaking-at-a-world.

This approach also allows us to define an arbitrary exact verifier of a counterfactual as follows:

$$s \Vdash A \boxdot\!\!\rightarrow C \Leftrightarrow \exists g \forall w \text{ such that } s \sqsubseteq w \text{ and } \forall t \in f(w, A), \qquad (\text{Exact } \boxdot\!\!\rightarrow^{+}_{f})$$
$$g(t) \Vdash C \text{ and } s = \bigsqcup \{u \mid t \mapsto_u g(t)\}$$

Furthermore, it makes $A \boxdot\!\!\rightarrow C$ true at a world w according to (CTC) whenever s is a truthmaker of $A \boxdot\!\!\rightarrow C$ at w according to $(\boxdot\!\!\rightarrow^{+}_{f,w})$, given that s will be part of w.

A deeper discussion of the philosophical merits, shortcomings and motivations behind all of these proposals lies outside the scope of this work, alongside a more in-depth exploration of their logical features. The present chapter aimed primarily at discussing some of the more general formal issues surrounding counterfactuals in TMS: what has been presented so far should suffice as groundwork for a more adequate, data-driven logic of counterfactuals.

6.9 Conclusion

In this chapter I delved into the complexities surrounding (SDA) in counterfactual logic, particularly within Fine's TMS. After an introduction of the principle and its behavior in some interesting intensional settings, I explored existing challenges to Fine's framework concerning determinable antecedents and extended them to disjunctive antecedents, in order to provide new evidence against the unrestricted validity of (SDA). The new evidence allowed me to reject the existing pragmatic solutions to the problems with disjunctive antecedents, and led me to treating failures of (SDA) as genuine semantic features of a large class of counterfactuals.

I argued that the best way to account for this lies in the rejection of the (URA) principle employed by Fine, through the combination of TMS with the similarity-based approaches developed by Stalnaker and Lewis. While some attempts to reconcile Fine's semantics with such approaches have already been made by Embry, they too encounter challenges. In response to these, I introduced a novel approach that combines TMS with state-selection functions. By allowing for a more fine-grained selection of antecedent content, the proposed framework addresses the shortcomings of previous solutions while preserving many core principles of counterfactual reasoning. Furthermore, it allows for the restricted validity of (SDA) over a class of models, without thereby validating other unpalatable inferences: something that the standard approaches could not achieve.

Finally, I sketched how the basic combined framework may be refined to deal with a broader class of counterfactuals: those embedding other counterfactuals in antecedent or consequent position; and those with impossible antecedents, namely, counterpossibles.

The resulting options should provide a solid starting point for further developments, among which the most obvious would probably be the development of an exact *falsemaker* clause for counterfactuals, which is needed to make the account fully compositional (with respect to a bilateral semantics). This and further related issues shall be addressed in future work.

Bibliography

Alonso-Ovalle, Luis. 2009. "Counterfactuals, Correlatives, and Disjunction." *Linguistics and Philosophy* 32(2):207–244.

Anderson, Anthony C. 2009. "The Lesson of Kaplan's Paradox about Possible World Semantics." In *The Philosophy of David Kaplan*, edited by J. Almog and P. Leonardi, 85–92. Oxford University Press.

Bacon, A., J. Hawthorne, and G. Uzquiano. 2016. "Higher-Order Free Logic and the Prior-Kaplan Paradox." *Canadian Journal of Philosophy* 46 (4-5): 493–541.

Bacon, Andrew. 2023. "Counterfactuals, Infinity and Paradox." In *Kit Fine on Truthmakers, Relevance and Non-classical Logic*, edited by F.L.G. Faroldi and F. Van De Putte, 349–388. Springer Cham.

Baker, Alan. 2024. "Counterpossibles in Mathematical Practice: The Case of Spoof Perfect Numbers." In *Handbook of the History and Philosophy of Mathematical Practice*, edited by B. Sriraman, 2261–2287. Springer.

Båve, Arvid. 2019. "Acts and alternative analyses." *The Journal of Philosophy* 116 (4): 181–205.

Bennett, Johnathan. 2003. *A Philosophical Guide to Conditionals.* Oxford: Oxford University Press. https://doi.org/10.1093/0199258872.001.0001.

Bernstein, Sara. 2016. "Omission Impossible." *Philosophical Studies* 173:2575–2589.

Berto, Francesco. 2022. *Topics of Thought– The Logic of Knowledge, Belief and Imagination.* Oxford University Press.

———. 2024. "Hyperintensionality and Overfitting." *Synthese* 203:117.

Berto, Francesco, R. French, G. Priest, and D. Ripley. 2018. "Williamson on Counterpossibles." *Journal of Philosophical Logic* 47:693–713.

Berto, Francesco, and Mark Jago. 2019. *Impossible worlds.* Oxford University Press.

Berto, Francesco, and Daniel Nolan. 2023. "Hyperintensionality." In *The Stanford Encyclopedia of Philosophy*, Winter 2023, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University.

Berto, Francesco, and Matteo Plebani. 2015. *Ontology and Metaontology – a contemporary guide.* London: Bloomsbury Academic.

Bonaguidi, Maria Beatrice. 2022. "Symmetry, locality and hyperintensionality." *Proceedings of the ESSLLI 2022 Student Session.*

Boolos, George. 1971. "The iterative conception of set." *The Journal of Philosophy* 68 (8): 215–231.

Briggs, Rachael. 2012. "Interventionist counterfactuals." *Philosophical studies* 160:139–166.

Brogaard, Berit, and Joe Salerno. 2013. "Remarks on Counterpossibles." *Synthese* 190:639–660.

Bueno, O., C. Menzel, and E. N. Zalta. 2014. "Worlds and Propositions Set Free." *Erkenntnis* 79:797–820.

Burge, Tyler. 1979a. "Individualism and the mental." *Midwest Studies in Philosophy* 4 (1): 73–122.

———. 1979b. "Sinning against Frege." *Philosophical Review* 88 (3): 398–432.

Burgess, John P. 1981. "Quick Completeness Proofs for Some Logics of Conditionals." *Notre Dame Journal of Formal Logic* 22 (1): 76–84.

Byrne, Darragh, and Naomi Thompson. 2019. "On hyperintensional metaphysics." In *Maurinian truths – essays in honour of Anna-Sofia Maurin on her 50th birthday*, edited by H. T. Wahlberg and R. Stenwall, 151–158. Lund: Lund University.

Carnap, Rudolf. 1947. *Meaning and Necessity.* Chicago: University of Chicago Press.

Chalmers, David. 2008. "Two-dimensional semantics." In *The Oxford Handbook to the Philosophy of Language*, edited by E. LePore and B. C. Smith, 574–606. Oxford University Press.

Chellas, Brian F. 1975. "Basic conditional logic." *Journal of philosophical logic*, 133–153.

Chihara, Charles S. 1998. *The worlds of possibility: modal realism and the semantics of modal logic.* New York: Oxford University Press.

Ciardelli, Ivano, Linmin Zhang, and Lucas Champollion. 2018. "Two switches in the theory of counterfactuals: A study of truth conditionality and minimal change." *Linguistics and Philosophy* 41 (6): 577–621.

Cresswell, Max J. 1975. "Hyperintensional logic." *Studia Logica* 34 (1): 25–38.

Davies, Martin. 1981. *Meaning, Quantification, Necessity: Themes in Philosophical Logic.* London: Routledge / Kegan Paul.

Deutsch, Harry. 2014. "Resolution of some paradoxes of propositions." *Analysis* 74 (1): 26–34.

Divers, John. 2006. "Possible-Worlds Semantics Without Possible Worlds: The Agnostic Approach." *Mind* 115 (458): 187–226.

Dorr, Cian. 2014. "Transparency and the Context-Sensitivity of Attitude Reports." In *Empty Representations: Reference and Non-Existence,* edited by M. García-Carpintero and G. Martí, 25–66. Oxford University Press.

Dosanjh, Ranpal. 2021. "Token-Distinctness and the Disjunctive Strategy." *Erkenntnis* 86(3):715–732.

Dunaway, Billy. 2013. "Modal Quantification without Worlds." In *Oxford Studies in Metaphysics, Volume 8,* edited by K. Bennett and Zimmerman D. W., 151–186. Oxford University Press.

Ellis, Brian, Frank Jackson, and Robert Pargetter. 1977. "An objection to possible-world semantics for counterfactual logics." *Journal of Philosophical Logic,* 355–357.

Embry, Brian. 2014. "Counterfactuals Without Possible Worlds? A Difficulty for Fine's Exact Semantics for Counterfactuals." *The Journal of Philosophy* 111 (5): 276–287.

Fine, Kit. 1975. "Vagueness, truth and logic." *Synthese* 30 (3/4): 265–300.

———. 2012a. "A difficulty for the possible worlds analysis of counterfactuals." *Synthese* 189:29–57.

———. 2012b. "Counterfactuals without possible worlds." *The Journal of Philosophy* 109 (3): 221–246.

———. 2017a. "A Theory of Truthmaker Content I: Conjunction, Disjunction and Negation." *Journal of Philosophical Logic* 46:625–674.

———. 2017b. "Truthmaker Semantics." In *A Companion to the Philosophy of Language,* edited by B. Hale, C. Wright, and A. Miller, 556–577. Wiley-Blackwell. https://doi.org/10.1002/9781118972090.ch22.

———. 2020. "Yablo on subject matter." *Philosophical Studies* 177 (1): 129–171.

———. 2023a. "'Defense of a Truthmaker Approach to Counterfactuals': Response to Andrew Bacon's 'Counterfactuals, Infinity and Paradox'." In *Kit Fine on Truthmakers, Relevance and Non-classical Logic,* edited by F.L.G. Faroldi and F. Van De Putte, 389–406. Springer Cham.

———. 2023b. "'Forms of Conditionality': Response to 'Truth-Maker Semantics for Some Substructural Logics' by Ondrej Majer, Vít Punˇcochárˇ and Igor Sedlár." In *Kit Fine on Truthmakers, Relevance and Non-classical Logic,* edited by F.L.G. Faroldi and F. Van De Putte, 223–230. Springer Cham.

Florio, Salvatore, and Øystein Linnebo. 2021. *The Many and the One: A Philosophical Study of Plural Logic.* Oxford: Oxford University Press.

Frege, Gottlob. 1892. "Sense and Reference." *Philosophical Review* 57 (1948): 209–230.

French, Steven. 1989. "Identity and Individuality in Classical and Quantum Physics." *Australasian Journal of Philosophy* 67:432–446.

Fritz, Peter. 2024. *Propositional Quantifiers.* Cambridge University Press.

Gillett, Carl. 2016. *Reduction and Emergence in Science and Philosophy.* Cambridge University Press.

Graeme, Forbes. 2020. "Intensional Transitive Verbs." In *The Stanford Encyclopedia of Philosophy,* Winter 2020, edited by Edward N. Zalta.

Hendry, Robin F. 2009. "Ontological reduction and molecular structure." *Studies in History and Philosophy of Modern Physics* 41:183–191.

———. 2017. "Prospects for Strong Emergence in Chemistry." In *Philosophical and Scientific Perspectives on Downward Causation,* edited by M. P. Paoletti and Orilia F., 146–163. Routledge.

———. 2023. "Structure, essence and existence in chemistry." *Ratio* 36 (4): 274–288.

Hewitt, Simon T. 2012. "Modalising Plurals." *Journal of Philosophical Logic* 41 (5): 853–875.

Hicks, Michael T. 2022. "Counterparts and Counterpossibles: Impossibility without Impossible Worlds." *The Journal of Philosophy* 119:542–574.

Hofweber, Thomas. 2022. "The case against higher-order metaphysics." *Metaphyisics* 5 (1): 29–50.

Jago, Mark. 2015. "Hyperintensional propositions." *Synthese* 192 (3): 585–601.

———. 2022. "Truthmaker Accounts of Propositions." In *The Routledge Handbook of Propositions,* edited by A. Russell Murray and C. Tillman, 231–244. Taylor & Francis.

Jenkins, Carrie S., and Daniel Nolan. 2012. "Disposition Impossible." *Noûs* 46:732–753.

Jenny, Matthias. 2018. "Counterpossibles in science: the case of relative computability." *Noûs* 52 (3): 530–560.

Kahle, Reinhard. 2011. "Modalities Without Worlds." In *The Realism-Antirealism Debate in the Age of Alternative Logics,* edited by S. Rahman, G. Primiero, and M. Marion, 101–118. Springer Dordrecht.

Kaplan, David. 1995. "A Problem in Possible Worlds Semantics." In *Modality, morality, and belief: essays in honor of Ruth Barcan Marcus,* edited by W. Sinnot-Armstrong, D. Raffman, and Asher N., 41–52. Cambridge University Press.

Kaufman, Stefan. 2017. "The Limit Assumption." *Semantics and Pragmatics* 10(18):1–29.

Kim, Dongwoo. 2024. "Exact Truthmaker Semantics for Modal Logics." *Journal of Philosophical Logic* 53 (3): 789–829.

King, Jeffrey. 1995. "Structured propositions and complex predicates." *Noûs* 29 (4): 516–535.

———. 1996. "Structured propostions and sentence structure." *Journal of philosophical logic* 25:495–521.

———. 2009. *The nature and structure of content.* Oxford University Press.

Kocurec, Alexander W. 2021. "Counterpossibles." *Philosophy Compass* 16 (11).

Korbmacher, Johannes. 2016. *Exact Truthmaker Semantics for Propositional Modal Logic(s).* Groningen: Unpublished presentation.

Kripke, Saul. 1980. *Naming and Necessity.* Harvard University Press.

———. 2011. *Philosophical Troubles - Collected Papers Volume I.* Oxford: Oxford University Press.

Lassiter, Daniel. 2018. "Complex sentential operators refute unrestricted Simplification of Disjunctive Antecedents." *Semantics and Pragmatics* 11:3–9.

Leitgeb, Hannes. 2018. "HYPE: a system of hyperintensional logic (with an application to semantic paradoxes)." *Journal of Philosophical Logic* 48:305–405.

Lenta, Giorgio. 2023. "Sources of hyperintensionality." *Theoria* 89 (6): 811–822.

———. 2024. "The Hyperintensional Variant of Kaplan's Paradox." *Philosophia* 52:187–201.

Lewis, David K. 1971. "Completeness and Decidability of Three Logics of Counterfactual Conditionals." *Theoria* 37 (1): 74–85. https://doi.org/10.1111/j.1755-2567.1971.tb00061.x.

———. 1973a. "Causation." *Journal of Philosophy* 70:556–567.

———. 1973b. *Counterfactuals.* Malden, Mass.: Blackwell.

———. 1986. *On the plurality of worlds.* Blackwell.

McGee, Vann, and Augustìn Rayo. 2000. "A puzzle about de rebus belief." *Analysis* 60 (4): 297–299.

McKay, Thomas, and Peter Van Inwagen. 1977. "Counterfactuals with disjunctive antecedents." *Philosophical studies* 31:353–356.

McLoone, Brian. 2021. "Calculus and counterpossibles in science." *Synthese* 198:12153–12174.

McLoone, Brian, C. Grützner, and M. T. Stuart. 2022. "Calculus and counterpossibles in science." *Synthese* 201:1–20.

Myhill, John. 1958. "Problems Arising in the Formalization of Intensional Logic." *Logique et Analyse* 1:78–83.

Needham, Paul. 2011. "Microessentialism: What is the argument?" *Noûs* 45:1–21.

Nolan, Daniel. 1997. "Impossible Worlds: A Modest Approach." *Notre Dame Journal of Formal Logic* 38 (4): 535–572.

———. 2014. "Hyperintensional metaphysics." *Philosophical Studies* 171:149–160.

———. 2017. "Causal Counterfactuals and Impossible Worlds." In *Making a Difference: Essays on the Philosophy of Causation,* edited by H. Beebee, C. Hitchcock, and H. Price, 14–32. Oxford University Press.

Nute, Donald. 1975. "Counterfactuals and the Similarity of Worlds." *Journal of Philosophy* 72 (21): 773–778. https://doi.org/10.2307/2025340.

———. 1980. *Topics in Conditional Logic.* Dordrecht: Reidel. https://doi.org/10.1007/978-94-009-8966-5.

Odintsov, Sergei, and Heinrich Wansing. 2021. "Routley star and hyperintensionality." *Journal of Philosophical Logic* 50 (1): 33–56.

Pagin, Peter. 2019. "A general argument against structured propositions." *Synthese* 196 (4): 1501–1528.

Plantinga, Alvin. 1978. *The Nature of Necessity.* Oxford: Oxford University Press.

Plebani, Matteo, Giuliano Rosella, and Vita Saitta. 2022. "Truthmakers, Incompatibility, and Modality." *Australasian Journal of Logic* 19 (5): 214–253.

Plebani, Matteo, Luca San Mauro, and Giorgio Lenta. 2024. "Counterpossibles in relative computability theory: a closer look." *unpublished manuscript.*

Pollock, John L. 1976. *Subjunctive Reasoning.* Dordrecht: Springer. https://doi.org/10.1007/978-94-010-1500-4.

Priest, Graham. 1992. "What is a Non-Normal World?" *Logique et Analyse* 35:291–302.

———. 2005. *Towards non-being: the logic and metaphysics of intentionality.* Oxford University Press.

Pruss, Alexander R. 2001. "The Cardinality Objection to David Lewis's Modal Realism." *Philosophical Studies* 104:169–178.

Rantala, Veikko. 1982. "Quantified Modal Logic: Non-Normal Worlds and Propositional Attitudes." *Studia Logica: An International Journal for Symbolic Logic* 41 (1): 41–65.

Restall, Greg. 1996. "Truthmakers, entailment and necessity." *Australasian Journal of Philosophy* 74 (2): 331–340.

Reutlinger, Alexander. 2017. "Does the counterfactual theory of explanation apply to non-causal explanations in metaphysics?" *European Journal for Philosophy of Science* 7:239–256.

Rosella, Giuliano. 2019. *A Truthmaker Semantics Approach to Modal Logic.* ILLC University of Amsterdam: MA Dissertation.

Rosella, Giuliano, and Jan Sprenger. 2023. "Causal Modeling Semantics for Counterfactuals with Disjunctive Antecedents." *Annals of Pure and Applied Logic* 75 (9).

Russell, Bertrand. 1902. *The Principles of Mathematics.* New York: W. W. Norton Company.

Salmon, Nathan. 1986. *Frege's Puzzle.* MIT Press.

Santorio, Paolo. 2018. "Alternatives and truthmakers in conditional semantics." *The Journal of Philosophy* 115 (10): 513–549.

Sbardolini, Giorgio. 2021. "Aboutness paradox." *The Journal of Philosophy* 118 (10): 549–571.

Schaffer, Jonathan. 2016. "Grounding in the image of causation." *Philosophical Studies* 173:49–100.

Seifert, Vanessa A. 2022. "Open questions on emergence in chemistry." *Commun Chem* 5 (49).

Sider, Ted. 2002. "The Ersatz Pluriverse." *The Journal of Philosophy* 99:279–315.

Silva, Francisca. 2021. *On the Topic of Impossibility: a question-sensitive impossible worlds approach to logical omniscience.* Universidade de Lisboa: MA Dissertation.

Skiba, Lukas. 2021. "Higher-order metaphysics." *Philosophy Compass* 16 (10): 1–11.

Skipper, Mattias, and Jens C. Bjerring. 2020. "Hyperintensional semantics: a Fregean approach." *Synthese* 197 (8): 3355–3558.

Soames, Scott. 1985. "Lost innocence." *Linguistics and Philosophy*, 59–71.

———. 1987. "Direct reference, propositional attitudes, and semantic content." *Philosophical topics* 15 (1): 47–87.

Stalnaker, Robert C. 1968. "A Theory of Conditionals." In *Studies in Logical Theory*, edited by Nicholas Rescher, 98–112. Oxford: Blackwell.

———. 1980. "A Defense of Conditional Excluded Middle." In *IFS Conditionals, Belief, Decision, Chance and Time*, edited by W.L. Harper, R. Stalnaker, and G. Pearce, 87–104. Dordrecht: Springer.

———. 1984. *Inquiry.* MIT Press.

Starr, William B. 2014. "A Uniform Theory of Conditionals." *Journal of Philosophical Logic* 43(6):1019–1064.

Tan, Peter. 2019. "Counterpossible non-vacuity in scientific practice." *Journal of Philosophy* 116 (1): 32–60.

Uzquiano, Gabriel. 2015. "A neglected resolution of Russell's paradox of propositions." *Review of Symbolic Logic* 8 (2): 328–344.

van Brakel, Jaap. 2000. *Philosophy of Chemistry.* Leuven University Press.

Van der Laan, David A. 2004. "Counterpossibles and Similarity." In *Lewisian Themes: the Philosophy of David K. Lewis*, edited by F. Jackson and G. Priest, 258–275. Oxford University Press.

Vetter, Barbara. 2013. "'Can' without Possible Worlds: Semantics for Anti Humeans." *Philosopher' Imprint* 13.

Warmbrōd, Ken. 1982. "A Defense of the Limit Assumption." *Philosophical Studies* 42(1):53–66.

Whittle, Bruno. 2022. "The iterative solution to paradoxes for propositions." *Philosophical Studies* 180 (5-6): 1623–1650.

Willer, Malte. 2015. "Simplifying counterfactuals." *Proceedings of the 20th Amsterdam colloquium*, 428–437.

Williamson, Timothy. 2003. "Everything." *Philosophical Perspectives* 17:415–465.

———. 2010. "Modal Logic within Counterfactual Logic." In *Modality: metaphysics, logic, and epistemology*, edited by B. Hale and A. Hoffman, 81–96. Oxford University Press.

———. 2013. *Modal Logic as Metaphysics.* Oxford: Oxford University Press.

———. 2018. "Counterpossibles." *Topoi* 37 (3): 357–368.

———. 2024. *Overfitting and Heuristics in Philosophy.* Oxford: Oxford University Press.

Wilson, Alastair. 2018. "Grounding Entails Counterpossible Non-Triviality." *Philosophy and Phenomenological Research* 46:716–728.

———. 2021. "Counterpossible reasoning in physics." *Philosophy of Science* 88 (5): 1113–1124.

Woodward, James. 2003. *Making Things Happen.* Oxford: Oxford University Press.

Woodward, James, and Christopher Hitchcock. 2003. "Explanatory Generalizations. Part I: A Counterfactual Account." *Noûs* 37:1–24.

Yablo, Stephen. 2014. *Aboutness.* Oxford: Princeton University Press.

Yagisawa, Takashi. 2010. *Worlds and Individuals, possible and otherwise.* Oxford University Press.

Yu, Andy D. 2017. "The modal account of propositions." *Dialectica* 71 (4): 463–488.