Thinking the Impossible

Arguments for Impossible Worlds in Semantics

MSc Thesis *(Afstudeerscriptie)*

written by

Tom Schoonen
(born March 13th, 1991 in Haarlem, The Netherlands)

under the supervision of **Prof. Francesco Berto** and **Dr. Paul Dekker**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:
July 08, 2016

Members of the Thesis Committee:
Dr. Maria Aloni
Prof. Francesco Berto
Prof. Arianna Betti
Dr. Paul Dekker
Prof. Robert van Rooij
Dr. Jakub Szymanik (*chair*)

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

This dissertation defends the use of impossible worlds in natural language semantics, by providing two arguments. First, a methodological argument is made, showing that the use of world postulates in semantics does not commit the semanticist to the ontological existence of worlds. Secondly, the argument from utility for impossible worlds is strengthened by providing a step towards a formal ordering of impossible worlds to improve the semantics of counterpossibles. Together, these arguments show that the use of impossible worlds in semantics is vindicated by their usefulness in capturing certain features of natural language.

Acknowledgement

I would first of all like to thank my supervisors Franz Berto and Paul Dekker for their patience and help throughout this project. They helped me find my way through my own thoughts concerning worlds and their status in semantic models and helped formulating these ideas in a clear and precise way.

I also want to mention Derek Ball, Brian Rabern, and Anders Schoubye for suggesting some very helpful literature and Arianna Betti who has been an invaluable academic mentor to me these last six months. Thanks to Saskia Watts and Thom van Gessel for reading a complete draft of this dissertation and thanks to my parents for their support. A special thanks to Thom, who has been a great sparring partner throughout the entire Master of Logic and this dissertation and who pushed me to get clear on the technical details when I was stuck pondering the philosophical picture.

And to Lieke, without whom this would have been impossible.

Contents

Introduction

One can't believe impossible things

Alice, Alice in Wonderland

> "There is no use in trying," said Alice. "One can't believe impossible things."
>
> "I dare say you haven't had much practice," said the Queen. "When I was your age, I always did it for half an hour a day. Why, sometimes I've believed as many as six impossible things before breakfast."

In the beginning of Lewis Carroll's *Alice in Wonderland*, Alice has a hard time accepting that it is possible to believe the impossible. Yet, in real life, it does not seem to require that much practice as the Queen has us believe. I can believe that it rains in Amsterdam, I can wonder what Lewis Carroll would write about if he were alive today, I can hope for the band to get back together and there are many more things that I can have an *intentional attitude* towards.[1] And, as Alice will later come to appreciate, I can even have intentional attitudes about *impossibilities*. I can seek a unicorn, even though unicorns are necessarily non-existent (if we believe Kripke 1980), I can believe Fermat's Last Theorem to be false, even though Andrew Wiles proved it to be true 20 years ago. I can even hope that Amy squared a circle.

As Priest (2005) notes, "[i]ntentionality is a fundamental feature—perhaps the fundamental feature—of cognition" (p. 5) and it seems that there is nothing that limits this feature to be directed towards possibilities. Though the research of intentionality itself belongs, more or less, to the field of philosophy of mind, the fact we can ascribe such attitudes through natural language expressions (e.g., 'Mary believes that Fermat's Last Theorem is false') gives rise to all sorts of problems in standard natural language semantics. In standard semantics (cf. Gamut 1991; Heim & Kratzer 1998), the semantic value of sentences is taken to be a set of possible worlds, namely, those worlds in which the sentence is true. So, the semantic value of, to take a classical example,

(0.1) Grass is green

[1] I will use 'intentional attitude' as opposed to 'propositional attitude' in order not to suggest that every intentional attitude is a propositional attitude. Following Montague's (2007, pp. 204-5) warning not to conflate the two by using them interchangeably. In line with this, I will follow Priest (2005) in calling verbs such as 'believes', 'knows', etcetera *intentional* verbs, as opposed to the, maybe more common, *intensional* verbs.

is the set of (possible) worlds in which grass is indeed green.

Such a possible worlds framework is very elegant, for it allows us to apply all sorts of set-theoretic operations on semantic values, as well as to define entailment as truth-preservation. Relatedly, the analysis of attitude reports that is often used in the possible worlds framework is, what Bach (1997) calls, the *relational analysis of attitude reports*. This analysis suggests that sentences of the form $\ulcorner a\ v's\ \varphi \urcorner$ are to be analysed as expressing a relation, $v's$, between an agent, a, and the semantic value of a sentence, φ.[2]

Given this picture, Alice' worry that one cannot believe impossibilities seems to be correct. For, consider the following sentence:

(0.2) Fermat's Last Theorem is false

As the semantic value of a sentence is the set of worlds in which it is true, the semantic value of (0.2) is the empty set. It is an impossibility that Fermat's Last Theorem is false, so there is no possible world in which 'Fermat's Last Theorem is false' is true. However, if the semantic value of (0.2) is the empty set, then this would be similar to the semantic value of *any* impossible sentence. If so, what then do we make of the following sentence:

(0.3) Mary believes that Fermat's Last Theorem is false

The problem for the semanticist is not so much that Mary can believe such an impossibility, but that someone can truthfully utter (0.3). How to model this and how to do so in a way that attitude ascriptions of different impossibilities can still differ in truth-value (as Mary does not believe that $2+2 \neq 4$, which also gets assigned the empty set as semantic value) is an open issue.

One possible way to solve the issue sketched above is to extend the worlds-framework beyond merely possible worlds. That is, maybe the semanticist should include *impossible worlds* in her framework to make sense of sentences such as (0.3). An example of an impossible world is the world the Mad Hatter, at some point, describes to Alice:

> If I had a world of my own, everything would be nonsense. Nothing would be what it is, because everything would be what it isn't. And contrary wise, what is, it wouldn't be. And what it wouldn't be, it would. You see?

This dissertation is an extensive argument for the use impossible worlds in natural language semantics. There is a variety of problems for possible worlds semantics, such as the logical omniscience problem, Frege's puzzle, etcetera. Most of these problems arise due to the fact that possible worlds are deductively closed (so, every agent believes all the consequences of her beliefs) and that necessary truths are

[2] In general, people argue that an attitude report reports a relation between an agent and a *proposition*. However, it is not an uncontroversial statement to claim that the semantic value of a sentence is a proposition. I will remain neutral on the issue here (cf. Dummett 1973; Lewis 1980; Rabern 2012a; Schoonen 2014).

true in all possible worlds (so all agents believe all logical truths). I argue that allowing impossible worlds in one's semantics is a very simple solution to most of these problems. The dissertation is centred around two original arguments that, taken together, aim to make a strong case for the acceptance of impossible worlds in models of natural language semantics.

The first argument aims to show that there are no great ontological costs connected to accepting impossible worlds in semantics. The argument, rests on an instrumentalist view concerning semantics, while respecting the value of *metaphysical* research as an orthogonal field of study. The second argument can be considered as a strengthening of 'new' arguments from utility. For we will suggest a step towards a formal ordering of impossible worlds in order to get to a better semantics of *counterpossibles*—i.e., counterfactuals whose antecedent is impossible. Thereby showing that impossible worlds are indeed useful in natural language semantics.

The dissertation is structured as follows. In the first chapter I will discuss possible worlds semantics in more detail. I will some of the limitations of possible worlds semantics, before turning to two accounts that claim to solve the problems of possible world semantics: *structuralism* and allowing *impossible worlds*. I argue that structuralism ultimately is not without some very serious problems of its own and turn to introduce an intuitive picture of adding impossible worlds. The second chapter builds up to the first original argument (presented in the third chapter), namely, as Nolan (1997) puts it, 'counting the costs'. That is, we will discuss what the ontological repercussions are of accepting impossible worlds in one's semantic models. There are many arguments concerning the ontology of impossible worlds and we will briefly consider some ontological accounts. However, through a discussion of Yablo's (2001; 2010) fictionalism, we will build up to my own account. In the third chapter, I argue for, what I call, *semantic agnosticism*. I argue that, as a semanticist, one should go for an instrumentalist point of view concerning the use of impossible worlds, while thereby not ruling out the study of the metaphysics and ontology of worlds as a valuable field of study in its own right; it is merely a non-related field of study. In the fourth chapter I will provide what I take to be the most basic semantics that one could have with impossible worlds and show that it satisfies all that we hope to get out of an impossible worlds semantics. In the last chapter, I aim to strengthen the argument from utility for the use of impossible worlds in semantics by suggesting a formal similarity ordering for counterpossibles. As such a formal ordering has not been suggested before, we will evaluate the utility of such an ordering for the semantics of counterpossibles. Finally, I conclude.

Before we dive into the impossible, let me briefly mention the notational conventions used throughout this dissertation.

I will use single quotes to distinguish mention from use; thus, 'Lewis' consists of five letters, whereas Lewis consists of flesh and blood. I use denotation brackets, $[\![\cdot]\!]$, to indicate the semantic values of linguistic items.[3] Thus, if we take names to denote their bearers, then $[\![$'Lewis'$]\!]$ = Lewis. I will suppress the quotes that indicate mention from use when context allows, for example, within denotation brackets. A complete notation would require the model, $\mathcal{M}$, and the context, c, assignment

[3] For a very interesting note on the history of this notation, see Rabern (forthcoming).

function, g, and index, i, with regards to which expressions are evaluated. If context allows, I will suppress these, however, for completeness:

I take a model to be an ordered triple of non-empty sets of worlds and objects and an interpretation function—i.e., $\mathcal{M} = \langle W, D, \mathcal{J} \rangle$.[4]

I follow Lewis (1970, 1980) in taking the context to be an ordered tuple of all parameters relevant to determine what is said.

I take the index to be an ordered tuple of all parameters that can be shifted by natural language expressions (cf. Lewis 1970, 1980).

Note that when I use $[\![\varphi]\!]^{\mathcal{M},w} = 1$, this is equivalent to when others might use $\mathcal{M}, w \models \varphi$. In some of the parts of the dissertation, I will use the former, whereas in others I will use the latter notation.

[4] Later, when we introduce impossible worlds, we will specify a subset of W, P, that contains only the possible worlds.

Chapter 1

Possible Worlds Semantics and its Limitations

> We might join Lewis in throwing our hands in the air upon an unentertainable supposition, but even obviously impossible propositions can be entertained
>
> Vander Laan (2004, p. 260)

This chapter will introduce *possible worlds semantics* and argue why it is we need to extend it with *impossible* worlds. That is, this chapter will show that there are limitations to possible worlds semantics and it will argue that a simple solution is the addition of impossible worlds to our semantic models.

The chapter is structured as follows: I will begin with a brief overview of the history of possible worlds semantics, after which I will discuss some of the well-known problems for such possible worlds semantics. I will then discuss a group of theories that argues that the addition of structure to propositions solves these problems, the *structured proposition theories*. However, I discuss two arguments against such theories (put forth by Ripley (2012) and Jago (2014)). I will then turn to a solution that will be the focus of the dissertation: the addition of impossible worlds. I will end this chapter with a section on why one should allow impossible worlds in her model.

1.1 Possible Worlds Semantics

In this section we will discuss possible worlds and their use in the semantics framework for formalizing natural language (e.g., Chierchia & McConnell-Ginet 1990; Gamut 1991; Heim & Kratzer 1998; Von Fintel & Heim 2011) and the limitations that such a semantics has in modelling certain features of natural language. This section draws heavily on work by Partee (1989) and Jago (2014, Chapter 1), as well as the introductions to most of the articles and books on impossible worlds.

1.1.1 Possible Worlds Semantics; A Short History

The notion of possible world traces back to, at least, Leibniz, however, it was only later used in formal models. For example, as Partee (1989) notes, Tarski used the notion *alternative models*, which is similar to what semanticists now call possible worlds (especially, in hindsight, given the development of model theoretic semantics). However, possible worlds semantics arguably finds its roots in Carnap (1956), where Carnap introduces the notion of *state-descriptions*. Even though state-descriptions are "crucially different from possible worlds" (Partee, 1989, p. 93), they are, arguably, what inspired the possible worlds semantics. Carnap argues that state-descriptions are sets of sentences such that for every atomic sentence of the language the set either includes the atomic sentence or its negation, not both and not more sentences (1956, p. 9). This makes state-descriptions ultimately linguistics objects, whereas "the possible worlds of possible-world semantics are parts of the model structures *in which languages are interpreted*" (Partee, 1989, p. 93, emphasis added).[1]

It has to be noted that, before possible worlds were used in the semantics of natural language, possible worlds were use most notable by Kanger (1957), Kripke (1959), and Hintikka (1962). Kripke, for example, refers to the points of evaluation in his modal logic as 'possible worlds' when he says that "a proposition is necessary if and only if it is true at all 'possible worlds'" (Kripke, 1959, p. 2), an insight he got from Leibniz (cf. Berto & Plebani 2015).

More related is Hintikka's (1962; 1969) work on epistemic and doxastic logic. Hintikka analysed these notions—i.e., belief and knowledge—in terms of sets of possible worlds. Namely, those (accessible) possible worlds in which the belief (or fact known) is true. Using the notion of possible worlds in the semantics for *natural language* was first done by Richard Montague in his seminal work on a semantics of a fragment of English (Montague, 1970a,b).

Where Chomsky applied mathematical methods to syntax, the revolutionary move for Montague (1970a,b) was to apply mathematical methods to semantics as well. As is stressed by multiple authors, Montague was also the first to provide an explicit mapping from syntax into semantics (cf. Partee 1989; Janssen 2016).

Montague thought that the main goal of semantics is to "characterize the notions of a true sentence (under a given interpretation) and of entailment" (1970b, p. 223, fn. 2). In this project, Montague saw no theoretical difference between formal languages and natural languages; he thought that the same mathematical methods could be applied to both. The method used by Montague to give a semantics for languages is that of *model theoretic semantics*. This, as Janssen (2016) nicely puts it, "means that, using constructions from set theory, a model is defined and that natural language expressions are interpreted as elements (or sets, or functions) in this universe".[2]

Montague's work has been of incredible value for the development of modern formal semantics and Montagovian models are still the standard in semantics (cf. Chierchia & McConnell-Ginet 1990; Gamut 1991; Heim & Kratzer 1998; Von Fintel & Heim 2011; Pickel 2015).

[1] Carnap's account is still mentioned in the literature on *possible worlds* ontology. This emphasises that the view he held comes very close to contemporary views about possible worlds.

[2] Note that, as Janssen (2016) points out, the Montagovian models are not meant to be *metaphysical* models, but merely models of language. This point will be important, and comes back, later in this dissertation.

1.1.2 Limitations of Possible Worlds Semantics

I will now briefly discuss some of the well-known limitations of possible worlds semantics. However, before we do, I want to briefly note another role possible worlds have been argued to play, namely, that of meaning.

Propositions and Meaning

Possible worlds also have been argued to play a fundamental role in the philosophy of language: the 'role' of meaning. This is, most famously, defended by Stalnaker (1976a,b). Stalnaker argued that the meaning of sentences can be represented by sets of possible worlds, namely, the set of worlds where the sentence is true—i.e., the *proposition*. The idea is that propositions are the meanings of sentences: they are the things that ultimately are true or false, the things that remain constant under translation, the referents of 'that'-clauses, and the things we believe, doubt, fear, know, etcetera.[3]

Informally, the meaning of 'Kangaroos have tails' is the set of possible worlds in which kangaroos have tails. Interestingly, this way of characterizing propositions goes nicely with the entailment of sentences, attitude ascriptions, and more pragmatic features of assertions.

I want to note that throughout this dissertation I aim to remain neutral on what the 'meaning' of sentences are and whether or not the standard worlds semantics approach is the correct one to capture natural language meaning. (See, for example, Dekker (2012) and Stokhof (2013) for a dynamic account of meaning or Lenci (2008) and Turney & Pantel (2010) for a distributional view on meaning.) I will therefore often talk of the 'semantic value' of sentences without specifying whether or how this relates to meaning.

Let us now turn to a short discussion of the limitations of possible worlds semantics, in order to do so, we will briefly provide a very standard syntax and semantics for a possible worlds semantics.

Syntax and Semantics

Here we will provide a sketch of a standard possible worlds first-order language. This is merely for illustration and can be skipped if the reader is familiar with possible worlds semantics. The reason we do describe it here is that we will sometimes refer back to it when discussing the impossible worlds semantics.

Our language consists of constants, $a, b, c, \ldots$, variables, $x, y, z, \ldots$, n-ary predicates, $P^n, R^n, \ldots$, a '$\Box$'-operator, a universal quantifier, '$\forall$', and two connectives: '$\neg$' and '$\wedge$'. Remember that we take a model to be an ordered triple of non-empty sets of worlds and objects and an interpretation function—i.e., $\mathcal{M} = \langle W, D, \mathcal{J} \rangle$.

[3] Note that some have been sceptical about the idea that all of these roles can be fulfilled by one entity. For example, Dummett (1973) and Lewis (1980) argue that the entities that figure in compositional semantics and the objects of intentional attitudes are distinct (cf. Rabern 2012a,b; Schoonen 2014).

Syntax:

- If π is an n-ary predicate and $\alpha_1, \ldots, \alpha_n$ are terms (i.e., constants or variables), then $\pi(\alpha_1, \ldots, \alpha_n)$ is a formula
- If φ is a formula and x is a variable, then $\forall x \varphi$ is a formula
- If φ and ψ are formulae, then $\neg\varphi$, $\varphi \wedge \psi$, and $\Box\varphi$ are formulae
- Nothing else is a formula.

The other operator, quantifier, and connectives, '$\vee$', '$\rightarrow$', '$\exists$', and '$\Diamond$', can be defined with the given connectives in the usual way. Note, however, that we can only get away with this as long as we restrict ourselves to *possible* worlds. We will see later, in the chapter on impossible worlds semantics, why this is the case.

Semantics:

- If a is a constant, then $\llbracket a \rrbracket^w = \mathcal{J}(a)$
- If x is a variable, then $\llbracket x \rrbracket^{w,g} = g(x)$, where g is an assignment-function from variables to objects
- If π is an n-ary predicate and $\alpha_1, \ldots, \alpha_n$ are terms, then $\llbracket \pi(\alpha_1, \ldots, \alpha_n) \rrbracket^w = 1$ iff $(\llbracket \alpha_1 \rrbracket^w, \ldots, \llbracket \alpha_n \rrbracket^w) \in \mathcal{J}_w(\pi)$
- If φ is a formula, then $\llbracket \neg\varphi \rrbracket^w = 1$ iff $\llbracket \varphi \rrbracket^w = 0$
- If φ and ψ are formulae, then $\llbracket \varphi \wedge \psi \rrbracket^w = 1$ iff $\llbracket \varphi \rrbracket^w = 1$ and $\llbracket \psi \rrbracket^w = 1$
- If φ is a formula and x a variable, then $\llbracket \forall x \varphi \rrbracket^{w,g} = 1$ iff for all assignment functions, g', such that $g'[x]g$, $\llbracket \varphi \rrbracket^{w,g'} = 1$ (where $g'[x]g$ is an assignment function, g', that differs at most from g in its assignment to x)
- If φ is a formula, then $\llbracket \Box\varphi \rrbracket^w = 1$ iff for all worlds, w', $\llbracket \varphi \rrbracket^{w'} = 1$

With this generic sketch of what a standard possible worlds semantics looks like, we can now turn to some of its limitations.

Logical Omniscience and Frege's Puzzle

Intuitively, attitude ascriptions report a cognitive relation between an agent and a proposition. For example, 'Lewis believes that kangaroos have tails' describes the relation of believing between Lewis and the proposition that kangaroos have tails. Formally, the two aspects of attitude ascriptions (i.e., Hintikka's epistemic logic and Stalnaker's account of propositions) work nicely together. For example, an agent, a (e.g., Lewis), has an attitude, v (e.g., 'believes'), towards a proposition, φ (e.g. 'kangaroos have tails'). The sentence 'Lewis believes that kangaroos have tails' is then true if and only if in all the accessible belief-worlds for Lewis ('$\forall w'$ s.t. $w\mathcal{R}^{\mathrm{a}}_{\mathrm{v}}w'$'), it is true that kangaroos have tails. Formally represented as:[4]

$$\llbracket a\ v\text{'s}\ \varphi \rrbracket^w = 1 \text{ iff } \forall w' \text{ s.t. } w\mathcal{R}^{\mathrm{a}}_{\mathrm{v}}w' : \llbracket \varphi \rrbracket^{w'} = 1$$

[4] These formal accounts are often inspired by work of Hintikka (1969); see for a common account Von Fintel & Heim (2011, Ch. 2).

Two of the most well-known problems for possible worlds semantics (logical omniscience and Frege's puzzle) are both related to attitude ascriptions.

The problem of logical omniscience comes in a variety of forms (cf. Priest 2005; Bjerring 2013; Jago 2014). In its most basic form, the problem comes down to this:

> Take any logical or mathematical truth, for example Fermat's last theorem:
>
> (1.1) There do not exist three positive integers a, b, and c, such that $a^n + b^n = c^n$, for any integer value of n strictly greater than 2
>
> Given that (1.1) is a mathematical truth, it is a necessary truth. That is, (1.1) is true in all possible worlds. It follows that (1.1) is thus true in all the belief-worlds or knowledge-worlds of any arbitrary agent. Thus, everybody knows and believes (1.1) (or any mathematical and logical truths for that matter). However, this is counter-intuitive, for there are many who do not know or believe that (1.1) is true.

More formally, Priest (2005) puts the problem as follows:

> For any sentence, φ, and any agent, a:
>
> if $\vDash \varphi$ then $\vDash a\ v's\ \varphi$

The conclusion that everybody knows all mathematical and logical truths is of course absurd. To borrow from Jago (2014), it seems fair to say that Frege did not know the falsity of his Basic Law V and did indeed believe in its truth at the time of his writing it. So, we either have to give up that mathematical and logical truths are necessary—i.e., true in all possible worlds—or we have to give up our analysis in terms of these possible worlds.

A somewhat related issue for possible worlds semantics and its account of attitude ascriptions is the problem of Frege's Puzzle. In its most general form, Frege's puzzle is taken to show that possible worlds are not fine-grained enough to capture attitude ascriptions, especially when rigid designators are involved (note that some take this to be an argument against rigid designators and direct reference, see Elbourne 2010). For example, consider the following sentences:

(1.2) The ancients believed that Hesperus is the brightest star in the morning sky

(1.3) The ancients believed that Phosphorus is not the brightest star in the morning sky

However, given that Hesperus is Phosphorus in every possible world (cf. Kripke 1980), (1.3) has the same semantic value as

(1.3′) The ancients believed that Hesperus is not the brightest star in the morning sky

This means that, given (1.2), the ancients held contradictory beliefs. Many more such examples are discussed in the literature, for example, Lois believing that Clark Kent can fly and that Clark Kent cannot fly. All these consequences seem too counter-intuitive to be accepted (although some, such as Soames 1987, 2008, seem to bite this bullet).

Another area where possible worlds both proved to be very useful and, simultaneously, have run into great problems is their application to counterfactuals.

Counterfactuals and Counterpossibles

In order to distinguish counterfactuals from indicative conditionals consider the difference between the following two sentences:

(1.4) If Oswald did not kill Kennedy, then somebody else did.

(1.5) If Oswald had not killed Kennedy, then somebody else would have.

Though of very similar structure, these two sentences seem to have different truth-values. (1.4) seems intuitively true, whereas (1.5) might well be false. Sentences such as (1.5) are often referred to as *counterfactuals*. Even though there does not seem to be a very clear definition of what counterfactuals are (at least, none that everybody agrees on), definitions are often something along the following lines (cf. Edgington 1995; Bennett 2003):

> A counterfactual is a conditional in the subjunctive mood with an antecedent that is known to be false.

(However, note that there can also be counterfactuals with a true antecedent.) The most common analysis of such counterfactuals is the Lewis/Stalnaker-analysis (cf. Stalnaker 1968; Lewis 1973; Sider 2010). On such an analysis, a counterfactual, '$\varphi \mathrel{\Box\!\!\rightarrow} \psi$', is true if and only if the consequent is true in the world *most similar to the actual world* where the antecedent is true. 'Most similar to the actual world' means, as eloquently put by Stalnaker, that "the world selected *differ minimally* from the actual world", i.e., that "there are no differences between the actual world and the selected world except those that are required [. . .] by the antecedent" (1968, p. 104, original emphasis). So, consider again the counterfactual from above and its analysis:

(1.5) If Oswald had not killed Kennedy, then somebody else would have.
(Let this sentence be represented by '$\varphi \mathrel{\Box\!\!\rightarrow} \psi$'.)

(1.5′) $[\![\varphi \mathrel{\Box\!\!\rightarrow} \psi]\!]^{w_@} = 1$ iff for any $w \in W$, if $[\![\varphi]\!]^{w} = 1$ and for any $w' \in W$ s.t. $[\![\varphi]\!]^{w'} = 1, w \preceq_{w_@} w'$, then $[\![\psi]\!]^{w} = 1$

So, the counterfactual (1.5) is *not* true, for if we take the actual world and make the minimal changes to make the antecedent true (that is, to make the world so that Oswald had not killed Kennedy), we end up in a world in which nobody else killed Kennedy (assuming that Oswald was operating alone). There are many subtleties concerning the analysis of counterfactuals such as context-sensitivity and a variety of constraints on the similarity ordering of worlds. We will get back to these in Ch. 5, where we will discuss the problem of counterpossibles in more detail.

Possible worlds semantics have proved to be quite useful for the analysis of counterfactuals (though not without its problems, see, for example, Goodman 2004 and Veltman 2005, who discusses the Tichy-problems). However, the Lewis/Stalnaker-analysis also seems to make the wrong predictions when impossible antecedents are involved: this is the problem of counterpossibles. For example, consider the following two sentences:[5]

(1.6) If Amy had squared a circle, Amy would be famous

(1.7) If Sarkozy had squared a circle, Amy would be famous

Given that the antecedent is true in no possible world, there is no closest possible world to evaluate the consequent of (1.6) and(1.7) in. Hence, they are both vacuously true. However, intuitively (1.6) and (1.7) differ in truth-value: (1.7) seems intuitively false, whereas (1.6) seems true. As it is, possible worlds semantics cannot account for such different intuitions.

The analysis of counterfactuals, and the problems of counterpossibles, are much more complicated. We will get back to a detailed analysis of both in Chapter 5.

More Perks and Limitations

The areas discussed above are only a fragment of the applications of possible worlds in semantics and philosophy of language. For example, Groenendijk & Stokhof (1984) have used possible worlds in the analysis of questions. In the original work of Groenendijk and Stokhof, questions partition the set of possible worlds in possible answers to the question asked. Also, Stalnaker (1978) used sets of possible worlds for his notion of *common ground*. This is the set of worlds that all participants of the discourse consider to be actual. After an assertion of a proposition, say φ, the common ground is updated with φ if all participants accept the assertion. But the above was not meant as an exhaustive list, merely as an indication of the usefulness of possible worlds semantics (cf. Stanley 2008). For an early account of the use of possible worlds in semantics, see Partee (1989).

On the other hand, there are also more problems for the possible worlds framework. For example, that of inconsistent fictions. Possible worlds semantics is said to be able to deal with truth in fictions (cf. Lewis 1978), however, the account quickly runs into problems concerning impossible fictions. Say that we write a novel about Amy, who squares a circle. If the story of when Amy squares the circle is told in the beginning of the book, this results in the empty set of worlds for the Lewisian modal analysis of such sentence. Then, the story of her acquired fame, told later in the book, becomes inexplicable (cf. Berto 2013). Possible worlds semantics cannot account for this.

This concludes a variety of perks and serious problems for possible worlds semantics. Even though some possible worlds semanticists have continued work on furthering the possible worlds programme, others suggest that these problems are so devastating that we cannot simply ignore them and have proposed alternatives to the possible worlds semantics framework. We will turn to these next.

[5] These are examples from Ripley (2012, p. 99).

1.2 Solution by Structure

In this section we will discuss the structuralist's solution to these problems. However, as we will note, structuralist approaches have problems of their own. We will discuss two problems for the structuralists and then suggest that there is another solution for the problems of possible worlds semantics that seems to be more intuitive and elegant.

Structuralists take as their starting point the claim that the problems described above for possible worlds are problems for possible worlds "no matter how fine-grained" (Soames, 1987, p. 52). Therefore, they argue, there has to be something more to propositions. What plays this additional role, according to these structuralists, is structure.[6]

In general, structured proposition theorists claim that we need to add structure to propositions in order to make distinctions between sentences such as:

(1.8) $2 + 2 = 4$

(1.9) $x^{\frac{1}{2}} = \sqrt{x}$

However, structure alone is also not enough, for we also want to distinguish between sentences with similar structures, such as:

(1.10) Eva ate a pie

(1.11) John saw a house

Arguably, (1.10) and (1.11) have a similar structure, however, we still want to say that they have different contents. Thus, structured proposition theorist hold that *both* the structure of the proposition and the content of its constituents are important (cf. Salmon 1986; Soames 1987; King 2007; Ripley 2012; King 2014). King (2014) nicely formulates the general idea of structured propositions theorists:

> [S]tructured proposition theorists hold that sentences express propositions that are complex entities (most of) whose constituents are the semantic values of expressions occurring in the sentence, where these constituents are bound together by some structure inducing bond that renders the structure of the proposition similar to the structure of the sentence expressing it.

This allows structured proposition theorists to have different views on either what the content of constituents is or on what the structure of such propositions is. For example, there are those who hold that the content of the constituents is their intension (cf. Carnap 1956; Lewis 1970; Cresswell & von Stechow 1982), there are those who hold that the content of the constituents is something such as their

[6] King (2014) notes that there are two main motivations for structured propositions: (i) the fact that possible worlds are too coarse-grained (these are the arguments we will focus on here) and (ii) the fact that we want to distinguish between terms that are directly referential (cf. Kaplan 1989) and terms that are rigid designators (cf. Kripke 1980). We will ignore this last issue here as it is of no great importance to the main project of this dissertation.

Fregean sense (cf. Chalmers 2011), and there are those who hold that the content of the constituents is their denotation (cf. Salmon 1986; Soames 1987, 1989; King 2007; Soames 2010). The latter are often called Neo-Russellians and we will discuss the Neo-Russellian theories in our discussion of structured proposition theories.

Finally, before we turn to how structured propositions are used to solve some of the problems for possible worlds semantics, it is very important to note that theories of structured propositions are ultimately *metaphysical* theories. That is, they are often theories of what propositions are—i.e., concerning the nature of propositions. It is the semantics of structured proposition theories that we will focus on (however, we cannot discuss the latter without briefly introducing the former).[7]

In order to understand how structured proposition theorists avoid the problems of possible worlds semantics, we need to briefly discuss the semantics that follows from their metaphysical picture of propositions (we follow Pickel 2015 who draws on Salmon 1986). In this semantic theory, predicates express a relation to their arguments and connectives are modelled as relations between propositions (often indicated with SMALL CAPS). Below we provide a very crude representation of the semantics of sentences for structured propositions (see Pickel 2015, p. 11 for a more elaborate account):[8]

If π is an n-ary predicate and $\alpha_1, \ldots, \alpha_n$ are terms:

$$[\![\pi\alpha_1, \ldots, \alpha_n]\!]^w = \langle \mathcal{J}_w(\pi), \langle [\![\alpha_1]\!]^w, \ldots, [\![\alpha_n]\!]^w \rangle\rangle$$

If φ and ψ are formulae:

$$[\![\neg\varphi]\!]^w = \langle \textsc{neg}, [\![\varphi]\!]^w \rangle$$
$$[\![\varphi \wedge \psi]\!]^w = \langle \textsc{conj}, \langle [\![\varphi]\!]^w, [\![\psi]\!]^w \rangle\rangle$$

With this semantics in place, we can now see how the structured proposition theorists avoid the problem of logical omniscience. Remember that (on most accounts) to hold a certain attitude towards something is to stand in a doxastic/epistemic relation to some proposition. So, if we can distinguish between two propositions, one can stand in a doxastic/epistemic relation to one, without standing in such a relation to the other. Let us return to the example above of the two mathematical statements: (1.8) and (1.9). On the structuralists' accounts these sentences render the following two propositions:[9]

(1.8) $2 + 2 = 4$

(1.8′) $\langle [\![=]\!], \langle\langle [\![+]\!], \langle [\![2]\!], [\![2]\!]\rangle\rangle, [\![4]\!]\rangle\rangle$

(1.9) $x^{\frac{1}{2}} = \sqrt{x}$

(1.9′) $\langle [\![=]\!], \langle\langle \textsc{pow}, \langle [\![x]\!], [\![\tfrac{1}{2}]\!]\rangle\rangle, \langle \text{sqrt}, \langle [\![x]\!]\rangle\rangle\rangle\rangle$

Given that (1.8′) and (1.9′) differ in structure and the content of their constituents, they express different propositions on the structuralists' theory. Thus, we can believe one without thereby believing the other.

[7] Remember that Montague-models are explicitly *not* metaphysical models.

[8] Pickel (2015) does not provide a semantics for '□'- or '◊'-operators. He does give the semantics of the quantifiers and the 'believe'-operator. See his work (p. 11).

[9] For simplicity's sake, I take identity, square root, addition, and power to be predicates.

This nicely seems to solve the issues concerning logical omniscience and the closure of doxastic and epistemic states under entailment. However, there are some problems for the structured propositions theorists as well.

Frege's Puzzle, again

The main problem for structured propositions theorists here is that of Frege's Puzzle and, what I will dub, the *connective*-objection.[10]

Structuralists often argue from a 'fineness-of-grain' argument (Ripley, 2012, sec. 1.3); ironically, structured propositions themselves cannot account for one of the main puzzles concerning fineness-of-grain, namely, Frege's Puzzle. (The argument where Frege's Puzzle is turned around against structured propositions is, as far as I am aware, due to Ripley (2012).)

Related to (1.2) and (1.3) above, consider the following two sentences:

(1.12) Hesperus is Hesperus

(1.13) Hesperus is Phosphorus

Given the structuralists' semantics given above, these sentences are rendered as follows:[11]

(1.12′) $\langle [\![\text{Hespersus}]\!], \langle [\![\text{is}]\!], [\![\text{Hesperus}]\!] \rangle \rangle$

(1.13′) $\langle [\![\text{Hespersus}]\!], \langle [\![\text{is}]\!], [\![\text{Phosphorus}]\!] \rangle \rangle$

This shows that (1.12) and (1.13) can only express different propositions if the semantic value of 'Hesperus' is not equal to the semantic value of 'Phosphorus'. However, as Ripley notes, 'Hesperus' and 'Phosphorus' "have the same structure, the same referent, and the same possible-world intension" (2012, p. 105). Thus, (1.12) and (1.13) express exactly the same proposition. This makes that the structured proposition theories *also* make the wrong predictions considering the following sentences:

(1.14) The ancients believed that Hesperus is Hesperus

(1.15) The ancients believe that Hesperus is Phosphorus

This is a very problematic result for the Neo-Russellian structuralists. Soames seems to bite the bullet and take this counter-intuitive result on board, others have proposed solutions for the structuralists. What is important, is that Ripley notes that this argument is not meant to show that structured proposition theorists cannot deal with this problem, it is only meant to show that *structuralism* in and of itself does not hold the solution to Frege's Puzzle.

[10] Note that another major problem for structured proposition theories is that they are not compatible with compositional, Montagovian semantics. However, in a recent article Pickel (2015) argues that we *can* develop a Montagovian semantics with structured propositions. Therefore, we will not discuss this problem here.

[11] Note that I take 'is' here as a copula, whereas one might also take it to be identity. However, the argument does not hinge on this.

For example, Ripley discusses the structuralist account of Chalmers (2011). Chalmers uses his signature, two-dimensionalism move in order to allow terms to be rigid designators, while they still differ in cognitive significance. In this case, the primary intension, for Chalmers, is a set of *epistemically possible scenarios*, at which 'Hesperus' might be what the person at the centre of the centred world takes it to be (i.e., is epistemically possible for the agent at the centre). This allows Chalmers to make very fine-grained distinctions, for example, for an agent at the centre, it might be epistemically possible for Hesperus and Phosphorus to be distinct.

However, as we saw above, it is metaphysically not possible for Hesperus and Phosphorus to be distinct, thus, these epistemically possible scenarios of Chalmers, might as well be thought of as impossible worlds. This is also what Ripley points out, he argues that Chalmers' account "provides these differences [i.e., between (1.14) and (1.15)] by considering circumstances beyond ordinary possible worlds; it is a circumstantialist solution" (2012, p. 106). What Ripley calls 'circumstantialist solution', I call 'impossible worlds solution'. Moreover, as Ripley points out, the structure in Chalmers' account does nothing to add to the solution.

The other proposed solutions by the Structuralists are discussed elaborately in Ripley (2012, sec. 2.2). I want to turn to another argument against structured propositions theories; an argument that I will dub the *connective*-argument.

The Connective-Argument

Jago (2014) notes that the argument by Ripley (2012) against structured proposition theorists, only holds against Neo-Russellians (as opposed to Fregean structuralists). Therefore, he sets out to provide an argument that holds against any form of structured content—Fregean, Russellian, or otherwise. The problem, according to Jago, is that a structured account of content is incompatible with "platitudes about how definitions fix content and meaning" (2014, p. 79). What makes Jago's argument very powerful is that it does not rely on features of language that are already taken to be difficult or contentious, but is based on a merely extensional language. As Jago notes "[i]f structuralism about content fails in such amenable linguistic territory, then it clearly fails for any natural language" (idem.).

The argument is based on structuralism and two very intuitive premises. Jago argues that these three conditions are not mutually compatible.

P1: "If we introduce a new term 't' to a simple [extensional] language using an explicit definition, then for any sentence 'A' and 'B' of that language which differ only in the substitution of the definiendum 't' for its definiens, 'A' and 'B' have the same semantic value" (Jago, 2014, p. 80)

P2: "The semantic value of a connective governed by a truth-table is the truth-function determined by truth-table"[12] (Jago, 2014, p. 81)

P3: Propositions are complex entities of the kind described above by King (2014)

[12] Jago elaborates on what he means by **P2** (his 3.12), however, for our discussion it only matters that the classical connectives do satisfy **P2**—i.e., the classical connectives are governed by a truth-table.

Taking these premises as our starting point, let us briefly summarize Jago's argument. Consider a simple language, $\mathcal{L}_1$, whose connectives are classical conjunction, '$\wedge$', and classical negation, '$\neg$'. By definition, '$\wedge$' and '$\neg$' satisfy **P2**. Let us now stipulate that there is a truth-function, *disj*, that takes pairs of truth-values and returns *true* if one of the input values is *true* and *false* otherwise. We then stipulate that there is no name or predicate in $\mathcal{L}_1$ such that it has *disj* as its semantic value. We can now prove, by induction on the complexity of sentences, that "[n]o sentence in $\mathcal{L}_1$ has any tuple containing *disj* as its semantic value" (2014, p. 81, Theorem 3.1).[13]

Now we extend $\mathcal{L}_1$ with a new two-place connective '$\odot$' such that

$$\varphi \odot \psi =_{\text{def}} \neg(\neg\varphi \wedge \neg\psi)$$

We call our extended language $\mathcal{L}_1^+$. We can now show that every sentence in $\mathcal{L}_1^+$ is expressible by a sentence in $\mathcal{L}_1$ by replacing every instance of '$\varphi \odot \psi$' with '$\neg(\neg\varphi \wedge \neg\psi)$' (by **P1**). Given that we have proved that there is no sentence in $\mathcal{L}_1$ that contains any tuple containing *disj*, it follows that there is no sentence in $\mathcal{L}_1^+$ that contains any tuple containing *disj*. Hence, $[\![\varphi \odot \psi]\!] \neq \langle disj, \langle [\![\varphi]\!], [\![\psi]\!]\rangle\rangle$.

However, given **P2** and our definition of '$\odot$', it follows that '$\odot$' has as its semantic value the function *disj*—i.e. $[\![\odot]\!] = disj$. Hence, $[\![\varphi \odot \psi]\!] = \langle disj, \langle [\![\varphi]\!], [\![\psi]\!]\rangle\rangle$. Contradiction.

Thus, **P1**, **P2**, and **P3** are mutually incompatible.[14] □

Jago (2014, sec. 3.4) goes on to discuss some possible responses by structured propositions theorists. However, we will not discuss those here. For it is not our purpose to judge the connective-argument as a knock-down argument against structured propositions. We do, however, take this (and Ripley's) argument to show that structured propositions are not without some very serious problems of their own.

We will now turn to what will be the starting point of this dissertation. Namely, extending possible worlds semantics to include *impossible* worlds. The following chapters of this dissertation can be seen as an extended, philosophical argument in favour of the use of impossible worlds in semantics. In the final section of this chapter, we will introduce the notion of impossible worlds.

1.3 Solution by Impossibilities

In this section, I will argue that the addition of *impossible worlds* seems to solve the problems for the possible worlds semantics. Note that this section will only sketch the intuitive solutions provided by impossible worlds semantics. A more detailed analysis will have to wait until Chapter 4, where an impossible worlds semantics will be presented. What we need to know about impossible worlds for this section

[13] See Jago (2014, p. 81) for the entire proof.

[14] One might argue that this does not show us much, for **P1** is exactly what the structured propositionalist takes issue with to begin with. However, in the argument presented by Jago, structuralists cannot account for this in an *extensional* language, while arguably one only wants to reject **P1** in intentional contexts.

is that impossible worlds might be incomplete or inconsistent. An example of the former is a world that neither a sentence, nor its negation true; an example of the latter can either be a world where both a proposition and its negation are true, or a world where something impossible is true.[15] As mentioned above, a more detailed discussion of impossible worlds will be presented shortly.

This section is mostly based on work by Priest (1992), Nolan (1997), Ripley (2012), Berto (2013), Krakauer (2013), and Jago (2014, Ch. 4).

Let us first consider the problems of logical omniscience and Frege's puzzle. These are problems of fineness-of-grain (cf. Ripley 2012) or hyperintensionality (cf. Jago 2014). The solution provided by impossible worlds semantics is similar for both problems.

Consider again the two logically equivalent sentences:

(1.8) $2 + 2 = 4$

(1.9) $x^{\frac{1}{2}} = \sqrt{x}$

Now imagine that we add only one incomplete world to the set of possible worlds, one that lacks any information concerning (1.9). Then the set of worlds where (1.8) is true is different from the set of worlds where (1.9), as there is at least one world where information concerning (1.9) is lacking, and only that information. Thus, we can believe one without thereby believing the other. Note that in this case we only need incomplete worlds, and not inconsistent worlds yet.[16] A similar solution is provided for Frege's puzzle. Consider again the relevant sentences:

(1.12) Hesperus is Hesperus

(1.13) Hesperus is Phosphorus

In this case we need to add only one inconsistent world to the set of possible worlds, one where Hesperus is still identical to itself and not identical to Phosphorus (see Yablo 1993 for an interesting discussion on the relation between conceivability and possibility, e.g., whether we can conceive of Hesperus not being equal to Phosphorus). Then the set of worlds where (1.12) is true is different from the set of worlds where (1.13) is, as there is at least one world where Hesperus is not Phosphorus and thus where (1.13) is false. Hence, we can believe one without thereby believing the other. In this case we do need inconsistent worlds.

As Berto (2013) notes, counterpossible reasoning might be the most important area where impossible worlds play a role. Consider again the counterpossibles about Amy and Sarkozy trying to square a circle, repeated below:

(1.6) If Amy had squared a circle, Amy would be famous

[15] When I talk about 'worlds' in the context of impossible worlds, it is often of no (great) importance whether the world is possible or impossible. So, you may read my use of 'world' such that it does not matter whether the world under discussion is impossible or merely possible.

[16] Note that we could also add an inconsistent world where one of the two mathematical truths fails.

(1.7) If Sarkozy had squared a circle, Amy would be famous

Imagine, for the sake of the argument, that we can order impossible worlds in a similar fashion to the way we order possible worlds. (This is clearly not as trivial as I make it sound, but we will come back to this in Chapter 5.) For now, imagine that we can hold everything else fixed, while making the antecedent true. So, in the case of (1.7), the only thing that would change is the fact that Sarkozy squared the circle (again, I will not go into the difficulties of this assumption here). Then, given that some innocent assumptions concerning how one acquires fame remain fixed, it seems that Sarkozy would be famous and not Amy. Hence, (1.7) comes out false. Similarly, in the world where Amy squares a circle, she receives the related fame. Thus, (1.6) comes out true.

So, the addition of impossible worlds seems to account for the different intuitions regarding (1.6) and (1.7).[17]

This section aimed to introduce the notion of impossible worlds in semantics and some intuitive ways of how impossible worlds semantics might solve some of the issues discussed above. At first sight, it seems as if the prospects for impossible worlds semantics look good and this is enough of a reason to develop such an account in more detail. Throughout this dissertation we will discuss how impossible worlds help solve these issues in more detail. For now, we will end this chapter with a brief note on why one should accept impossible worlds into ones semantical toolkit.

1.3.1 Why Impossible Worlds

Clearly, one needs to argue *why* one would accept impossible worlds into her semantical toolkit. It is important to note that most theorists (e.g., Berto 2013; Vander Laan 1997; Krakauer 2013; Jago 2014) use the arguments, presented below, to show why one should 'believe' in impossible worlds. As will become apparent in due course of this dissertation, I will argue for a weaker claim, namely, that these are arguments why one should 'accept' impossible worlds 'in her semantical toolkit'. The reason for this distinction will become clear in Chapter 3 of this dissertation. For now, we will stick with the traditional sentiment.

I stand behind Will's (who is one of the participants of Stalnaker's 1996 dialogue on impossible worlds) eloquent words in that I believe that there is "no argument for the existence of merely possible worlds that is not matched by a parallel and equally compelling argument for the existence of impossible worlds" (Stalnaker, 1996, p. 194). This sentiment is often echoed by impossible world theorists: e.g., Vander Laan (1997, p. 598) says "that many or most of those who believe in possible worlds already believe in impossible worlds"; Nolan (2013, p. 361) says that "the motivations for employing possible worlds come with related motivations to employ impossible worlds as well"; and Krakauer (2013, p. 990) says that "there is no principled reason why we shouldn't also believe in incomplete or inconsistent sets of sentences, propositions, or states of affairs". (See also Priest 1992 and Nolan 1997.)

[17] A discussion of counterpossible will be a large part of the remainder of this dissertation. I will discuss counterpossibles in more detail in Chapter 5.

I agree with the above sentiment. If one believes (for whatever reason) in possible worlds, then there is no principled reason not to believe in impossible worlds. To elaborate, I will discuss three arguments in favour of possible worlds, extended to impossible worlds: the argument from ways, the argument from counterpossible reasoning, and the argument from utility.

However, before we do, note that most of these arguments seem to be in line with the *parity thesis* (cf. Berto 2010)—i.e., a thesis that impossible worlds are of the same kind as their possible better halves. We will discuss the parity thesis in more detail in Chapter 3.

Let us first discuss 'the argument from ways' (cf. Vander Laan 1997, p. 598) and its extension to an argument for impossible worlds. The argument is most famously, and eloquently, put forward by David Lewis and, as one can never quote Lewis enough, here is his original argument at length:

> I believe there are possible worlds other than the one we happen to inhabit. If an argument is wanted, it is this: It is uncontroversially true that things might have been otherwise than they are. I believe, and so do you, that things could have been different in countless ways. But what does this mean? Ordinary language permits the paraphrase: there are many ways things could have been besides the way that they actually are. On the face of it, this sentence is an existential quantification. It says that there exist many entities of a certain description, to wit, 'ways things could have been'. I believe things could have been different in countless ways. I believe permissible paraphrases of what I believe; taking the paraphrase at its face value, I therefore believe in the existence of entities which might be called 'ways things could have been'. I prefer to call them 'possible worlds' (1973, p. 84)

As can be seen from Lewis' reasoning, there is a fairly straightforward way to extend this argument to an argument for impossible worlds. Namely, as there are many ways things could have been, there are also many ways things could not have been. Many impossible worlds theorists have indeed extended the argument thusly (cf. Vander Laan 1997; Berto 2013; Krakauer 2013). For example, "there exist entities that might be called 'ways things could not have been' [. . .] I prefer to call these 'impossible worlds'" (Vander Laan, 1997, p. 598). First of all, I want to flag that at this point I do not want to commit myself to the *existence* of entities, especially not through there use in natural language (that is, within the Quinean methodology for ontological commitments, see Sec. 2.1). Secondly, as some have noted (cf. Berto 2013; Krakauer 2013), the argument from ways is, in and of itself, not very convincing. As Berto (2013) notes, Lewis himself did not *only* rely on this argument and the argument from ways "is hardly convincing".

According to Berto (2013), one of the main reasons for believing in impossible worlds is their use in accounting for counterpossible reasoning. That is, reasoning about impossible, counterfactual situations (such as (1.6)). As Berto notes, there is reasoning about counterpossible situations that is correct, and some reasoning that is not. It seems that we should be able to account for these different intuitions about such reasoning.

Jago (2014) generalizes this argument to discourse in general (the following argument is adapted from Jago 2014). Consider Russell's letter to Frege, in which he explains the paradox he deduced from Frege's Basic Law V. Jago notes that if Frege's Basic Law V is false, then it is necessarily so, that is, it is impossible. Two things follow from this if one merely accepts possible worlds, namely that Frege should have known that Basic Law V is false when he wrote it (logical omniscience) and that the discussion between Frege and Russell is meaningless. Thus, to engage in meaningful discourse about mathematical and logical falsehoods (especially if it is an unknown falsehood at the time), we should allow for impossible worlds.

The above argument is, in a slightly different form, also used for the acceptance of possible worlds. It is best known as the argument from utility and is also originally put forward by Lewis. As Vander Laan (1997, p. 600) puts it, the argument is put "in Quinean terms: improvements in utility and economy of ideology are sometimes worth controversial ontology". If this argument holds for possible worlds, which many philosophers seem to think it does, then the extended argument should also hold for impossible worlds. As Berto (2013) says, the argument from utility stresses "impossible worlds as a device by means of which particular linguistic, logical and philosophical issues can be regimented and analyzed". The last chapter of this dissertation aims to show this through work on the semantics of counterpossibles. For now, the above aims to show that impossible worlds, intuitively, seem to be at least as useful as possible worlds and even solve problems that (merely) possible worlds semantics has. Hence, the argument from utility, if it has any validity, should also be accepted for impossible worlds.

1.4 Concluding the Introduction

This chapter provided an introduction to the topic of impossible worlds in semantics, as well as a motivation why it is worth exploring the topic. We began by discussing the use of *possible* worlds in natural language semantics and some of the limitations of such an analysis or, as Ripley (2012) puts it, we pointed to some of the egdes where possible worlds semantics breaks down.

Eventually, we introduced the main topic of this thesis, *impossible worlds semantics*. Intuitively, adding impossible worlds to our semantics seems a good way to solve some of the problems for possible worlds semantics.

However, this was merely an introductory chapter. Much more needs to be said and in much more detail. We will do so in the following chapters; the main (original) body of the dissertation. In the following two chapters, we will discuss some ontological issues concerning impossible worlds. That is, we will evaluate the costs of accepting impossible worlds in one's semantics. I will ultimately argue, after a discussion of Yablo's (1998; 2001; 2010) figuralism, for a form of *semantic agnosticism*. I will try to delineate my account more clear through a comparison with Yablo's figuralism and Divers' (2006) agnosticism.

After we have counted the costs, and concluded that there are almost none, we will discuss a very simple semantics that makes use of impossible worlds. We will spell out some desiderata that we want such a semantics to satisfy and conclude that the semantics presented there does indeed satisfy them. Finally, we will turn to a technical issue concerning impossible worlds semantics, namely, issues concerning the semantics of counterpossibles (cf. Nolan 1997; Goodman 2004; Vander Laan 2004; Williamson 2007; Brogaard & Salerno 2013; Bjerring 2014a). We will provide a step forwards, toward a better understanding of the semantics for counterpossibles, by suggesting a similarity ordering for impossible worlds. And thus providing a positive argument from utility for the use of impossible worlds in semantics.

Chapter 2

Semantics and Ontology

> True, the use of a term can sometimes be reconciled with rejection of its object
>
> Quine (1960, p. 210)

In this chapter, we will start to discuss the costs of allowing impossible worlds into one's semantics. That is, this chapter discusses a particular meta-ontology, namely fictionalism, only in order to build up to the original account presented in the next chapter (which is inspired by Yablo's fictionalism). In order to properly do so, we have to get very clear on the subtle distinction between ontology and meta-ontology. I will not go into the ontological accounts that follow the standard Quinean meta-ontology, for the account that I prefer does not fit this framework. I will elaborately discuss fictionalism, as the account that I will argue for in the next is more closely related to fictionalism than it is to any of the Quinean ontologies.

2.1 Ontology versus Meta-ontology

We need to make a very important distinction: the distinction between first-order and second-order reasoning about ontology. Or, the distinction between *ontology* and *meta-ontology*.[1] As an introduction to this distinction, let us briefly return to the rebirth of ontology in analytic philosophy: Quine's "On What There Is".

Quine (1948) starts his seminal paper with the following paragraph:

> A curious thing about the ontological problem is its simplicity. It can be put in three Anglo-Saxon monosyllables: 'What is there?' It can be answered, moreover, in a word—'Everything'—and everyone will accept this answer as true. However, this is merely to say that there is what there is. There remains room for disagreement over cases; and so the issue has stayed alive down the centuries (p. 21)

[1] I follow, for no particular reason, Van Inwagen (1998) in using the hyphenated version of the word, whereas Berto & Plebani (2015) use 'metaontology'.

Quine proceeds in a dialogue with McX and Wyman to argue for such particular cases. For example, Quine argues that there is no such thing as Pegasus, or, there is nothing to which 'Pegasus' refers. However, it seems that for the sentence 'Pegasus does not exist' to be meaningful, 'Pegasus' has to refer to something. To explain away this apparent paradox, Quine adopts a particular *methodology* to uncover the ontological commitments of sentences. Quine believes that we should translate sentences of natural language (with unwelcome ontological commitments) into canonical notation (i.e., First-Order Logic), in order to see what the real ontological commitments of such a sentence are. This is due to the fact that for Quine "[t]o be is, purely and simply, to be the value of a variable" (1948, p. 32). (Quine gets out of the paradox by claiming that 'Pegasus' translates into 'an x such that x Pegasizes', such an x does not exist, so the sentence 'Pegasus does not exist' is indeed true.) This is the root of the distinction we are about to make.

Note that Quine's dictum, "to be is the be the value of a variable" , does not mean to tell us anything about *what* we take to be viable values of variables. It merely suggests *a method* for finding out what is (in the existential sense of 'is'). Van Inwagen (1998) tells us that we should keep these two issues clearly separated. The investigation what, if anything, can be the value of a variable (and thus, what is), is first-order ontology. The question whether or not being the value of a variable is the right methodology of getting at what there is, is second-order ontology. Or, *meta-ontology.*

Berto & Plebani (2015) describe meta-ontology as follows:

> [I]f the key question for ontology, as Quine told us, is 'What is there?', then the (twofold) key question for meta[-]ontology is "What do we mean when we ask 'What is there?"', and 'What is the correct *methodology* for ontology?' (p. 2, emphasis added)

Let us consider one example to make the distinction at hand vividly clear. Consider the following sentence and its translation into canonical notation:

(2.1) There is a chair

(2.1′) $\exists x(x \text{ is a chair})$

One may argue that (2.1) commits us to the existence of concrete objects. However, as Berto & Plebani (2015) note, such commitment is, in and of itself, "not [a] logical entailment" (p. 41). For example, an ontologist might agree with Quine that to be is to be the value of a variable, but this only commits her to a certain meta-ontology. She might still believe that whatever (2.1′) commits us to is simply not a concrete object. So, agreeing with Quine only forces one to explain to what the translation of (2.1′) commits us. For example, she might be sceptical of concrete objects in general and believe that chairs can be reduced to "bundles of properties or universals" (idem.). In this case, our ontologist's Quinean meta-ontology does not commit her to concrete objects.

The issue is that there are two points where one can disagree with the claim that (2.1′) commits us to concrete objects. First, one may argue that (2.1′) does not capture the ontological commitments of (2.1). This would be to disagree on the meta-ontology—the methodology of ontology. Secondly, one may argue that (2.1′) *does* capture the ontological commitments, however, she might hold that this commits us to properties/universals/etcetera instead of concrete objects. This would be a dispute in ontology, that is, *what* do translations such as (2.1′) commit us to.

This point cannot be emphasised enough as it still seems that not everybody has the distinction between ontology and meta-ontology clear (for example, Meinongianism, a particular meta-ontology, is often mentioned in the same breath with genuine realism and ersatzism, both particular ontologies). It is especially of great importance for this dissertation as the account that I propose in the next chapter is of a meta-ontological nature, not an ontological one.

Quinean Ontologies

Most ontologies of impossible worlds are in the standard, Quinean meta-ontological framework. The two main ontological accounts are both realist accounts—that is, worlds exist, they differ only in what they are. For the genuine modal realist, worlds are concrete objects, of the same kind as the actual world is. Ersatz modal realists, on the other hand, argue that other worlds are abstract objects of some sort (there is a great variety of ersatz account, differing on *what* kind of abstract objects worlds are).

I will not discuss these two ontologies here, for, as we will see, the account that I will argue for is a meta-ontological account, more similar to fictionalism (though it is not a fictionalist account). See Lewis (1986) and Divers (2002) for a genuine realist account of possible worlds and Yagisawa (1988, 2010) and Kiourti (2009) for an extended account of impossible worlds. For arguments against extended modal realism and extended ersatz accounts, see Vander Laan (1997) and Jago (2012, 2014, 2015). There are also those who argue for a hybrid account, letting possible worlds be concrete objects, whereas impossible worlds are abstract objects. For such an account, see Berto (2010) and Krakauer (2013). For a good overview of these accounts and the distinction between ontology and meta-ontology, see Berto & Plebani (2015).

2.2 Modal Fictionalism: Fictive Worlds

As mentioned, most ontologies discussed above adhere to the ‘standard’, Quinean meta-ontology. However, there are accounts of possible worlds that give up this Quinean meta-ontology. For example, *modal fictionalism.* From our discussion of fictionalism, we will build towards the original meta-ontological account presented in this dissertation. We will first discuss motivations for adopting a different meta-ontology through a discussion of fictionalism. To introduce how such an account would work for (im)possible worlds, we will discuss so called modal fictionalism; fictionalism with regards to possible worlds (cf. Rosen 1990; Nolan 2016). This brings us to one of the main problems for modal fictionalism (it has also been taken to be a problem for fictionalism in general): *The Bomb* (cf. Yablo 2001). There is an account of fictionalism presented by Stephen Yablo (2001; 2010) that is inspired by this problem (Yablo writes that he learned to love the bomb). Yablo presents a particular meta-ontology that is called ‘Figuralism’ (following Yablo 2010, p. 5).

This discussion of figuralism will only be for exposition and an introduction to a different meta-ontology. It is important to get clear on this, for the original account that will be presented in the next chapter is inspired by elements of Yablo's figuralism. In order to clearly explicate my own account, it is important to get clear on the many subtleties that make up Yablo's account.

2.2.1 Fictionalism; Attitudes of Frivolity

Fictionalism is a view that concerns discourses that seem to commit us to (possibly) unwanted entities, for example, mathematical talk of numbers or modal talk of worlds. Eklund (2011) describes fictionalism as "the view that claims made within [a] discourse are not best seen as aiming at literal truth but are better regarded as a sort of 'fiction'". Or, as Yablo (2001) puts it, "sentences are put advanced in a fictional or make-believe spirit" (p. 74). The reason is that fictionalists want to be able to talk of numbers and other contentious entities, *without accepting such entities in their ontology*. As Stanley (2001) says, the "fictionalist holds that the best semantic theory for a discourse may not be a good guide to the ontological commitments of the person who uses that discourse" (p. 51). This means that the Quinean meta-ontology, i.e., that the translation of sentences into canonical notation entail the ontological commitments of such sentences, no longer holds.

Ironically, as Berto & Plebani (2015) point out, the fictionalist idea is spun off of a discussion on the ontological commitments of fictional talk by Quine (1953):

> One way in which a man may fail to share the ontological commitments of his discourse is [...] by taking an attitude of frivolity. The parent who tells the Cinderella story is no more committed to admitting a fairy godmother and a pumpkin coach into his own ontology than admitting the story as true (p. 103)

Fictionalists take this idea and propose that when we engage in discourse concerning undesirable entities, we also talk with "an attitude of frivolity". For example, consider the following two sentences:

(2.2) Bilbo is a hobbit

(2.3) Two is an even prime

(2.2) is a clear example of a sentence concerning a fictional character. The problem with such fictional sentences is that even though they feel intuitively true, they are literally false (due to the lack of a referent for 'Bilbo').[2] Fictionalists argue that

[2] Lewis (1978) suggests that, in order to accommodate our intuitions, we analyse the sentences as follows:

(2.2′) In the stories of Tolkien, Bilbo is a hobbit

(2.2′) is true, thus, Lewis suggests, we should analyse fictional sentences with a silent modal operator as in (2.2′). However, Stanley (2001) argues that the Lewisian analysis of fictional sentences "sits uncomfortably with ontology in the [...] fictionalist spirit" (p. 37). This is because Lewis analysis such operators in terms of possible worlds, which the modal fictionalist wants to get rid of. Rosen (1990), however, takes the modal fictionalist's operator to be primitive, in which case there is no real problem.

something similar is going on in (2.3). That is, strictly speaking (2.3) is false, but when the speaker is uttering (2.3), she "is normally indulging in a pretence" (Berto & Plebani, 2015, p. 85).

There is one main issue that needs to be addressed by all fictionalist accounts. Consider the following sentence pair:[3]

(2.3) Two is an even prime

(2.4) There is a man in the closet

The fictionalist claims that (2.3) is false but good in mathematical discourse. However, it seems that if you utter (2.4) and it is false (i.e., there is no man in the closet), you have made a mistake. A fictionalist has to explain the different standards for (2.3) and (2.4). Most fictionalists argue that in the case of (2.3) the talk of undesirable entities is allowed to be false, *because they are useful tools or representational aids* (cf. Balaguer 1996).

Finally, the fictionalist needs an account for the relation between the way fictional things are and the way things really are. This is often called *the principle of generation* (cf. Stanley 2001; Berto & Plebani 2015 and Walton 1990, who has worked extensively on the principle of generation in fiction). This is of great importance, for it is said that this principle is what links truthful utterance *in the fiction* with reality. For example, Stanley (2001) argues that such a principle is needed "to link real world objects and events with objects and events within the pretence" (p. 39). Berto & Plebani (2015) on the other hand argue that such a principle tells us "what we should pretend to be the case, or what we should take as true in the fiction, in the given circumstances" (p. 88). I take it that both Stanley and Berto and Plebani mean something along similar lines, something that is more explicitly mentioned in Yablo (2001). The fictionalist needs to be able to tell a story about what is correct by the fictionalist lights—e.g., why is '2+2=4' correct and why is '2+2=5' not.[4] "Schematically, we can say that S is correct iff $C(S)$, where C is the condition the fictionalist puts forward as making for correctness" (Yablo, 2001, p. 75).

Fictionalists differ on what makes for a good condition. A generic example would be something along the following lines:

(2.2) Bilbo is a hobbit

(2.2′) Bilbo is a hobbit if and only if, according to the Tolkien *Lord of the Rings* pretence, Bilbo is a hobbit

(See Yablo 2001 and Eklund 2011 for good overviews of the possible conditions that fictionalists propose.)

[3] Examples adopted from Berto & Plebani (2015).

[4] One of Stanley's arguments against fictionalism is that such an account cannot be properly given. He argues that fictionalist accounts "tend to over-generate, predicting that discourses can be felicitous that can never be felicitous" (2001, p. 60). We will ignore this here, as this will have no bearing on the account that will be presented in this thesis.

With this general account of fictionalism in mind, let us turn to the application of fictionalism to the undesirable entities of this dissertation: (im)possible worlds. Even though fictionalism has flourished most in the philosophy of mathematics, it recently has reared its head in the debate concerning possible worlds. Fictionalism with regards to possible worlds is called *modal fictionalism* and we will turn to discuss it next.

2.2.2 Modal Fictionalism and the Bomb

Modal Fictionalism is the fictionalist's meta-ontology applied, not to modality, but to talk of possible worlds. (Nolan (2016) notes that one can also be a fictionalist about modality, he dubs this *broad modal fictionalism.*) Proponents of modal fictionalism argue that it has the benefits of modal realism, without paying for it in ontological coin. So, in line with the fictionalist spirit, when we say: 'there is a possible world such that. . . ', this is meant as put forth in some pretence spirit. An immediate question that arises, as Rosen (1990) points out, "is to specify the story the fictionalist plans to exploit" (p. 332). That is, in terms used above, what is the relevant principle of generation for these contested sentences.

The first major proponents of modal fictionalism, Rosen (1990), takes the relevant principle of generation to be one of Lewisian genuine modal realism. So, when one says (2.5), the truth conditions of the utterance are (2.5′):[5]

(2.5) There is a possible world where φ

(2.5′) There is a possible world where φ iff, according to Lewis' *On the Plurality of Worlds* pretence, there is a possible world where φ

According to Rosen, with such a (silent) prefix modal fictionalism is indistinguishable from modal realism (i.e., Rosen argues that we should paraphrase the left-hand side of (2.5′) into the right-hand side when we engage in modal discourse).

With this preliminary sketch of modal fictionalism, we will now turn to a very influential argument *against* modal fictionalism (and fictionalism in general) known as the Brock-Rosen Objection, or, as Yablo (2001) calls it, The Bomb.[6]

Consider the following two assumptions; one is a statement, that, according to modal fictionalism, is true and the other is a standard axiom of modal logic:

(NEC): $\Box\varphi$ is true if and only if according to the modal pretence, φ is true at all possible worlds

(T): $\Box\varphi \rightarrow \varphi$

Given these two assumptions, we can derive the following contradiction:

[5] A subtlety that is noted by Nolan (2016) is that it may not be the case that φ itself is true in the story of possible worlds, but only a translation of φ into the language of possible worlds. I will ignore this subtlety here, but one may read on as if on each correct instance 'φ's are replaced with their translations, 'φ^*', into the language of possible worlds.

[6] In the following paragraph, I will mostly follow Yablo (2001) and Nolan (2016) in their exposition, however, see Nolan (2016) for references to the original argument.

(1) According to the modal pretence, at all worlds it is true that there are many other worlds

(2) Necessarily, there are many other worlds (1, **NEC**, MP)

(C) There are actually many other worlds (2, **T**)

(**1**) is true on any modal fictionalist account (whether the modal pretence assumes genuine worlds or ersatz worlds is of no great importance), however, the conclusion, (**C**), contradicts the fictionalists main selling point, namely, that, actually, there are no possible worlds.

Many take this to be bad enough already, however, Yablo (2001) argues that we can also turn the argument around.

(1′) Let φ be modal sentence, then φ is true iff according to the modal pretence φ is true

(2′) There are really no possible worlds

(C′) According to the modal pretence, there are no possible worlds (1′, 2′, MP)

(1′) and (2′) are true opinions of any modal fictionalist. (1′) simply describes her paraphrase strategy and (2′) expresses her opinion about the status of possible worlds. However, (**C′**) is clearly not true, for the fictionalist uses the modal pretence because *in* the modal pretence, there *are* possible worlds.

These results are very problematic for the modal fictionalist (see Yablo 2001 for the argument against other forms of fictionalism). I will specifically focus on Yablo's solution to the problem in the next section (see Nolan 2016 for other fictionalist solutions).

2.3 How Yablo Learned to Love the Bomb

In this section we will discuss Yablo's (1998; 2001; 2010) account of fictionalism. That is, I will distil what I take to be the best aspects of Yablo's fictionalism, as Yablo often discusses multiple versions of fictionalism (cf. Yablo 2001, 2010) and has defended a variety of them. In the introduction to his collected works on metaphysics, *Things; Papers on Objects, Events, and Properties* (2010), he distinguishes between three forms of fictionalism he has defended or defends. Let us briefly introduce some terminology Yablo uses, in order to get a clear exposition of the different accounts discussed by Yablo.

Let φ be a sentence concerning some undesirable entities. Then $|\varphi|$ expresses its *literal* content—i.e., the content that fictionalists deem false—and $||\varphi||$ expresses its *concrete content*. Yablo defines concrete content as follows:

> $||\varphi||$ is the proposition true in a world w iff φ is true at some v concretely indiscernible from w, albeit perhaps richer than w in [the contested entities in question] (2010, p. 3)

With this in mind, let us turn to the three forms of fictionalism, which Yablo has defended or defends, that he describes in his introduction to *Things*.

> *Object Fictionalism* (or *Figuralism*)
>
> The asserted content when uttering φ is the "real-world fact" that makes φ true in the pretence. That is, $||\varphi||$ plays "the role of φ's metaphorical content, if we understand metaphors as moves in prop-oriented make believe games"
>
> *Presuppositionalism*
>
> The asserted content when uttering φ is "the 'logical remainder' when the content of $|\pi|$ of operative presuppositions is subtracted from $|\varphi|$". That is, we define $|\varphi| - |\pi|$ as the part of the literal content that is not about the operative pressupositions. Then we get that $||\varphi||$ is the part of $|\varphi|$ that is not about whether the contested entities exist.
>
> *Subject-Matter-Ism*
>
> The asserted content when uttering φ is the part of $|\varphi|$ "that IS about the subject matter under discussion" (2010, p. 5, original emphasis)

When explicating Yablo's view, I will focus on Figuralism, as presented in Yablo (2001). Note that Yablo's account is very complex and seated in many subtle distinctions. I will try to describe most of them as clearly as possible when relevant to this dissertation.

2.3.1 Figuralism

As already pointed out above, Yablo emphasizes the distinction between the use of X as representational aids or things-represented, where X is introduced to talk about some subject (say, numbers to talk about mathematics). Yablo phrases this distinction eloquently with the use of some examples as follows:

> There are actually two roles X's can play. Sometimes they function as *representational aids*. This is how butterflies function in 'I had butterflies in my stomach,' and numbers function in 'the number of Martian moons is 2.' Other times they [i.e., X's] function as *things-represented*. This is how butterflies function in 'the butterflies were splattered all over the windscreen,' and how numbers function in 'there are no numbers'
> (2001, p. 81, original emphasis)

Why 'standard' fictionalism runs into trouble, Yablo argues, is that they never consider "for a moment that X's will function as things-represented" (idem.). So, in order to avoid some of the problems, we have to allow for the fact that X's can be a representational aid (or not) and they can be things-represented (or not). To accommodate this, Yablo does away with the "all-purpose governing" fiction, and allows for a different kind of pretence. (Yablo (2001) talks here of 'games', where different games allow for different demands on sentences, I will follow the more neutral terminology of 'pretence', as it is used throughout this dissertation. Also, it seems to me that this already hints towards subject-matter-ism, we will come back to this.) It is important to note that the context demands what pretence is applied to a sentence, as different pretences can be applied to the same sentence (cf. Yablo 2001, p. 82). This is what Yablo, at that point in his paper, calls *reflexive fictionalism.*

Consider the following two examples by Yablo (2001); when talking to philosophers one might truthfully want to say that there are no numbers (disengaged discourse), however, when talking to a mathematician, one might truthfully want to say that the number of even primes is one (engaged discourse). As it stands, Yablo argues, reflexive fictionalism cannot account for, what he calls, *engaged* discourse. Yet we do want an account that covers both of these cases.[7] Yablo argues that we only need to modify the account for the engaged discourse and does so through the use of *parasitic* pretences. I will not go into more detail with regards to these here and turn to one final distinction Yablo makes before he arrives at his figuralism.

What is important, Yablo argues, is that the fictionalist mathematician and the Platonist mathematician can now have meaningful discussions about even primes. However, this immediately gives rise to the question of how it is possible that "mathematicians can happily communicate despite having different views of the nature, and even existence, of mathematical objects" (2001, p. 84). Yablo's answer is somewhat related to Carnap's distinction between questions that are internal and external to a given framework (cf. Berto & Plebani 2015, Ch. 5). Yablo applies this distinction in such a way that the Platonist gets 'real truth', whereas the fictionalists gets 'real agreement'. Yablo notes this and concludes that the two 'real's are significantly different. Although both 'real's "look back to real content", they do so in significantly different lights—that is, they "draw on different aspects of the notion" of real content (ibid., p. 85). Crucially, Yablo argues that these different notions can come apart.

The best way to characterise the two aspects of real content involved, is by considering two different answers to the question: 'What makes real content real?'

Answer 1: What makes real content real is that it concerns real, actual things.

Answer 2: What makes real content real is that it is really asserted.

Yablo calls the former *objectual reality*, whereas the latter is what Yablo calls *assertional reality.* Even though these two notions are often run together, it is of crucial importance to realize that these notions "can come apart" (Yablo, 2001, p. 85). When we do pull these aspects apart, we can explain the communication between the platonic mathematician and her fictionalist friend.

This is, roughly, Yablo's version of fictionalism. It originated from a re-appreciation of the Bomb. For example, if we would derive the contradiction of the Bomb within mathematics, i.e., with numbers, we would end up with a sentence such as:

(2.6) According to the number pretence, the number of numbers is 0

[7] Again, it seems to me that this already hints towards a sort of subject-matter-ism.

As Yablo says, this does not only seem to contradict the starting assumption of fictionalism, the embedded sentence (‘the number of numbers is 0’) also seems self-refuting. How can the number of numbers be 0, as 0 is itself a number? According to Yablo’s fictionalism, we should treat the first occurrence of ‘number’ as literal, while the second occurrence as in the pretence. Yablo argues that both are used in a different pretence, or, put differently, the first occurrence is things-represented, while the second is a representational aid.

In general there are two ways to interpret a fictionalist account (I will follow Stanley 2001 in his exposition of this distinction). The distinction boils down to whether we take the fictionalist hypothesis to be *prescriptive* or *descriptive*. That is, *revolutionary* fictionalists take the fictionalist approach to be used as a reconstruction of the discourse used, it is therefore prescriptive. *Hermeneutic* fictionalists argue that the fictionalist approach is “how the discourse is *in fact* used” (Stanley, 2001, p. 36, emphasis added), it is thus descriptive. This leaves us with the question how to interpret Yablo’s fictionalism.

Yablo himself suggests that one might expect the account sketched above to be a revolutionary account. Remember that revolutionary fictionalists propose to paraphrase all the problematic utterances and hermeneutic fictionalists argue that the problematic utterances are actually meant to be taken as in the fictionalists’ proposal. However, if Yablo’s proposal would be revolutionary, it seems too complex and *ad hoc*. As Yablo (2001) puts it, “[a] revolution with this many rules is unlikely to generate a whole lot of fervor” (p. 85). Thus, Yablo concludes, his account should be seen as a hermeneutic account, yet, with a twist. Yablo argues that it is not the case that we have been relative reflexive fictionalists (as he dubs his account) all along. However, he argues that we have been something very similar. That is, according to Yablo “[w]e are people apt on occasion to speak figuratively” and there “is nothing in relative reflexive fictionalism not found already in figurative speech” (2001, p. 85). This is why this version of his fictionalism is dubbed *figuralism*: Yablo argues that discourse concerning undesirable entities is on a par with metaphorical discourse.

Yablo provides us with a list of examples that he takes to strengthen his view. Here are some of them:

> ‘Prime numbers are mostly odd’ ≈ ‘Stomach butterflies do not sit still but flutter about’
>
> ‘The number of F’s = the number of G’s iff there are as many F’s as G’s’ ≈ ‘My bottom line is your bottom line iff both of us are prepared to such-and-such and neither is prepared to settle for anything less’

As the sentences on the right-hand side of the similarity symbol do not commit us to the existence of stomach butterflies and bottom lines, so do the left-hand side sentences not commit us to the existence of numbers, Yablo argues.

Yablo continues to give another argument for fictionalism (i.e., that it would better reflect practice), however, I will not go into that here (see sections 11 and 12 of Yablo 2001).

2.4 Conclusion

Note that, so far, we have not yet applied Yablo's figuralism to possible worlds nor to impossible worlds. It is fairly trivial to extrapolate figuralism to possible worlds and impossible worlds the like. On such an account, we can say meaningful sentences about (im)possible worlds, while believing that such objects do not exist. However, I believe that, in some sense, this account is too strong.

In the next chapter, we will turn to my account, which aims to make less strong *ontological* claims, yet has similar benefits *for semanticists.*

Chapter 3

Semantic Agnosticism

> The central question is whether impossible worlds or the like are of any use, especially for the purpose of semantic enquiry. If they are of no use, then who cares whether they exist or what they are like? And if they are of some use, then we should be able to find a place for them within our ontology, *if only as a convenient fiction*
>
> Fine (2013, p. 4, emphasis added)

As figuralism seems to reject the existence of worlds, we will retreat to a slightly weaker claim, namely, that the semanticists does not know (or should not care) whether worlds exist. In this chapter, I will argue for my own, meta-ontological, account of the impossible worlds that are used in formal semantics. I will first give a sketch of the account that I have in mind. Afterwards, I discuss the account of Divers (2006), that seems to be very similar, yet which is significantly different. Then, I will provide more explicit arguments for my own account. I discuss a possible objection and argue that this objection points to a very subtle distinction and that, when we make this distinction, the argument no longer holds. Finally, I position my account in the fictionalist/agnosticist landscape.

3.1 Semantic Agnosticism

Let us, for exposition, briefly return to the mathematician. Remember that Yablo (2001) wondered how mathematicians with different ontological beliefs could have meaningful discussions and (dis)agreements. I think that this is indeed the right question to ask, yet that the, somewhat convoluted, conclusion that Yablo draws is the wrong one to draw.[1] I think that the right conclusion from the mathematicians' meaningful discussion is this:

> *Qua* mathematician, she does not care whether or not numbers exist or what their nature is. So, when she engages in a discussion with other

[1] As mentioned before, after Yablo (2001), Yablo came to hold a variety of different fictionalist accounts that already seem to come closer to the account that I present here.

> mathematicians, *qua* mathematicians, about mathematics, the different *metaphysical* views do not affect the discourse. It might still turn out that our mathematician, *qua* philosopher, also has meaningful discussions about the existence and nature of numbers, yet these discussions are completely *orthogonal* to each other.

To put the point more explicitly, it seems to me that it does not matter for her work, *qua* mathematician, whether or not numbers exist. An often heard wonderment in philosophy of mathematics classes (especially from science students who take such a course) boils down to this: 'how would the answers to these questions [about the nature or existence of numbers] affect the work of mathematicians?' My point is that they would not affect the work of mathematicians, *qua* mathematician. If numbers turn out to exist, the mathematician, *qua* mathematician, might rejoice that her work relates to reality, but she will not alter her work. If numbers turn out to not exist, the mathematician, *qua* mathematician, will probably shrug her shoulders and continue to use mathematical objects in her proofs.

However, she might, despite all the above, still have very fierce, meaningful, discussions concerning views on the nature and existence of numbers, *qua* philosopher. I believe that these two views are orthogonal in that the one should not necessarily influence the other (it might, but it should not necessarily be so).

The story is nicely explicated with the example of numbers and mathematicians and I believe that the same holds for semanticists. In a way, this goes with the view that semantics, qua *pure semantics*, can be thought of as applied mathematics. For example, Divers (2006) describes pure semantics as "pieces of mathematics" (p. 188). The point that I want to make is that the story extends to applied semantics with its interpretation in terms of worlds.[2]

It does not matter for the work of a semanticist, *qua* semanticist, whether or not worlds exist. A semanticist should, *qua* semanticist, keep on using worlds in her models, despite the outcome of ontological discussions about the nature and existence of worlds. However, she might, as many semanticists, also have philosophical interests and fiercely discuss the nature and existence of worlds, *qua* metaphysician. Yet, these two practices (i.e., using worlds in her model, while holding a, possibly incompatible, ontological view) are orthogonal. Note that Fine (2013) makes a similar remark when he says:

> Philosophers have been intrigued by the ontological status of impossible worlds. Do they exist and, if they do exist, then do they have the same status as possible worlds? To my own mind, these questions are of *peripheral interest*. The central question is whether impossible worlds or the like are of any use, especially for the purposes of semantic enquiry. (p. 4, emphasis added)

Though of similar spirit, I do not think the issues of metaphysics are *peripheral*, I take them to be *orthogonal* to the issues in semantics. Finally, note that taking these

[2] For a detailed analysis of the distinction between pure and applied semantics, see Dummett (1974) and Copeland (1979, 1983).

issues to be orthogonal is also where Yablo (2001) (or any fictionalist) and I differ. Where the fictionalist holds that one is allowed to use worlds while *denying* their existence, I want to argue that the semanticists need not (necessarily) deny their existence. She simply does not care (*qua* semanticist) whether or not worlds exist. That is, as a semanticist, she can remain agnostic. Hence, semantic agnosticism.

The above sketches the intuitive idea behind my view. As we will proceed through this chapter, the view will become more precise and more clearly delineated. In order to do so, we will first briefly discuss an account by Divers (2006).

3.1.1 Divers' Agnosticism

As it is, the above is not much more than a statement of the account that I want to defend. I will provide more arguments for semantic agnosticism, and point to some similar thoughts expressed in the literature, presently. However, before I do, it is important to distinguish between my account and the similar account of Divers (2006), which he dubbed *agnosticism.*

Divers is driven by similar motivations as I am and nicely formulates the issue at stake as "the question of whether it is *only* a realist about possible worlds who is entitled to deploy a possible worlds semantic theory and to claim the benefits that it affords" (2006, p. 188, original emphasis). Divers wants to defend an account of where one is allowed to use possible worlds semantics, yet believe that there is only one world and that there are no other possible worlds (not even ersatz worlds). His argument proceeds through a distinction between *pure semantics* and *applied semantics* (cf. Dummett 1974; Copeland 1983; Priest 2006; Berto 2007).

When one opts for a pure semantics, there are no ontological commitments to possible worlds—as pure semantics are 'uninterpreted', mathematical formalisms; "pieces of mathematics" (Divers, 2006, p. 188)—and one is free to use such semantics as she wishes, whatever ontological leaning she has. Applied semantics, however, *does* involve ontological commitments, due to the fact that "a semantics gives an account of meaning *only once* the mathematical formalism of the semantics itself has been explained *in terms of concepts relating to the actual or intended use* of the sentences of the language for which the semantics is given" (Copeland, 1983, p. 202, emphasis added).

Divers proceeds to provide such an applied semantics for propositional modal logic and then argues that the source of the ontological commitments are "one and the same" as the source of some meta-logical results pertaining to validity (the full technical details of which will not concern us here, as we merely want to compare his particular (meta-)ontological view). Divers spells out his agnostic account in such a way that the agnostic may not be in a position to assert certain semantic and meta-logical truths, "[y]et, she accepts, there is a matter of the fact" (2006, p. 208). This is where I object, as I believe that the semanticist should be in a position to assert semantic and meta-logical truths, *without* any ontological commitments.

We may paraphrase Divers' account so that the semanticists may not be in a position to make any ontological claims about worlds, yet she does believe there to be a matter of the fact. When paraphrasing Divers' account, it comes indeed very close to what I have in mind. However, the two are not completely equivalent. Divers goes through a lengthy argument to show that the agnostic is allowed to make certain semantic and meta-logical assertions, as on Divers' account the semanticist's agnosticism does put some constraints on her semantics. This is what I object to.

Moreover, Divers explicitly states a particular ontology that his agnosticist has, namely, that "[t]he agnostic believes in the existence of the (one) actualized possible world and *does not believe* in the existence of any others" (2006, p. 206, emphasis added). I would argue for a stronger form of agnosticism in that the semanticist, *qua semanticist*, does not care whether there is one world or many.[3]

So, it seems that Divers' agnosticism is not as 'agnostic' as I would want it to be. I will leave the discussion of Divers' agnosticism here and return to a more clear delineation of my own account and provide some arguments in favour of it.

3.2 Semantic Agnosticism: The Arguments

I hope that the explication of Yablo's (2001) figuralism and Divers' (2006) agnosticism helps to delineate my own account (we will return to a more detailed description of the relation between these two accounts and my own later on in this chapter). In brief, the view that I aim to defend is that the usage of worlds in the model of a semanticist, should not ontologically commit her to such worlds. The discussion of the use of worlds in the model and of the one on the existence and nature of worlds are two orthogonal discussions. One can use many worlds in her semantics (*qua* semanticist), while, for example, believing that only the actual world exists (*qua* metaphysician). This is a form of instrumentalism, though it is limited to instrumentalism *in semantics*. That is, instrumentalists often allow for unobservable objects "simply as instruments for the prediction of observable phenomena", while simultaneously holding that "unobservable things have no literal meaning at all" (Chakravartty, 2015). However, my point is precisely that there may be a matter of the fact what things there are in the world and whether these include possible worlds, impossible worlds, or not, but the semanticist should not be bothered by this and use whatever 'instruments' she needs to model interesting features of natural language use.

In this section, I will first briefly discuss the view that I describe above and some pointers to, what I take to be, similar sentiments in the literature. Then, I will provide an argument for this view, namely, the argument from utility. Remember that this argument was already presented in Chapter 1 in order to argue that one should 'believe' in (im)possible worlds. Given the account presented here, those arguments need to be re-evaluated. In order to re-evaluate the arguments for worlds in semantics, we will briefly go over the project of semantics and semantic theories, especially in relation to propositions and semantic values. This discussion will help

[3] However, Divers also argues that "the agnostic will not go as far as [...] asserting the non-existence of [...] other, non-actualize, possible worlds" (idem.). Divers does not really go into these ontological issues, so it has to be noted that it is not clear whether or not Divers would disagree (or agree) with me on these matters.

to outline why the semanticist, *qua* semanticist, should not be concerned with issues in metaphysics and will ultimately lead us to re-evaluate the arguments for the use of worlds in semantics.

An aim of semantics is to provide semantic values to expressions in such a way that the semantic value of a complex expression is a function from the semantic values of its parts (and, possibly the syntactic structure). In most semantics, the semantic value of a sentence is a truth-value (or a set of worlds in which the sentence is true). There are a couple of subtleties that need to be noted here, that are quite often ignored and/or conflated.

First of all, we have to be aware that the semantic value of an expression, is not necessarily what that expression denotes. For example, intentional attitude reports are often argued to report a certain cognitive relation of an individual towards a *proposition*. On the semantic level, this idea is reflected by the common semantics, which Bach (1997) calls the *relational analysis of attitude reports*. This analysis suggests that attitude reports express a relation between an individual and a proposition; importantly, on this view the semantic value of the embedded 'that'-clause *is* the proposition in question. For example, Soames (1988, pp. 105-106) says, "attitudes are relations to propositions [. . .] To believe that 'S' is to believe the proposition that S." The idea seems to be the following: attitude reports relate an agent, a, to a proposition, φ, by sentences of the form $\ulcorner a\ v$'s $\varphi\urcorner$, hence, the reasoning goes, the semantic value of 'φ' *must* be the proposition that a has an attitude towards. This is an example of what is often called *propositionalism* (cf. Montague 2007; Schwarz 2016)—e.g., "[t]he semantic value of a 'that'-clause is a proposition" (Bach, 1997, p. 222).

The problem is, that semantic value of an expression need not be equivalent with what 'entity' the expression denotes. For example, Dummett (1973) and Lewis (1980) already argued that the semantic value might come apart from, what they called, the assertoric- or propositional content. Applied to attitude reports, it is argued in Schoonen (2014) that the objects of our beliefs are different from the semantic values of the embedded 'that'-clauses of our attitude reports. So, even thought the objects of one's beliefs may be propositions, the semantic value of the embedded 'that'-clauses need not be. This line of thought can be extrapolated to semantic values in general. Schwarz (2016) puts it as follows:

> Semantic values are here construed as abstract mathematical entities somehow involving possible worlds. These entities are clearly not things competent speakers "grasp", in any substantive sense of the term. They are too coarse-grained to capture intuitive differences in meaning or cognitive significance. They are presumably not involved in the computational processes that underlie our linguistic competence. They are meant to play a different kind of role, but it is not always clear what that role is. Historically, they were often used to give a compositional semantics of intensional (especially modal) constructions. (p. 2)

This passage aims to emphasize that semantic values are merely postulates of the semantic theory and not seated in 'reality' in any way (cf. Ball 2016; Schwarz 2016).

This is precisely the point that I try to make, and Schwarz (2016) goes on to put it very explicitly when he says that "the 'possible worlds' of semantics should not be identified with the possible worlds of metaphysics. They should *not* be taken as independently given at all, but *rather treated as theoretical postulates*" (pp. 2-3, emphasis added).[4]

This point, though recently emphasised in the meta-semantics literature, was already noted in early literature in semantics. For example, remember that Montague did not mean for his models to be models of metaphysics, but of features of language such as entailment (Montague, 1970a,b). Similarly, Lewis (1970) develops his semantics in *Semantic Markerese*, a stipulated system of symbols, in order not to alarm "the ontologically parsimonious" (p. 19). Also Gamut (1991) express, more abstractly, a similar point in that "philosophical objections should never be allowed to weigh heavily *if the aim is the description of natural language*" (p. 47, emphasis added). As well as Sider (2010), who keeps emphasising throughout his book that "model theory isn't metaphysics" (p. 206). When Sider introduces modal propositional logic models, he emphasises that we should not confuse model theory with metaphysics. While introducing the non-empty set of worlds that is in the model, he says:

> This is certainly a vivid way to talk about these models. But officially, W, is nothing but an non-empty set, any old non-empty set. Its members needn't be the kinds of things metaphysicians call possible worlds. They can be numbers, people, bananas—whatever you like. [...] Officially, then, the possible-worlds talk we use to describe our models is just talk, not heavy-duty metaphysics.

Schwarz (2016) makes the point that I aimed to make very explicitly, however, there is a qualification that I want to make. As I pulled apart the discussions one might have *qua* semanticist and *qua* metaphysician, it might very well turn out that the worlds that serve as the semantic postulates, are indeed the worlds that the metaphysician talks about, however, the point is that they need not be. As Lewis (1980) says, "[i]t would be a convenience, nothing more, if we could take the propositional content of a sentence in context as its semantic value" (p. 95).[5]

[4] Schwarz makes similar points throughout his paper:

> "Accepting such models does not mean to believe that possible worlds somehow figure in the fundamental fabric of reality. It does not contradict the assumption that reality is at the bottom completely physical"
>
> (p. 15)
>
> "In general, we should not identify the possible worlds in our linguistic model with a fixed set of independently specified entities [i.e., the set of metaphysically possible worlds]. The possible worlds are simply part of the model; the construction of the logical space is a modelling choice" (p. 16)
>
> "[P]ossible worlds are theoretical postulates" (p. 19)

[5] Schwarz (2016) does say something along these lines when he says that "it might turn out that the space of possible worlds assumed in the philosophical analysis of metaphysical modality can do double duty as the space of possible worlds in pragmatic models of language use [...] But on the face of it, the two job descriptions look very different" (p. 15).

3.2.1 Arguments from Utility

So, given that the worlds used in semantics need not be the metaphysical worlds, but are independently stipulated theoretical postulates, why then is the semanticist justified in stipulating/using these postulates? An obvious point to start, is to look at the arguments for *believing* in worlds discussed in Chapter 1 and then see if these still hold for the postulate worlds.

Note that on semantic agnosticism, one only needs weaker conclusion from the argument from utility. In the original arguments, the usefulness of worlds had to vindicate their acceptance in one's ontology. I, on the other hand, only need the usefulness of worlds to vindicate the use of such *theoretical postulates*. This is a much weaker conclusion that is needed than the conclusion that is originally argued for.

In section 1.2.3 we discussed a number of arguments for worlds, the first of which was the argument from ways. Remember that the argument from ways, as formulated by Lewis (1973), concerned the quantification over the claim that there seem to be many ways the world could have been. Lewis' argument, essentially, rests on a Quinean meta-ontology, when he says that "this sentence is an existential quantification. It says that there *exists* many entities [. . .] I therefore believe in the existence of [these] entities" (1973, p. 84). However, as we saw above, I reject the Quinean meta-ontology, at least, for the formal semanticist and the use of world postulates in her model. So, it seems, that on a methodological level, this argument no longer holds and the semanticist cannot use it for her world postulates.[6]

Secondly, we discussed the argument from utility. This argument, in a variety of forms, was made by Berto (2013) and Jago (2014) for impossible worlds through their use in counterpossible reasoning and contentful discourse about necessary falsehoods, respectively. Remember that Vander Laan (1997) said that the argument is put "in Quinean terms: improvements in utility and economy of ideology are sometimes worth controversial ontology" (p. 600). However, on the account presented above, we rejected the Quinean meta-ontology and therefore the 'ontological costs' are no longer relevant for the discussion. Hence, it seems that the conclusion that we aim for from the argument from utility is weaker than the original conclusion argued for. The original argument claimed that, because of their utility, we have to accept worlds into our ontology. However, for our argument, it suffices for one to accept *world postulates into her semantics toolkit*, as opposed to worlds into her ontology. The point is, as Fine (2013) already noted, either worlds are not useful and the semanticist should not care whether they exist or they are useful and we better make it work with our ontology, which we did by suggesting that their acceptance as useful does not commit us to worlds ontologically. Moreover, it seems that this kind of reasoning—that is, using abstract, theoretical postulates in a scientific theory—is less controversial. As Schwarz (2016) notes, "[a]ppealing to abstract entities in theoretical models is common practice and harmless" (p. 15). Let us go into this reasoning, in relation to semantics, with a bit more detail.[7]

[6] This is even besides the fact that 'ways the world could have been' might not be the entities that play the part of world postulates as we have seen in the arguments from Schwarz (2016).

[7] The following paragraphs are inspired by the work of Ball (2016) and Schwarz (2016), both

To fully determine the utility of these worlds postulates, we have to look at, specifically, their usefulness with regards to the semantic theory and we cannot resort to examples of their usefulness in metaphysics or philosophy of mind. In order to fully determine the usefulness of worlds to a semantic theory, we should consider what (exactly) the aim of a semantic theory is. Obviously, this meta-theoretical question is worthy of a dissertation of its own, therefore, we will only note a point that has been paid a lot of attention to recently.

Recently, meta-theoretical reflections on semantics suggest that semantic theories may aim for something slightly different than original was claimed. That is, recently there seems to be a tendency to take semantics as a 'special science' that aims to represent certain *patterns or features of language.* For example Schwarz says:

> The point of assigning a given set of worlds to a given signal is that this allows us to elegantly capture certain behavioural patterns in the relevant community (2016, p. 11)

Ball (2016) makes similar claims on multiple occasions and, interestingly, Stalnaker (1984b) made similar remarks while discussing related issues. Stalnaker discusses the "charge of ontological extravagance" (p. 4) and argues that what we are really after is *capturing interesting linguistic behaviour.* That is, an agent has certain intentional attitudes towards certain situations and linguistic behaviour is "a particularly rich source of evidence" for these attitudes. To capture this, we need to be able to distinguish between situations that an agent has attitudes towards and the possible worlds framework is merely "useful for characterizing and expressing an agent's attitudes" (Stalnaker, 1984b, p. 4).

On such characterisations of semantics, the aim of a semantic theory is to capture interesting features of language. The point is then that worlds are extremely useful in capturing/representing these features of language.[8] As Schwarz notes, "possible worlds are theoretical postulates whose purpose in a semantic model is to capture interesting patterns in an ultimately non-linguistic and non-intentional world" (2016, p. 19) and, as he noted earlier in his paper, "the point of models in special science is to capture interesting and robust patterns in the world that somehow emerge from the chaotic complexity of microphysical interactions. *If abstract parameters help achieve this goal, that is enough to vindicate their use*" (Schwarz, 2016, p. 5, emphasis added).

This characterises the semantic instrumentalism noted earlier: the semanticist should use whatever tool she needs to capture the interesting features of natural language without worrying about the ontology of those tools. Limiting this to *semantic* instrumentalism, the semanticist may, while doing metaphysics, research the matters of the fact concerning the ontology of worlds, objects, etcetera.[9]

of which are to appear in a forthcoming book on the philosophy of semantics (Eds. Ball, D. and Rabern, B.). I would like to thank Derek Ball and Brian Rabern for pointing me to these articles and sending me advanced copies.

[8] Of course, something that I do not discuss here is the possibility of disagreeing what features of language one should focus on. Possibly, there are features of language that are not easily captured by worlds semantics (for example, lexical meaning similarity). For the purposes of this dissertation, we will leave this discussion aside.

[9] Quine (1960) has a similar approach to semantics, that is, his approach is very behaviouristic,

So, we have argued that the world postulates that the semanticist uses or talks about, *qua* semanticist, need not be the worlds the metaphysician is concerned with (or need not have ontological commitments). That is, the metaphysician may ask questions about the nature of (possible) worlds, questions that the semanticist simply does not care for. It seems that due to propositionalism, people often take the semantic values of sentences to be propositions, however, it might be that these two things come apart. Then, we argued that the semanticist is allowed to use such world postulates based on their utility. That is, we argued that their usefulness is enough to vindicate their use. Let us turn to an example that aims to show that there are features of language use that warrant the acceptance of *impossible worlds* in one's semantic toolkit.

Consider the fact that it is a feature of language that I can truthfully and, more importantly, meaningfully utter the sentence 'Amy believes Fermat's Last Theorem to be false', even though it is impossible for Fermat's Last Theorem to be false. And there are many more of such examples. If we both know that Superman is Clark Kent and I want to point out to you how ridiculous it is that Lois does not, I might, truthfully, say to you 'Lois believes that Clark Kent can and cannot fly'. Or, if we both believe Kripke in that unicorns necessarily do not exist, I might still, truthfully, say to you 'Eva seeks a unicorn'. So, we want our semantic theory *not* to assign the null-extension to the embedded sentences, as this would imply that they are meaningless. Extending the sentiment expressed by the arguments from Ball (2016) and Schwarz (2016), it seems that the fact that impossible worlds postulates help model these features of language, is enough to 'vindicate their use' in our semantic model. Or, as Kit Fine (2013) says:

> [W]e wish to state semantical clause over an extended space of possible and impossible states *without regards to how the extension was made* (p. 5, emphasis added)

3.2.2 Objection: Intuitionism and Non-Standard Semantics

Before we conclude this chapter, let us turn to a possible objection one might make at this point.[10] More specifically, one might make an objection against the example used to pump intuitions for semantic agnosticism. The example concerned the mathematician and the philosopher of mathematics and I claimed, in the example, that the mathematician would not let her work be influenced by the findings of the philosopher on the ontology of numbers. Therefore, I concluded, ontology and mathematics are two orthogonal issues. Here is where one might object and the objection might run along the following lines:

> *Objection.* Brouwer held a certain view of mathematical objects and based his view of mathematics on a particular philosophy of mind (cf.

based on the observable features of language use. Quine goes so far as to reject notions such as 'meaning' and 'intensions' as "creatures of darkness" (Quine, 1956, p. 180). I will not go so far, but it is interesting to see some similarities between Quine's behaviouristic approach and these recent accounts of semantics.

[10] Thanks to Franz Berto for raising this objection and helping me get clear on how to respond to it.

> Van Atten 2015). He believed mathematical entities to be activities in the mind or mental constructions. His view of mathematical objects had great consequences for his mathematics and it led to a rejection of (large parts of) classical mathematics, while allowing for a form of constructive mathematics. Thus, the objector may conclude, it is not true that the ontology of mathematical entities does not affect the mathematicians work. Hence, the objector may conclude, the analogy at the beginning of the chapter with semantics, fails.

This is an argument against the analogy between mathematics and semantics in that in mathematics, it turns out, one's ontology *does* affect one's semantics.

To avoid this argument and restore the analogy, we have to discuss a subtlety that we have not discussed yet. Namely, the distinction between 'standard' mathematics and, what we will call, 'non-standard' mathematics. Standard mathematics is the textbook mathematics that you will find in the introductory textbooks on the topic and that is taught at high-school or early stages of university. Non-standard mathematics is mathematics that is non-classical (e.g., intuitionistic mathematics).

What the objection shows, is that the arguments made in this chapter pertain only to *standard* mathematics. That is, standard mathematics is compatible with many ontologies and the mathematician can remain agnostic about the nature of mathematical objects. However, when doing non-standard mathematics, the ontology of mathematical objects might become of importance, as we see in the example of Brouwer.

As Van Atten (2015) notes, "Brouwer was prepared to follow his philosophy of mind to its *ultimate* conclusions [...]. In thus granting philosophy priority over *traditional* mathematics, he showed himself a revisionist" (emphases added). It thus does not seem strange to classify his mathematics as non-traditional, or non-standard.[11]

With this distinction in place, the analogy remains. For standard semantics (cf. Chierchia & McConnell-Ginet 1990; Heim & Kratzer 1998) the semanticist can remain agnostic about the nature of (possible) worlds and keep using her world postulates freely. However, also in semantics, there are non-standard semantics (cf. Lewis 1968, 1980), where ontology and semantics *do* interact. For example, Lewis (1968) has a semantics on which names are non-classically rigid,[12] deviating from the post-Kripkean (1980) picture that most standard semantics adhere to. Furthermore, Yagisawa's (1988) extended genuine modal realism (which is an extension of Lewis' account) does affect the semantics in that it allows contradictions at impossible worlds to spill over to contradictions at the actual world.

Chierchia & McConnell-Ginet (1990), who present a standard semantics, put the point very explicitly, when they say the following:[13]

[11] One might argue, that Brouwer's non-standard mathematics arose *because* of his ontology of mathematical objects. I do not want take sides on what caused what; what matters here is the distinction between standard and non-standard mathematics.

[12] I call it 'non-classically rigid', for one might argue that his counterpart theory captures some aspects of rigidity. I mean to indicate that it differs from the rigidity that Kripke (1980) argues for.

[13] Note that this quote is in itself also an argument for my general point presented above.

> The formal apparatus of possible worlds, [. . .], was introduced in Kripke (1959) as a tool for investigating the semantic properties of certain formal systems. There has since been, and continues to be, much controversy in the philosophical literature over what assumptions that apparatus requires. In accepting Lewis's point we do not deny that possible worlds might raise deep metaphysical issues, but we think that the formal apparatus can be adopted *without resolving these issues*, just as we can successfully use the notion of an individual in set theory and logic without resolving all the thorny problems it raises, for example, the mysteries of criteria for identifying individuals (p. 207, emphasis added)

Distinguishing between standard and non-standard semantics or mathematics thus restores the analogy. In both cases, the mathematician/semanticist does not need to care about the nature of her objects of study when she works in the standard framework. However, when working in non-standard frameworks, the ontology might influence the semantics (e.g., Brouwer's intuitionism and Yagisawa's extended genuine modal realism).

3.3 Semantic Agnosticism in the Fictionalist Landscape

Earlier, I said that I hoped that the exposition of Yablo's (2001) figuralism and Divers' (2006) agnosticism helped to delineate my own account. Now that the presentation of my account and the arguments in favour of it are complete, let us return to where exactly the account presented here fits in the philosophical landscape.

What should be clear by now, is that the account presented above is a meta-ontological account and not an ontological account (as it does not make any claims as to what really belongs to the furniture of our world). Secondly, what also should be clear, is that it is not a Quinean meta-ontology, as translations of sentences into canonical notation do no longer entail ontological commitments of such sentences, at least for semantic theories. I also already noted that there is, at least, one serious difference between the account presented here and standard modal fictionalism. Where modal fictionalists claim that talk of worlds is literally false and that worlds do not exist; on the account above one speaks truly when one talks of worlds postulates (though, she may not be talking about the metaphysical worlds) and we are truly agnostic about the existence of metaphysical worlds (not agnostic in Divers' sense).

Yet, there are also some similarities between my account and fictionalist accounts. For example, remember that fictionalists use talk of undesirable entities as they present useful tools or representational aids. This is exactly the argument I made for the use of world postulates. And as Yablo (2001), who distinguishes between representational aids and things-represented, I argue that the semanticist uses worlds as representational aids *qua* semanticist, yet that she may also use talk of worlds as things-represented *qua* metaphysician. This is very similar, I believe, to Yablo's distinction between *engaged* and *disengaged* talk. Remember that Yablo argued that when talking to philosophers one might truthfully want to say that there are no numbers (disengaged discourse), however, when talking to a mathematician, one might truthfully want to say that the number of even primes is one (engaged discourse). This is paralleled on my account. this would be accommodate by arguing that in the disengaged discourse, the mathematician is arguing that there are no numbers *qua* metaphysician, whereas in engaged discourse, she is arguing that there is an even prime *qua* mathematician.

This seems, I believe, to be very similar to what Yablo (2010) calls 'Subject-Matter-Ism':[14]

> *Subject-Matter-Ism*
>
> The asserted content when uttering φ is the part of $|\varphi|$ "that IS about the subject matter under discussion" (2010, p. 5, original emphasis)

In the first case, the subject matter is worlds (and their ontological status), whereas in the latter case it is language (and the use of certain theoretical postulates to model it). The discussions with regards to the two are orthogonal. So it seems that the view that I defend here might be something similar to what Yablo has in mind with his Subject-Matter-Ism.

3.3.1 *Semantic* Instrumentalism

I want to briefly turn to the notion of instrumentalism, which has been mentioned above. Instrumentalism is contrasted with realism in philosophy of science, either with respect to objects or with respect to laws. The former is most related to the issues discussed above.

For example, Kuhn is often mentioned as an instrumentalist in that he "thinks that he can set aside the issues of objective truth and real theory-independent existence" (Curd et al., 2013, p. 204). In general, as mentioned above, instrumentalists allow for unobservable objects "simply as instruments for the prediction of observable phenomena", while simultaneously holding that "unobservable things have no literal meaning at all" (Chakravartty, 2015). This seems similar to the fictionalist's position with regards to unwelcome ontological entities: they are useful for some objective or other, but that they are literally non-existent.

However, my point is precisely that there may very well be a matter of the fact what things there are in the world and whether these includes possible worlds, impossible worlds, or not, but the semanticist should not be bothered by this and use whatever 'instruments' she needs to model interesting feature of natural language use. Note that this characterises *semantic* instrumentalist. Limiting this to *semantic* instrumentalism, the semanticist may, while doing metaphysics, research the matters of the fact concerning the ontology of worlds, unobservable objects, etcetera.

3.3.2 The Parity Thesis

The parity thesis was already hinted at earlier, when providing arguments for believing in impossible worlds. There, it was said that most philosophers who believe

[14] Unfortunately, Subject-Matter-Ism is not been worked out 'officially', as Yablo (2010) himself points out, so it is hard to really judge the comparison.

that there are impossible worlds do so because there is no 'principled reason' to believe in possible worlds, while not in impossible worlds. The parity thesis extends this line of thought to; whatever it is that possible worlds are, impossible worlds are of the same kind.

According to Berto (2010, 2013) the parity thesis was put forth first by Rescher & Brandom (1980), who mention that they want to remain agnostic about the ontology of (im)possible worlds, and that the only important thing is to treat possible and impossible worlds on a par with regards to their ontological status. Priest (1997) echoes this when he says that there is "absolutely no cogent (in particular, non-question-begging) reason to suppose that there is an *ontological* difference between merely possible and impossible worlds" (p. 581, original emphasis).[15]

Note that Priest explicitly states his endorsement of the parity thesis in ontological terms. In this brief section, we will consider how the account presented above (where the use of semantic world postulates has no impact on one's ontology) affects the parity thesis. In particular, we will consider two possible interpretations of the parity thesis. One, that seems prevalent in the literature, taking it as a metaphysical thesis and one, more liberal, taking it as a general thesis concerning 'worlds' (whatever they may be).

With Priest, who endorses a metaphysical version of the parity thesis, most of the literature concerning the parity thesis seems to take it as a metaphysical thesis. For example, Berto, who argues against the parity thesis, concludes that "possible and impossible worlds, *metaphysically speaking*, are *not* of a kind" (2010, p. 475, first emphasis added). However, in this chapter, we spend quite some time arguing that the world postulates used by the semanticist need not necessarily be the worlds the metaphysician is concerned with. In particular, we argued that the semanticist's use of world postulates in her model does not commit her to accepting worlds into her ontology. So, it seems that if the parity thesis is an explicitly metaphysical thesis, it need not concern us here.

Yet, there might be an interpretation of the parity thesis on which it does relate to the account sketched above. Namely, we could also interpret the parity thesis more broadly and include the 'world postulates' with regards to it. If this is indeed the case, then there are two possible responses to it.

First of all, we want to emphasize (again) that the world postulates used in a semantic model *need not be* the worlds that the metaphysician concerns herself with. In this case, we reject (a version) of the parity thesis. Note that this rejection is not a rejection of the claim that possible and impossible worlds are alike, but a rejection of the claim that worlds and world postulates are of the same kind. The former are part of the furniture of reality (if they exists), whereas the latter are merely theoretical postulates used to represent interesting facts about language.

[15] However, note that not all agree with the parity thesis. In particular, Berto (2010) argues that the parity thesis in untenable. That is, Berto proceeds to discuss a range of ontological accounts of possible worlds (e.g., genuine modal realism, linguistic ersatzism) and argues that each of these accounts runs in to undesirable consequences when extended to impossible worlds in line with the parity thesis. See section IV of Berto's (2010) paper "Impossible worlds and propositions: against the parity thesis".

Secondly, if we only focus on the world postulates, we accept some version of the parity thesis. For consider again the arguments we made for the acceptances of impossible world postulates. The argument was one of utility, however, by the same argument we accept merely possible world postulates. It seems then that on the account above, possible and impossible worlds are of a kind: they are both theoretical postulates to help represent facts about language. This way the parity thesis is almost trivially satisfied, for it seems to me that on the account above there is no real difference between impossible and possible world postulates, only that the former is capable to represent impossibilities, yet this seems merely a terminological difference.

3.4 Conclusion

In this chapter I've argued for a particular meta-ontology for a semanticist to deploy with regards to the world postulates she uses in her model. The idea is that the world postulates she uses do not commit her to the acceptance of their metaphysical counterparts. I then proceeded to argue for this account from utility. The world postulates are extremely useful in representing interesting features of language.

There are, however, arguments that aim to show that impossible worlds are not as useful as they may seem. For example, Williamson (2007, forthcoming) argue against the utility of impossible worlds for counterpossibles and Bjerring (2013, 2014b) has argued that impossible worlds are not successful in dealing with issues concerning epistemic content. The remainder of this dissertation aims to engage with some of those arguments. In particular, we will focus on strengthening the case for impossible worlds semantics for counterpossibles. With this, we aim to strengthen the argument from utility (in that we show that impossible worlds *are* useful).

Chapter 4

Impossible Worlds Semantics

> [W]e can, for example, keep even classical logic while making adequate room for thinking about impossibilities
>
> Nolan (1997, p. 535)

In this, very brief, chapter, we will present a general, simple framework for an impossible worlds semantics. The semantical framework presented here is based on the work of Priest (2005), in his book *Towards Non-Being*. Note that, even though Priest already aims to provide an as simple as possible semantics, the version presented here is an even further simplification of Priest's work. I aim to present the most simple version of an impossible worlds semantics that is able to deal with the problems discussed above for possible worlds semantics. That is, everything remains as it is in the classical **S5** modal logic, except at impossible worlds, which are only accessible through intentional operators.[1]

There are more impossible worlds semantics (or circumstantialists' semantics) in the literature (cf. Rantala 1982; Edelberg 1994; Ripley 2012), however, we will not discuss or compare these. The aim of this chapter is merely to provide a very simple framework for impossible worlds semantics. If needed, one is free to modify the details whatever way she wants to.

We will first present some desiderata of an impossible worlds semantics, after which we will present our general framework for an impossible worlds semantics. After we have shown that this semantics satisfies the desiderata, we will continue to show that this semantics is able to deal with the problems for impossible worlds that were mentioned in Chapter 1.

4.1 An Impossible Worlds Semantics

On the semantics that Priest (2005) provides (or at least the material that we borrow from him), impossible worlds are almost completely anarchistic. That is, nothing

[1] Remember that, in line with saying 'intentional attitude' as opposed to 'propositional attitude', I use 'intentional operator', where others might have used 'intensional operator'.

is governed at the impossible worlds by any logic. Jago (2012, 2014) argues that this, at first, is what we want from impossible worlds so that we can, later, impose constraints on them to get out exactly the type of content that we want. Even though this may be what we want *at* impossible worlds, it seems that there are some desiderata that we want our semantics to satisfy too. Some are rather trivial, while others might be less so.

> *Classicality*: We want our semantics to behave purely classical at all *possible* worlds. That is, we want a characterisation of logical consequence and validity that is classical, i.e., it complies with classical logic for the usual extensional logical operators and with standard (normal, **S5**) modal logic for the modal operators.
>
> *Impossible Belief*: We want our semantics to allow for agents to have impossible beliefs. For example, John may seek a squared circle and the ancients may believe that Hesperus is not Phosphorus, and Amy may fear that Fermat's Last Theorem is false.
>
> *Consistent Actuality*: We do not want there to be any contradictions at the actual world. This is, as it is, a different formulation of classicality (or, at least, it follows from classicality). However, it is important to note, for we do not want sentences of the form 'A believes that φ and not φ' to entail that 'A believes that φ and A does not believe that φ' (i.e., we do not want contradictory beliefs to 'spill over' to actual contradictions).

With these desiderata in mind, we can now give the general framework of the semantics.

4.1.1 Impossible Worlds Semantics

We will consider a first-order language with identity. Our language will consist of constants, $a, b, c, \ldots$, variables, $x, y, z, \ldots$, and n-ary predicates, P^n, R^n. We also have intentional-operators, $v, u, \ldots$, that signify intentional attitude verbs such as 'fears', 'thinks', etc. Finally, we have the standard modal operators, '$\Box$' and '$\Diamond$', and two quantifiers, the existential quantifier, $\exists$, and the universal quantifier, $\forall$. We take our model, $\mathcal{M}$, to be an ordered quadruple; $\mathcal{M} = \langle W, P, D, \mathcal{J} \rangle$.

W is a non-empty set of all worlds
P is the set of all possible worlds, i.e. $P \subset W$
D is a non-empty set of objects, i.e., the domain
$\mathcal{J}$ is the interpretation function, and

We define the set of all impossible worlds, I, as follows: $W - P$, thus $I \subseteq W$. So, I and P are exclusive and exhaustive. Thus, $I \cup P = W$ and $I \cap P = \emptyset$. Furthermore, $\mathcal{J}$ assigns to every non-logical symbol a denotation at *possible* worlds. So,

if c is a constant, then $\mathcal{J}_w(c) \in D$
if P is an n-ary predicate and $w \in W$, then $\mathcal{J}_w(P, w) \subseteq D^n$
if v is an intentional operator, then $\mathcal{J}_w(v)$ is a function that maps each $d \in D$ to a binary relation on W. I will write $\mathcal{J}_w(v)(d)$ as $\mathcal{R}_v^d$

At impossible worlds, $\mathcal{J}$ assigns truth values to formulae directly and arbitrarily. So, at impossible worlds, w, for every φ, $\mathcal{J}_w(\varphi) \mapsto \{1, 0\}$.

Now, before we turn to the semantics itself, we need to specify what counts as a well-formed formulae of the language by giving a syntax.

Syntax:

If π is an n-ary predicate and $\alpha_1, \ldots, \alpha_n$ are terms (i.e., constants or variables), then $\pi(\alpha_1, \ldots, \alpha_n)$ is a formula

If φ and ψ are formulae, then $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$, $\varphi \rightarrow \psi$, $\exists x\varphi$, $\forall x\varphi$, $\Diamond\varphi$, and $\Box\varphi$ are formulae

If a is a term, v an intentional operator, and φ a formula, then a v's φ is a formula

Nothing else is a formula.

As at impossible worlds everything is arbitrary, we cannot define connectives in terms of each other. For, if we would define '$\vee$' in terms of '$\neg$' and '$\wedge$' we would run into trouble due to the impossible world that makes $\varphi \vee \psi$ true, while also making $\neg(\neg\varphi \wedge \neg\psi)$ false. Therefore, we explicate all connectives in the language. However, at possible worlds, the semantics of some connectives could be defined in terms of the semantics of others as usual.

We can now specify what it means for a sentence, φ, to be true relative to a world, w, the assignment-function, g, and a model, $\mathcal{M}$. I will write $[\![\varphi]\!]^{\mathcal{M},w,g} = 1$, where I will always suppress the mention of the model and, when context allows, also suppress the assignment-function. It is important to note that I use denotation brackets, $[\![\cdot]\!]$, to indicate the semantic value of *linguistic items.* So, in the case of constants, the semantic value is their denotation, which gets assigned through the interpretation function. Similarly for variables. Finally, if we take the semantic value of sentences to be their truth-value, the semantic value of a sentence is either '1' or '0'.

Semantics:

If a is a constant, then $[\![a]\!]^w = \mathcal{J}(a)$

If x is a variable, then $[\![x]\!]^{w,g} = g(x)$, where g is an assignment-function from variables to objects. Note that the interpretation of variables is world-independent (I only leave the world-parameter in for uniformity). So, variables are rigid in this semantics.

If P is an n-ary predicate and $\alpha_1, \ldots, \alpha_n$ are terms, then $[\![P(\alpha_1, \ldots, \alpha_n)]\!]^w = 1$ iff $([\![\alpha_1]\!]^w, \ldots, [\![\alpha_n]\!]^w) \in \mathcal{J}_w(P)$.

This is all still pretty straightforward. From here on out we will discriminate between evaluating a formula at a possible world and evaluating a formula at an impossible world. As mentioned above, the semantics of terms and formulae at impossible worlds goes directly through the interpretation and is completely arbitrary. So, we can give one single clause for the semantics at impossible worlds:[2]

If $w \in I$, then for any φ, $[\![\varphi]\!]^w = \mathcal{J}_w(\varphi)$

[2] This arbitrariness is similar to that in the semantics of Nolan (1997), given in his Appendix.

We will now turn to the rest of the semantics, that is, the recursive semantics for formulae evaluated at possible worlds.

If φ is a formula and $w \in P$ then $[\![\neg\varphi]\!]^w = 1$ iff $[\![\varphi]\!]^w = 0$

If φ and ψ are formulae and $w \in P$ then $[\![\varphi \wedge \psi]\!]^w = 1$ iff $[\![\varphi]\!]^w = 1$ and $[\![\psi]\!]^w = 1$

If φ and ψ are formulae and $w \in P$ then $[\![\varphi \vee \psi]\!]^w = 1$ iff $[\![\varphi]\!]^w = 1$ or $[\![\psi]\!]^w = 1$

If φ and ψ are formulae and $w \in P$ then $[\![\varphi \rightarrow \psi]\!]^w = 1$ iff $[\![\varphi]\!]^w = 0$ or $[\![\psi]\!]^w = 1$

Note that the semantics of the connectives at possible worlds is classical. So, for example, for possible worlds, '$\rightarrow$' could have been defined in terms of negation and conjunction. However, due to the fact that there are impossible worlds, we need to have all connectives explicitly in the language, for there might be an impossible world where $\mathcal{J}_w(\varphi \rightarrow \psi) \neq \mathcal{J}_w(\neg(\varphi \wedge \neg\psi))$. Thus, at impossible worlds the definition in terms of other connectives breaks down. So, we leave every connective explicitly in the language, even though the semantics at possible worlds might be defined in terms of each other. (The same goes for the quantifiers and the modal operators.)

Now, I will assume, with Priest, that for the alethic modal operators there is no accessibility relation and the '$\Box$' and '$\Diamond$'-operator get defined, at possible worlds, as true at all or true at some possible world, respectively. This gives us an **S5** modality for the possible worlds portion of the semantics. (Anyone who has different feelings about alethic modalities should change this relation accordingly. Nothing hinges on assuming **S5**, this is merely for simplicity.) This means that we could simply say, as Priest does, that a formula holds necessarily if the formula holds in all worlds.

If φ is a formula and $w \in P$ then $[\![\Box\varphi]\!]^w = 1$ iff for all possible worlds $w' \in P$, $[\![\varphi]\!]^{w'} = 1$

If φ is a formula and $w \in P$ then $[\![\Diamond\varphi]\!]^w = 1$ iff for some possible world $w' \in P$, $[\![\varphi]\!]^{w'} = 1$

For the semantics for the quantifiers, we will use the notation '$g'[x]g$' to indicate an assignment function, g', that differs *at most* from g in its assignment to x.

If φ is a formula and x is a variable and $w \in P$ then $[\![\forall x\varphi]\!]^{w,g} = 1$ iff for all g' such that $g'[x]g$, $[\![\varphi]\!]^{w,g'} = 1$

If φ is a formula and x is a variable and $w \in P$ then $[\![\exists x\varphi]\!]^{w,g} = 1$ iff for some g' such that $g'[x]g$, $[\![\varphi]\!]^{w,g'} = 1$

Note that the relation, $\mathcal{R}$, that is used in the semantics for intentional verbs is specific to the object holding the attitude (i.e., the denotation of the term of the attitude ascription) and specific to the type of relation, v. We might want to impose specific conditions on the accessibility relation when the relation concerns, for example, knowledge (for example, we might want knowledge to be factive, and thus the relation to be reflexive), while we might want to impose other conditions on the accessibility relation for other relations (for example, we might not want belief to be factive). I will not go in the details of these specifications here, but one can look at Priest (2005, Ch. 1) for some ideas.

If α is a term, v an intentional operator, and φ a formula and $w \in P$ then $[\![\alpha\ v's\ \varphi]\!]^w = 1$ iff for all $w' \in W$ such that $w\mathcal{R}_v^{[\![\alpha]\!]}w'$, $[\![\varphi]\!]^{w'} = 1$

Finally, the semantics for identity:

If α and β are terms and $w \in P$ then $[\![\alpha = \beta]\!]^w = 1$ iff $[\![\alpha]\!]^w = [\![\beta]\!]^w$

Given the rigidity of variables and constants on this semantics, objects that are identical are so in all possible worlds. However, given that, at impossible worlds, the assignment of truth-values goes arbitrarily, we can have an impossible world where $a = b$ is true and Pa is true, but where Pb is false (or, where $\neg Pb$ is true). So, even though, in every possible world where Hesperus is the brightest object in the evening sky, Phosphorus is also the brightest object in the evening sky, I can believe the former without believing the latter, for there is an impossible world where Pb is not true (or $\neg Pb$ is true), while $a = b$ and Pa are true.

With the semantics in place, we now need to define the notions of logical validity and logical entailment. Again, we follow Priest (2005) in defining logical validity and logical entailment as truth (preservation) at all *possible* worlds. As Nolan (1997) says, it is "[b]etter, [...], to only worry about *possibilities* when considering what notion of logical consequence our logic should capture" (p. 548, emphasis added).

Entailment: $\Psi \vDash \varphi$ iff for every model and every assignment of variables, and every *possible* world where all members of Ψ are true, φ is true.

We can then specify validity as a special case of entailment, namely entailment from an empty set of premises. We will define validity as follows:

Validity: $\vDash \varphi$ iff for every model, every assignment of variables, and every *possible* world, φ is true.

4.2 Impossible Worlds Semantics in Action

With the general framework for the semantics in place, we can now turn to the question whether this semantics satisfies the desiderata specified above and how it solves some of the issues we discussed for possible worlds semantics.

4.2.1 Satisfying the Desiderata

Classicality. Given the semantic clauses for the connectives at possible worlds and the lack of accessibility relation on the alethic modal operators in combination with the fact that entailment is defined as 'truth-preservation at *possible worlds*' and validity as 'truth at all *possible* worlds', it is quite easy to see that these all behave classically. By definition, impossible worlds only make their mark inside the scope of an intentional operator. Hence, this semantics, as with Priest's (2005) semantics, provides the 'standard' semantics when sentences are evaluated at possible worlds.

Impossible Belief. We want our semantics to allow for reports of agents that hold beliefs that are impossible. As, for example, the ancients believing that Hesperus is not Phosphorus or the fact that one might believe a conjunction without believing any individual conjunct (or the other way around, e.g., the preface paradox), or believe of one object that it does and does not have a property (e.g., Superman can fly and Clark Kent cannot fly).

We will show that it is possible to have each of these beliefs in our model by constructing a model on which it is true. First, the belief of the ancients that Hesperus is not Phosphorus (which, if we believe Kripke, is a necessary falsehood, i.e., an impossibility). Consider a belief-relation for an ancient such that $\mathcal{R}_v^{[\![a]\!]^w} = \{\langle w_1, w_2\rangle\}$, where w_1 is the actual world and w_2 an impossible world. Then, let 'H = P' be true at w_1 and 'H $\neq$ P' be true at w_2. Then we get that

$$[\![\text{H} \neq \text{P}]\!]^{w_1} = 0, \text{ while} \qquad [\![a \text{ believes that H} \neq \text{P}]\!]^{w_1} = 1$$

Now consider a version of the preface paradox, where an agents believes that, on a whole, there is at least one false sentence in her book (i.e., $\neg a\ v's\ (\varphi_1 \wedge \cdots \wedge \varphi_n)$). However, she has checked each sentence individually and believes of each sentence that it is true (i.e., $a\ v's\ \varphi_i$, for each $i < n$). Consider a belief-relation for this agent such that $\mathcal{R} = \{\langle w_1, w_2\rangle\}$, where w_1 is the actual world and w_2 an impossible world. Now, let $\varphi_1 \wedge \cdots \wedge \varphi_n$ be false at w_2 and φ_i be true at w_2 for each $i < n$. Then we get that:

$$[\![a \text{ believes that } \varphi_1 \wedge \cdots \wedge \varphi_n]\!]^{w_1} = 0, \text{ while}$$
$$[\![a \text{ believes that } \varphi_i]\!]^{w_1} = 1, \text{ for each } i < n.$$

Finally, the contradictory belief-ascription. That is, say Lois does not know that Clark Kent is Superman and believes of the latter that he can fly, while she does not believe that of the former. We do both know that Clark Kent is Superman and I want to let you know that, unbeknownst to her, Lois holds a contradictory belief. So I say to you 'Lois believes that Superman can and Lois believes that he cannot fly'. Consider a belief-relation for Lois such that $\mathcal{R}_{\text{believe}}^{\text{Lois}} = \{\langle w_1, w_2\rangle\}$, where w_1 is the actual world and w_2 an impossible world. Let 'φ' represent 'Superman can fly' and let φ and $\neg\varphi$ be true at w_2. Then, we get that:

$$[\![\text{Lois believes that } \varphi]\!]^{w_1} = 1 \text{ and } [\![\text{Lois believes that } \neg\varphi]\!]^{w_1} = 1$$

Consistent Actuality. We have shown above that it is possible to report contradictory beliefs in our model, however, we do not want that holding such beliefs lead to true contradictions at possible worlds. Specifically, we do not want that both φ and $\neg\varphi$ are true at any $w \in P$. We will prove this by induction on the complexity of φ. Note that this will be fairly trivial for all the regular connectives, for we have shown above that these are classical. I will therefore not do it for all the connectives, but only for a few. The real trouble is the case where we access impossible worlds, i.e., the case where φ is of the form $a\ v's\ \psi$.

> *Base case:* Take an arbitrary $w \in P$ and let φ be an atomic sentence. Then, by definition, $\neg\varphi$ can only be true if φ is false. So, if φ is an atomic sentence, φ and $\neg\varphi$ cannot both be true at w.

Negation: Take an arbitrary $w \in P$ and let φ at w be of the form $\neg\psi$, where ψ and $\neg\psi$ cannot both be true (induction hypothesis). Then, φ can only be true, when ψ is false and for $\neg\varphi$ to be true, by definition of '$\neg$' at possible worlds, ψ has to be true. Hence, φ and $\neg\varphi$ cannot both be true at w.

We can check by the truth-conditions of the other connectives, that φ and its negation cannot both be true at a possible world if φ is of the form: $\psi \vee \chi$, $\psi \wedge \chi$, $\psi \rightarrow \chi$, $\forall x\psi$, $\exists x\psi$, $\Box\psi$, $\Diamond\psi$. Let us now turn to the most important step.

Intentional operator: Take an arbitrary a, v, and $w \in P$ and let φ at w be of the form $a\ v's\ \psi$. If $a\ v's\ \psi$ is true, then all v-accessible-worlds, w', for a are such that ψ is true at w'. Yet, if $\neg a\ v's\ \psi$ is true, then there exists a v-accessible-world, w'', for a such that ψ is not true at w''. But, it cannot be the case that there exists a world in W such that a sentence is both true and not true at that world.[3] Hence, these two can never be true simultaneously.

This shows that, for any φ, φ and its negation cannot both be true at a possible world. □

Informally, we can also argue for why contradictory beliefs never lead to contradictions at a possible world. For, consider why we introduced the framework as it is. We wanted to model beliefs without, necessarily, modelling beliefs that logically follow from them. For example, we want to be able to say that $a\ v's\ (\varphi \wedge \psi)$, while it might not be the case that $a\ v's\ \varphi$ or that $a\ v's\ \psi$. Or, similarly, we want to be able to say that $a\ v's\ ((\varphi \rightarrow \psi) \wedge \varphi)$, while it might not be the case that a believes ψ. And this is what the above framework does.

As it seems that there are counterexamples to beliefs that 'should' follow from believing certain sentences containing a connective, it seems only natural that we also have that $a\ v's\ \neg\varphi$, while it might not be the case that $\neg(a\ v's\ \varphi)$.

4.2.2 Solving the problems of Possible Worlds Semantics

We will now briefly go over some of the problems for possible worlds semantics that we have discussed in Chapter 1. Note that most of the solutions follow pretty trivially.

Let us first look at logical omniscience. Take again sentence (1.1), repeated below:

(1.1) There do not exist three positive integers a, b, and c, such that $a^n + b^n = c^n$, for any integer value of n strictly greater than 2

As we saw, (1.1) is a mathematical truth and thus true at all worlds in P. Remember that logical omniscience can be characterised as the following inference: if $\vDash \varphi$ then $\vDash a\ v's\ \varphi$. Now, let φ be (1.1). We can, given the semantics given above, provide a counter-model against logical omniscience;

[3] Note that our impossible worlds can make a sentence and its negation both true, but it cannot be the case that one sentence gets assigned two truth-values.

> Let W be $\{w_1, w_2\}$, where $w_1 \in P$ and $w_2 \in I$. Let $\mathcal{R} = \{\langle w_1, w_2 \rangle\}$, and let φ be true at w_1 and false at w_2.
>
> Then, for all $w \in P$, $w \vDash \varphi$. Hence, by definition of validity, $\vDash \varphi$.
>
> But, at w_1, it is not the case that for all w', such that $w_1 \mathcal{R} w'$, $w' \vDash \varphi$, for w_2 is such that $w_1 \mathcal{R} w_2$ and $w_2 \nvDash \varphi$. Thus, $w_1 \nvDash a\ v's\ \varphi$ (by definition of intentional operators) and hence, by definition of validity, $\nvDash a\ v's\ \varphi$ □

(See Priest 2005, p. 23 for some related proofs.)

Secondly, attitude ascriptions and Frege's puzzle. This we have already seen as the semantics above satisfies the *Impossible Belief* desideratum. The solution for inconsistent fictions (mentioned in Chapter 1) is solved in the same way as Frege's puzzle is solved. Namely, we allow the fiction-operator to access all the worlds, so also the impossible ones. Then, it can be the case that there is a $w' \in I$ such that it makes true the impossible fiction.

4.2.3 Objection: Intentional Inferences

One may object that the semantics presented here is not strong enough. In the sense that, as it is, the semantics allows for *no* inferences concerning attitude reports whatsoever. For example, on the semantics presented here, it does not follow from 'Mary believes that the rose is red' that 'Mary believes that something is red'. This might indeed seem very counter-intuitive. I think that there are two ways to respond to this objection.

First, one may argue that it *should* indeed be the case that we cannot make inferences inside the scope of (certain) intentional operators. That is, one may believe that for every inference within the scope of an intentional operator, you can think of a potential counterexample (even existential generalisation). Such counterexamples will probably involve altered mental states or very extreme situations or illusions. However, this seems to be a very radical position to take.

On the other hand, one might agree that it is *too* counter-intuitive that the semantics are so arbitrary. She might then retort that the semantics presented here is merely a starting point and that we should look at how to improve upon the system in order to allow for the inferences that we want, while maintaining the lack of logical omniscience.[4] In this case, she would explicate that the semantics here is only meant as a starting point, a 'working hypothesis'. Similarly, Nolan remarks that his account of impossible worlds "is at least a good working hypothesis" (1997, p. 542).

Personally, I agree with Nolan that this lack of inferences that can be made within the scope of an intentional operator is not necessarily a drawback. As Nolan notes of his own system:

> Not many interesting theorems or inference patterns emerge from this system, but I do not think that this is a drawback, but is a reflection

[4] A place to start would be Jago (2009, 2014), who introduced an ordering of worlds to deal with just such problems.

> of how ill-behaved impossible worlds are, [. . .]. However, while this may limit the algorithmic usefulness of the theory to an extent, the system modeled [sic.] may still be powerful enough to carry on our hypothetical consideration of impossibilities. (1997, p. 567)

However, I do not want to take the radical stance of arguing that this is how it *should* be. Remember that in the previous chapter I argued that an aim of semantics is to capture interesting features of language. It is indeed a feature of language that if one ascribes a belief of a rose being red to someone, the addressee can infer that that someone believes there to be something that is red. In line with the instrumentalist view of the previous chapter, I believe that *if* we can improve the semantics presented here in order to gain better results, this should definitely be done.

The semantics as presented here seems to be good starting point as it is a very simple semantic framework, hoping to serve as many as possible and allowing many subtleties to be filled in. It is thus not a weakness of the semantics that so little can be inferred, it is more a result of the fact that the semantics provided is aimed to be as general as possible. If we can polish this semantics to capture more interesting features of language, we should.

4.3 Conclusion

This was a very brief chapter in which we discussed a simple framework for impossible worlds semantics. The main result of this chapter is that our semantics does not allow for contradictory beliefs to ‘spill over’ to the actual world (or to any possible world for that matter).[5] Furthermore, the semantics provided in this chapter satisfy the desiderata set out at the beginning of this chapter. That is, this semantics allows for contradictory beliefs, is classical in everything except within the scope of an intentional operator, and preserves consistency at possible worlds.

In the next chapter, we will discuss counterpossibles. We will do this to the end of showing that impossible worlds semantics are indeed useful. If we succeed in this, this will strengthen the argument made in Chapter 3, for the use of impossible worlds in semantics.

[5] As opposed to, for example, the semantics that follows from extended modal realism.

Chapter 5

Counterlogicals: A Case for Impossible World Semantics

> Perhaps much of philosophy is vacuous, uninformative and fallacious. But if it is, it is not for systematic misuse of the counterfactual
>
> (Brogaard & Salerno, 2013, p. 644)

The previous chapter sketched a very simple impossible worlds semantics that seems to capture some of our initial intuitions about how such a framework should behave. In Chapter 3, I argued that semanticists should be allowed to use impossible world postulates in their theories without the ontological commitments, as these are simply two different things. I argued for this through arguments of utility and the alleged usefulness of impossible worlds in formal semantics. However, if we really want our semantics, and especially the impossible worlds therein, to be useful, it should be able to deal with *counterpossibles*. For, as Berto (2013) notes, counterpossibles are one of the main areas where impossible worlds are employed.

Counterfactuals, on the standard analysis, are evaluated for a truth-value by means of a *similarity ordering*. On such an ordering, worlds are ordered based on their similarity with the actual world, with respect to relevant features of the context (we will explicate this below). Even though the literature on impossible worlds and their applicability to counterpossibles is growing (cf. Nolan 1997; Vander Laan 2004; Goodman 2004; Brogaard & Salerno 2013; Bjerring 2014a; Williamson forthcoming), most of this literature focusses on the 'intuitive' and 'conceptual' analysis of counterpossibles. Almost no one has provided a similarity ordering for the impossible worlds in question. Even though I will not aim to provide the final solution, in this chapter we will apply an ordering of impossible worlds that is used to deal with other problems and apply it as a similarity ordering for counterpossibles. That is, we will take the ordering as it is and evaluate its applicability as a similarity ordering for counterpossibles.

This chapter is structured as follows. In the first section we will introduce the notion of counterpossibles and the problems for the Lewis/Stalnaker-analysis of these.[1] We will discuss the similarity ordering for counterfactuals and flag that this is a notoriously difficult notion. Secondly, we will turn to the problem of counterpossibles and explicate the often suggested *extended* Lewis/Stalnaker-analysis. The problem with this suggested analysis is that there has never been any similarity ordering proposed, which is where we aim to contribute. We will turn to an ordering suggested in a different field and evaluate it as a similarity ordering for counterpossibles. The ordering in question is that of Jago (2009, 2014) (henceforth: Jago-ordering) and is a response to problems of triviality in impossible worlds frameworks for epistemic agents (cf. Bjerring 2013, 2014b; Bjerring & Schwarz 2016). In the last section, we will evaluate the Jago-ordering as a similarity ordering for impossible worlds. We will do so by discussing its scope and comparing it with more informal accounts of counterpossibles such as those in Goodman (2004) and Bjerring (2014a). Finally, we conclude by suggesting some avenues for future research.

Before we dive in, note the following terminological remark. Throughout this chapter I talk about ‘counterfactuals’, ‘counterpossibles’, and ‘counterlogicals’. On most occasions, context will help clarify what class of sentences I mean. However, in general, when I talk of ‘counterfactuals’ I intent to talk about counterfactuals with a *possible* antecedent. When I talk about ‘counterfactuals in general’, I often intent any counterfactual, with any antecedent. ‘Counterlogicals’ indicate the counterfactuals with specifically *logically* impossible antecedents, whereas ‘counterpossibles’ indicate all counterfactuals with impossible antecedents.

5.1 Counterfactuals and Counterpossible Problems

Remember that we discussed the Lewis/Stalnaker-analysis of counterfactuals in the first chapter and some of the flaws that this analysis has, especially when impossible antecedents are concerned. Counterfactuals are conditional sentences often in the subjunctive mood and often with a (known) false antecedent. For the present purposes, we take it that counterfactuals are of the form ‘If φ were the case, then ψ would have been the case’—formally represented as ‘$\varphi \mathrel{\Box\!\!\to} \psi$’.

The standard analysis for such sentences originated with Stalnaker (1968) and Lewis (1973). This Lewis/Stalnaker-analysis, as it is often called, rests on the idea that we can ‘order’ worlds based on similarity. Before we turn to the semantics that they propose, I will briefly discuss two important differences between the analysis proposed by Stalnaker (1968) and by Lewis (1973). These differences are found in the restrictions put on the similarity ordering. For example, Stalnaker’s semantics satisfies the Limit Assumption and the Uniqueness Assumption, whereas Lewis’ semantics does not (see Edgington 1995; Bennett 2003; Goodman 2004; Sider 2010).[2]

[1] There might be different analyses of counterfactuals, where either one does not use a similarity ordering, or one where counterfactuals do not receive a truth-value at all (e.g., Veltman 2005). However, for this chapter I will follow most of the literature with the assumption that the Lewis/Stalnaker analysis is on the right track.

[2] See also the Appendix of Nolan (1997), who discusses these and some other constraints with respect to his semantics, which also includes impossible worlds.

1. Uniqueness Assumption. Stalnaker (1968) defends the uniqueness assumption, which states that there is exactly *one* world that is most similar to the actual world. He does so, because he wants to defend the Conditional Excluded Middle ($(\varphi \mathrel{\Box\!\!\rightarrow} \psi) \vee (\varphi \mathrel{\Box\!\!\rightarrow} \neg\psi)$). Lewis (1973) rejects this based on the following sentence pair:

   (5.1) If Bizet and Verdi were compatriots, then Bizet would be Italian

   (5.2) If Bizet and Verdi were compatriots, then Bizet would be French

   Lewis' objection has to do with his aim to also provide a semantics for '$\varphi \mathrel{\Diamond\!\!\rightarrow} \psi$', which, according to him does not work on Stalnaker's system. We will not go into the details of his argument, for these subtleties do not really matter when we turn to counterpossibles. One may either assume the Uniqueness Assumption or not. In the semantic here, we do so merely for simplicity.

2. Limit Assumption. Lewis (1973) notes the limit assumption, that is, that there can never be an infinite regress of more similar worlds. For example, given that I am 1.87m tall, which world where I am *not* is most similar? It seems that there is always a world even more similar *ad infinitum*. Stalnaker's semantics satisfies the limit assumption, whereas Lewis' does not. For simplicity we will assume the limit assumption, but nothing hinges on this.

The semantics that will be presented below, satisfies the limit assumption and the uniqueness assumption, however, this is merely for simplicity. If one feels that the semantics of counterfactuals should not satisfy these assumptions, she is free to change the semantics accordingly. Let us now turn to the Lewis/Stalnaker-analysis of counterfactuals. On their analysis, a counterfactual '$\varphi \mathrel{\Box\!\!\rightarrow} \psi$' is true if and only if in the most similar world (or worlds) where the antecedent is true, the consequent is also true.[3] In other words, $\varphi \mathrel{\Box\!\!\rightarrow} \psi$ is true in the actual world if and only if ψ is true in the most similar world in which φ is the case (Sider, 2010, p. 203). Consider the famous example of Lewis (1973, p. 1):

(5.3) If kangaroos had no tails, they would topple over

This sentence is, intuitively, true (according to most of the literature). On the similarity-analysis, this indeed comes out correct. For, consider two worlds: one where kangaroos are exactly as they are in the actual world, but without tails and a different world where kangaroos also have wings and crutches, but lack tails. In the first world, kangaroos would topple over if they lacked tails, whereas in the second world they would not (as the crutch or wings would prevent that). As the former scenario differs less from the actual world than the latter, (5.3) is true at the actual world. Stalnaker elegantly puts this as follows, "truth conditions" require "that the world selected *differ minimally* from the actual world", which means that "there are no differences between the actual world and the selected world except those that are required [...] by the antecedent" (Stalnaker, 1968, p. 104, original emphasis).

Formally, we can capture this, assuming a similarity ordering '$\preceq_@$' (where '$w_@$ is the actual world), as follows:

[3] I will use the notions 'most similar world' and 'closest world' interchangeably, both referring to the same similarity ordering.

$[\![\varphi \mathrel{\Box\!\!\rightarrow} \psi]\!]^{w_@} = 1$ iff for any $w \in P$, if $[\![\varphi]\!]^{w} = 1$ and for any $w' \in P$ such that $[\![\varphi]\!]^{w'} = 1$, $w \preceq_{w_@} w'$, then $[\![\psi]\!]^{w} = 1$

Ordering Possible Worlds

It is important to note (also for what is to come) some characteristics of the similarity ordering for counterfactuals. Consider the following counterfactual:[4]

(5.4) If Edinburgh were in Italy, it would rain less in Edinburgh

We can imagine two possible scenarios that are relevant for (5.4). First, imagine the borders of Italy, extending all the way to include the Southern part of Scotland. This scenario is one, in which the antecedent of the counterfactual is true, yet, given that the geological location of Edinburgh is exactly the same, the consequent seems false. On the other hand, we can also consider a scenario where Edinburgh, with all its pubs, its Castle, and its inhabitants, is now located within the original Italian borders. In that case, the antecedent would also be true and, this time, the consequent would be as well. Arguably, it depends on the context, which of these two scenarios is used in the evaluation of (5.4). That is, which facts should remain fixed in the ordering is context-sensitive.

The example above aims to illustrate the context-sensitivity of the similarity for counterfactuals. When discussing counterpossibles, we will also resort to the context from time to time. This example shows that such a move is not a weakness of the account for counterpossibles, but inherent to counterfactuals in general.

Secondly, it is an open problem to provide a *uniform* ordering (and account) for all counterfactuals. For example, it is not altogether clear what 'relevant facts' are and there are examples that aim to show that the Lewis-ordering makes the wrong predictions in certain cases (see for example Veltman's 2005 discussion of the Tichy-problems). Furthermore, Bennett (2003) argues that *counterlegals*, counterfactuals whose antecedent is contrary to the actual laws of nature (Bennett, 2003, § 87), are in need of a special treatment. Bennett concludes that for such counterfactuals (whose antecedent is still representable by *possible* worlds), we might need to turn away from the classical analysis involving worlds and propositions.

These examples aim to highlight the extreme complexity of a similarity ordering and the fact that for counterfactuals in general it is still an open issue whether it is possible to have a uniform similarity ordering. This is something we need to keep in mind when, later on, we turn to our proposed similarity ordering for impossible worlds.

5.1.1 The Problem of Counterpossibles

The Lewis/Stalnaker-analysis has, in some form or other, been the predominant analysis of counterfactuals. However, there is one problem for such accounts and

[4] Example adapted from Sider (2010, Ch. 8).

some argue that this problem calls for an extension of the analysis: including impossible worlds in the similarity ordering. Consider the following counterfactuals:[5]

(1.6) If Amy had squared a circle, Amy would be famous

(1.7) If Sarkozy had squared a circle, Amy would be famous

(5.5) If Intuitionistic Logic were the one true logic, then the Law of Excluded Middle would fail

(5.6) If Intuitionistic Logic were the one true logic, then the Law of Excluded Middle would be valid

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

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

Suppose that the antecedents of all the above counterfactuals are impossible. On the Lewis/Stalnaker-analysis, this means that all these sentences are (vacuously, trivially) true. However, intuitively, one of each pair of examples, (1.6-1.7), (5.5-5.6), and (5.7-5.8) has one sentence that is true and one that is false. This seems to be a serious flaw of the Lewis/Stalnaker-analysis, one that is echoed in the literature on counterpossibles (see also Nolan 1997; Goodman 2004):

> "The claim that counterpossibles must be trivially true [. . .] is an insistence on one way of measuring similarity, a failure to recognize that other ways [i.e., including impossible worlds] may be more appropriate on certain occasions"
> (Vander Laan, 2004, p. 263)

> "Counterpossibles are trivial on the standard [Lewis/Stalnaker] account. [. . .] That is one piece of evidence against the standard account and vacuism more generally" (Brogaard & Salerno, 2013, pp. 642-3)

> "Notoriously, however, [Lewis/Stalnaker] semantics deems all counterpossibles trivially or vacuously true. [. . .] But if at least some counterpossibles can be non-trivially false, and some non-trivially true, as the considerations above suggest, then the Lewis/Stalnaker semantics seems inadequate"
> (Bjerring, 2014a, pp. 328-9)

Many of these philosophers have argued for an extended Lewis/Stalnaker account for counterpossibles. That is, if we allow our similarity order to also range over impossible worlds, then we could account for the intuition that some counterpossibles are false. Intuitively, this seems quite straightforward. Consider a world where Hobbes would have squared the circle (secretly or not), it seems then that in that world the sick children in the mountains of South America could not have cared less. So, intuitively, (5.8) is true and (5.7) is false. However, given that the antecedent requires an impossible world, it might not be all that clear what impossible worlds are

[5] Examples (1.6) and (1.7) are from Ripley (2012); examples are (5.5) and (5.6) are found in Bjerring (2014a); and examples (5.7) and (5.8) are from Nolan (1997).

more similar than another. That is, it is not at all a trivial matter how impossible worlds should be ordered for similarity. Given that, as we noted above, ordering merely possible worlds is already a complex matter, it is to be expected that there are many phenomena with counterpossibles that are very context-sensitive and are equally hard to evaluate as it is for some counterfactuals. We suggest, however, that some progress can be made towards a, somewhat more formal, ordering of impossible worlds, which is also able to capture some of this context-sensitivity.

Arguably, Nolan (1997) is somewhat of a *locus classicus* on the subject of the use of impossible worlds in counterpossible semantics. However, Nolan, and many that followed him, "doesn't commit [himself] one way or another to what similar-in-relevant-respects amounts to" (Brogaard & Salerno, 2013, p. 651).[6] That is, most of the literature on impossible worlds in counterpossible semantics focusses on the conceptual, intuitive picture, however, almost nobody provides a definitive similarity ordering for impossible worlds. Remember that even for possible worlds, it is still an open problem to come up with a complete similarity-matrix, yet, for counterpossibles no such ordering has ever been proposed.

Note that there are those who argue against a non-vacuous account of counterpossibles (most notably, Williamson 2007, forthcoming). I take it that the examples above are sufficient to show that counterpossibles are, at least sometimes, non-trivially true or even false. So, I will not further engage in the discussion whether or not counterpossibles are non-trivially true or not. For more arguments in favour of non-trivial counterpossibles see Nolan (1997) and for some rejections of Williamson's (2007) arguments, see Brogaard & Salerno (2013).

Before we properly start this chapter, we need to note something that many have argued for in the context of counterpossibles. That is, many authors who defend a non-trivial account of counterpossibles have argued that for such a semantics *partial* impossible worlds are necessary. For example, Bjerring (2014a) has a very elaborate argument against complete impossible worlds for counterpossibles, and so does Vander Laan (2004, p. 265). The problem, they argue, is that if our semantics includes the notion of a 'false consequent', it will make the wrong predictions. Better, they argue, to have the consequent being 'non-true'.[7]

It seems that we will have to allow for partial impossible worlds, also because Jago uses partial worlds on his account (Jago, 2014, p. 203). However, as Restall

[6] Note, however, that Nolan does emphasise how important a clear grasp of the similarity-ordering is. He says that "[i]t is clear that if the notion of similarity-in-relevant-respects between worlds is a useful device for thinking about the truth-conditions of some class of conditionals, then we must have a better implicit grasp of the relevant respects to take as similar" (1997, p. 544).

[7] Vander Laan (2004, p. 265), for example, makes the distinction between the following two definitions:

> A counterfactual is true iff some world where both the antecedent and consequent are true, is more similar to the actual world than every world where the antecedent is true and the consequent is *false*.
>
> A counterfactual is true iff some world where both the antecedent and consequent are true, is more similar to the actual world than every world where the antecedent is true and the consequent is *not true*.

has pointed out to Nolan, we can captures this on the current semantics with only two truth-values:

> A world where φ would be assigned 'both' is, in [our] formulation, a world where φ and $\neg\varphi$ are both assigned 1, and a world where φ would be assigned 'neither' is, in [our] formulation, a world where φ and $\neg\varphi$ are both assigned 0 (Nolan, 1997, p. 562, fn. 27)

In this chapter we will aim to provide a starting point for a well-defined ordering for counterpossibles. We will do so by evaluating the applicability of an ordering of impossible worlds suggested by Jago (2009, 2014). As Jago orders the space of worlds very specifically, based on logical structure, we will start our evaluation by evaluating the applicability to counterlogicals: counterfactuals with a logically impossible antecedent. We will then move on to see if we could use the Jago-ordering for other types of counterpossibles as well.

Let us start with introducing the ordering as presented by Jago.

5.2 Triviality and Ordering Epistemic Space

To introduce the ordering of worlds that Jago (2009, 2014) suggests, we have to introduce the context in which it is originally presented. Jago presents his ordering as a solution to, what he calls, *the problem of bounded rationality* (see also Bjerring 2010, 2013; Jago 2013; Bjerring 2014b): a problem for accounts that use impossible worlds to model epistemic content. [8]

We will briefly discuss the problem of bounded rationality here, only as a set up for the ordering that Jago presents as a solution. We will then take this ordering and evaluate it as a candidate for a similarity ordering on impossible worlds.

The Problem of Bounded Rationality

In a nutshell, the problem of bounded rationality is that while "both the normative principles of rationality and the fact of non-omniscience seem non-negotiable", together "they take us into the realm of philosophical paradox" (Jago, 2014, p. 166). That is, agents seem to have some rationality, yet once we allow full-blown anarchic impossible worlds in epistemic space, the resulting epistemic space does not seem to be able to capture this. It becomes trivial. On the other side of the spectrum, agents do not seem to be logically omniscient, yet if we only allow for possible worlds, the resulting picture predicts that they are. As Jago (2014) puts it:

> This is the *problem of bounded rationality.* It is the conflict between normative principles of rationality and the fact that the agents with which we are concerned have limited cognitive resources. It is a problem specifically for assigning contents to attitudes of rational but cognitively bounded agents, such as ourselves (p. 165, original emphasis)

[8] Note that no such account has been argued for here. The account present here is merely a formal semantic framework that uses impossible worlds to model certain features of language use. I do not want to take a stance on the relation between formal semantics and 'content' (whatever it may be).

Jago suggests a solution that involves an ordering of the impossible worlds based on their logical structure. That is, Jago suggests structuring *all of epistemic space*, not the worlds themselves. We will focus on the general picture presented in Jago (2009), for a more detailed account and how it solves the problem of bounded rationality see also Jago (2014, Ch. 6 & 7).

5.2.1 Jago's Solution Through Ordering Epistemic Space

To properly discuss Jago's ordering, we first have to make a notational remark. Jago defines $|w|$ as the *truth set* of w, which is simply the set of all truths at w. For our purposes, we can think of the truth set of a world w to be the set of all sentences that get assigned '1' at that world. Jago then divides epistemic space in two, non-overlapping, sets of worlds: those consistent and closed under deduction and those that are inconsistent (and hence not closed under deduction), W^C and W^O respectively.

Now, consider a total order on the set of worlds, $\preceq$, such that $w \preceq w'$ implies that if we rule out w' as an epistemic possibility, then we should also rule out w.[9] (In terms that will be introduced later on, w is more obviously impossible than w'.)

Consistent worlds are the maximal elements with regards to this ordering. So, for a consistent world, w, and all worlds w', $w' \preceq w$. Minimal elements are those that contain an 'obvious' impossibility in their truth set. Therefore, Jago defines a set of all these 'obvious' impossibilities, x. If a world contains an obvious impossibility, then the intersection of the truth set of that world with x will not be empty. Thus, Jago defines the semantic value of x, $[\![\mathsf{x}]\!]$, as the set of such worlds—i.e., $[\![\mathsf{x}]\!] = \{w \mid \mathsf{x} \cap |w| \neq \varnothing\}$.[10]

Now that the maximal and the minimal elements of the set of worlds are ordered, we need to come up with a procedure to order two arbitrary worlds that are neither maximal, nor minimal. Jago's way of doing this is specifically tailored to the problem of bounded rationality and rational agents.[11] Accordingly, he argues that, just as we can expect agents to recognize certain obviously impossible worlds immediately (i.e., the content of $[\![\mathsf{x}]\!]$), we can *also* expect agents to have the ability to make some simple inferences. Note, as Jago (2009) stresses, that having the ability to apply some inference rules is *not* the same as having one's beliefs closed under that inference rule. Let us define the set of inferences an agent can apply as $\mathfrak{R}$.[12] The semantic value of $\mathfrak{R}$ is a binary relation on W. A pair of worlds, $\langle w, w' \rangle$, is a member of $[\![\mathfrak{R}]\!]$ iff there is an instance of a rule in $\mathfrak{R}$ such that the set of premises of the instance of rule are a subset of the truth set of w (i.e., $\{\varphi_1, \ldots, \varphi_n\} \subseteq |w|$), the

[9] I follow Jago (2009, p. 334, fn. 17) in taking '$\preceq$' to be a total order, for it seems that any two worlds can be compared. If this turns out to be wrong, the ordering can be changed accordingly.

[10] It is not entirely clear why Jago talks about 'the semantic value of x', for this does not seem to correspond to the use of 'semantic value' as it is been used throughout this dissertation. In this section on Jago, I just follow him in his terminology, though it has to be noted that this is not the same as my use of 'semantic value'.

[11] Jago (2009) takes to be only *one of possibly many ways* of doing it, however, in Jago 2014, he seems to believe more firmly that the way he presents there is *the* way to do it.

[12] Jago (2009) uses '$\mathcal{R}$', but, in order to avoid confusion with the accessibility relations used in the semantics of this dissertation, I will write '$\mathfrak{R}$'.

conclusion is a member of the truth set of w' but not of w, and the truth set of w' differs only from that of w in its inclusion of the conclusion in its truth set (i.e., $\psi \notin |w|$ and $|w'| = |w| \cup \{\psi\}$). Informally, this means that there is an instance of an inference rule in $\mathfrak{R}$ of which an application allows the agent to get from w to w' in one proper step. (Note that this is why Jago needs worlds that are *not* closed under deduction, for if all worlds were closed under deduction, then this characterisation of 'getting from w to w'' is void.)

Informally, we now want to capture is what the minimal number of applications of inferences rules of $[\![\mathfrak{R}]\!]$ is to get from an impossible world to a world whose truth-set contains a member of $[\![\mathsf{x}]\!]$—i.e., a world an agent can judge to be obviously impossible. In order to do so, we define a partial function from worlds to the set of natural numbers, where the number that a world is mapped to represents the number of applications of inference rules that are needed to get from that world to a world that represents an obvious impossibility. We can then use this function to order the worlds. A world that reaches an obviously impossible world with *less* applications of rules from $\mathfrak{R}$ is *more* obviously impossible than a world that needs more applications of rules from $\mathfrak{R}$.

So, we define a partial function, f, from impossible worlds to the set of natural numbers $\mathbb{N}$, such that $f(w) = n$ iff there exists a sequence of worlds, $w_0, w_1, \ldots, w_n$, such that $\langle w_i, w_{i+1}\rangle \in [\![\mathfrak{R}]\!]$ for each $i < n$; $w_0 = w$; and $w_n \in [\![\mathsf{x}]\!]$, but such that there does not exist a sequence of worlds $w_0, w_1, \ldots, w_m$, where $m < n$, with the same properties (Jago, 2009, p. 335). This captures the intuitive description from above.

Given all of this, we can now define a the order on W, '$\preceq$', as follows:[13]

For any worlds w and w', $w \preceq w'$ iff

- either w' is a consistent world (for then no impossibility whatsoever will be inferable from it),
- $w \in [\![\mathsf{x}]\!]$ (for then w is itself already an obvious impossibility), or
- $f(w) \leq f(w')$ (for then the number of applications of inference rules to get to an obvious impossibility is less for w than for w')

This is the simple version of Jago's solution to the problem of bounded rationality. This version is presented in Jago (2009), whereas a more detailed analysis is given in Jago (2014). As Jago (2014) is far more restrictive, we followed the simpler version.

Bjerring (2014b) presents an argument against the worlds-frameworks for modelling a notion of epistemic possibility that he centres around Jago's (2009) ordering. Note that Bjerring's argument is especially aimed at an account of epistemic content. He suggests that a worlds-account of epistemic content cannot account for the minimal level of rationality we can expect from an agent. Though this is indeed a

[13] Note that on this ordering, if you rule out a possible world as an epistemic possibility, you have to rule out *all* possible worlds. This seems a bit too extreme. Jago never mentions this, presumably because he is only concerned with ordering impossible worlds. As the account below does not concern the 'ruling out' of worlds, we will leave this aside.

very pressing, and interesting, problem for worlds-frameworks, this is not what is at stake here.[14]

We will now turn to a critical evaluation of the applicability of Jago's ordering to counterpossibles. In doing this, we will first look at the scope of the ordering. Then, in the next section, we will look at the relation between the ordering and some informal accounts of counterpossibles, and we will try to foresee some potential problems.

5.2.2 Scope of the Jago-ordering

Here, we will discuss the scope of counterpossibles the Jago-ordering might be applicable to. Remember, that Jago (2009, 2014) provides us with an ordering that is based on inference steps. That is, the ordering is based on the logical content of the worlds. In particular, it would not tell us how blatantly impossible it is for Amy to square a circle or for Kripke not been born from his own parents.[15] Consider the following two negated tautologies (i.e., impossibilities):

(5.9) $\neg((p \leftrightarrow (q \wedge \neg p)) \rightarrow \neg q)$

(5.10) $\neg(p \rightarrow p)$

It is easier to deduce an explicit impossibility from (5.10) than it is from (5.9) (the former almost immediately leads to $p \wedge \neg p$). So, (5.10) is more obviously impossible than (5.9) and, thus, is less similar to the actual world than (5.9). These are the types of sentences the Jago-ordering has something to say about.

So, imagine that there is a counterpossible, $\varphi \mathrel{\Box\!\!\rightarrow} \psi$, where ψ is such that (5.9) is derivable from it and $\neg\psi$ is such that (5.10) is derivable from it. Then, according to the Jago-ordering, a world where φ and ψ are true is more similar to the actual than a world where φ and $\neg\psi$ are true. Hence, $\varphi \mathrel{\Box\!\!\rightarrow} \psi$ would be true, whereas $\varphi \mathrel{\Box\!\!\rightarrow} \neg\psi$ would be false.

Of course, one might wonder whether there are any sentences in natural language that, intuitively, would translate into something as (5.9), but we will leave this aside for now and return to it in the conclusion of this chapter.

What is important to note, however, is that the ordering does not seem to be helpful for any of the examples used above, which were used to argue for counterpossibles. This is indeed true, which shows that there is still a lot to be done. Furthermore, Jago (2009, 2014) never claims to provide a *general* similarity ordering, but explicitly argues that he provides an ordering based on *logical structure*. It thus seems unfair to expect from the ordering that, *as it is*, it is able to make predictions concerning worlds that make true 'Amy squared a circle' and 'Kripke was born from different parents', for these are, *logically*, on a par.

[14] Bjerring seems to take issue with the fact that Jago's solution includes vague cut-off points for epistemically possible scenarios. As this is has nothing to do with the similarity ordering in itself, we will leave the objection aside. See Bjerring (2014b) and Jago (2014, Ch. 7.3) for detailed versions of the argument and Jago's response.

[15] Jago (2009) argues that something being a square *and* a circle is blatantly impossible, i.e., in the set $[\![\mathbf{x}]\!]$, however, one might disagree with him on this.

In line with this, we explicitly stated at the beginning of this chapter that in adapting the Jago-ordering we would only provide a framework for *counterlogicals*. And the order does seem to capture these, even though this might only be a small portion of the interesting counterpossibles. Of course, we do hope to find a complete ordering of impossible worlds, logical and non-logical. This would provide us with a semantics that can deal with counterfactuals such as (5.7) and (5.8), repeated below:

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

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

As it is, this cannot be done. However, given that, as we already saw, it does not seem possible to provide one, uniform, similarity ordering for *possible* worlds, it is likely that we neither get a uniform similarity ordering for all counterpossibles. We will briefly get back to this thought in the concluding section of this chapter.

Semantics for Counterlogicals

We will now turn to a very 'naive' clause for the truth-conditions of counterlogicals following the ordering of Jago, presented above. We will leave out the semantics for when the antecedent is possible, for then one can plug in her own favourite semantics for counterfactuals. Now, let a counterpossible be true if there is no world less obviously impossible in which the antecedent is true and the consequent false, than all the worlds where both the antecedent and the consequent are true. Or, conversely, all the worlds in which the antecedent is true and the consequent false (or, the negation of the consequent true) should be more obviously impossible than the worlds where the antecedent and the consequent are both true. Formally, this can be spelled out as follows:[16]

$[\![\varphi \mathrel{\Box\!\!\rightarrow} \psi]\!]^{w_@} = 1$ iff for any $w \in W$, if $[\![\varphi]\!]^{w} = 1$ and for any $w' \in P$ s.t. $[\![\varphi]\!]^{w'} = 1, w \preceq_{w_@} w'$, then $[\![\psi]\!]^{w} = 1$, where, '$\preceq_{w_@}$' is the Jago-ordering if φ is a counterlogical.[17]

Note that I have suppressed whether or not we are evaluating the counterfactual at possible worlds or impossible worlds. For simplicity the above clause is only when we evaluate counterfactuals at *possible* worlds. According to Bjerring (2014a), "...it is not so clear what it means to evaluate counterfactuals in arbitrary impossible worlds" (p. 330). However, in line with the semantics provided in Chapter 4, we would evaluate counterfactuals as atomic sentences, directly with the interpretation function. Remember that we said that:

If $w \in I$, then for any φ, $[\![\varphi]\!] = \mathcal{J}_w(\varphi)$

[16] In order for the definition to be more uniform with the semantics for counterfactuals, I have reversed the ordering relation. That is, $w \preceq w'$ means that w is *less* obviously impossible than w'.

[17] Note that Jago's ordering is absolute, thus not world sensitive. Hence, the world-subscript is strictly speaking not necessary.

So, also when φ is of the form '$\psi \boxright \chi$'.

One might worry that allowing impossible worlds to be accessed in the counterfactual semantics affects our aim to keep the semantics, in general, as classical as possible. However, this worry is unwarranted. For Nolan (1997) already notes "that the admission and use of impossible worlds [for counterpossibles] need not have much in the way of logical ramifications" (p. 546).

The last thing to say about the semantics provided above concerns '$\preceq$'. As it is, '$\preceq$' is taken from Jago (2009) without any modifications, which means that we have a set of inferences schemas, $[\![\mathfrak{R}]\!]$, and a set of 'obvious impossibilities', $[\![\mathsf{x}]\!]$. Initially, we will keep both these sets fixed. That is, we assume that our meta-language is classical—i.e., that there are only classical inference schemas in $[\![\mathfrak{R}]\!]$ and that there are only 'classically obvious impossibilities' in $[\![\mathsf{x}]\!]$. We will discuss later on what it would mean if we start to modify what is in these sets. For example, what happens if, according to a dialetheist, it is *not* the case that for every sentence φ, '$\varphi \wedge \neg\varphi$' is in $[\![\mathsf{x}]\!]$. Or, if according to some other logician, certain inference schemes should be added to or removed from $[\![\mathfrak{R}]\!]$. Throughout the rest of this chapter, we sometimes ponder these options.

5.3 Evaluating Jago's Ordering for Counterpossibles

Above we have seen how Jago's (2009; 2014) ordering works and discussed its scope when applied as a similarity ordering for counterpossibles. In this section, we will critically evaluate the applicability of the ordering that Jago (2009, 2014) provides to counterpossibles. The idea behind trying to apply this ordering to counterpossibles, is inspired by (often informal) remarks about the use of impossible worlds for counterpossibles. For example, Nolan (1997) and Goodman (2004) argue that we can have a relatively simple ordering of impossible worlds, namely, those with very clear, obvious violations to our logic are less similar to the actual world than worlds with less clear, less obvious violations. This is exactly what the Jago-ordering does, ordering impossible worlds based on subtle and blatant violations. The idea would be that 'small violations' are less obviously impossible or, in other worlds, that from 'small violations', clear impossibilities are less easily deduced. It seems then, that the Jago-ordering is indeed a viable candidate to play the role of similarity ordering for counterlogicals.

5.3.1 Small Violations to the Logic

Before we compare the Jago-ordering to some informal accounts of counterpossibles, it is interesting to note the following. Consider again Goodman's intuitive description of the use of impossible worlds for counterpossibles. Goodman says that "[i]mpossible worlds in which 'small violations' to the logic of our world occur are closer to the actual world than worlds in which 'large violations' to the logic of our world occur" (2004, p. 54). However, let us pause and think about this claim for a moment. What would a 'small' violation to classical logic look like, as opposed to a 'large' violation? In classical logic, *ex falso quodlibet* is valid, that is from a contradiction ('*ex falso*', more literally, 'from falsehood'), follows everything ('*quodlibet*').

That is, as soon as there is a violation of the logic, everything follows, our logic *explodes*, as it is sometimes called. Hence, it looks like there is no way of distinguishing between 'small' or 'large' violations, any violation leads to explosion.

However, Jago's ordering *does* seem to be able to make sense of the distinction between small and large violations. The ordering is designed to do just this, to order worlds from subtly impossible to blatantly impossible (or, from small violations to large violations). The way Jago's ordering *is* able to do so, is by looking beyond 'direct explosion' and evaluating how many applications of inference schemes are needed before an impossibility leads to an obvious impossibility.

It seems then that Jago's ordering has quite some intuitive potential to capture some of the ideas about counterpossibles. In the remainder of this chapter we will explore some of these and some possible problems for the order as similarity-ordering. In order to evaluate the applicability of the Jago-ordering as a similarity-ordering, we will compare the ordering with some informal accounts of counterpossibles. We will start with some findings of Bjerring and then turn to some worries of Goodman. In the end, we hope to have a clear judgement of whether or not the Jago-ordering is suitable as a similarity-ordering for counterpossibles.

5.3.2 The Jago-ordering and Bjerring's Spheres

Bjerring (2014a) is also concerned with sentences such as (5.5-5.6), repeated below, and with getting counterpossibles about different logics correct.

(5.5) If Intuitionistic Logic were the one true logic, then the Law of Excluded Middle would fail

(5.6) If Intuitionistic Logic were the one true logic, then the Law of Excluded Middle would be valid

For a related triple of sentences, Goodman (2004) concludes that, intuitively, the sentence where the consequent violates the facts of the actual world less than the consequent of a different sentences should, given that the antecedents are identical, be preferred (p. 60). However, this seems strange to me, for, given that the antecedents of (5.5) and (5.6) are identical, then (5.6) should be preferred to be correct, yet, it seems that we know enough about the workings of intuitionistic logic that we should judge (5.5) to be true. As Nolan (1997) says, "[w]e are often able to say quite exactly what would be the case, logically speaking at least, in the closest impossible world where an actually false logic is true," (p. 545). Bjerring provides a solution for sentences such as (5.5) and (5.6) that he takes to be the most promising way of extending the Lewis/Stalnaker analysis.

Bjerring (2014a) argues, on the basis (5.5) and (5.6), that the impossible worlds used for counterpossibles cannot be complete. And, as we have seen, we should indeed not (only) use complete impossible worlds for the semantics of counterpossibles. Bjerring then proceeds to consider two possible amendments one could make in order to avoid his criticism. Namely, one could either use partial impossible worlds or go for a *stratified* modal space (we will get back to what this means in a bit). Bjerring (2014a) argues that simply adding partial impossible worlds is not enough for properly accounting for certain counterpossibles. Specifically, counterpossibles whose antecedent seem to endorse some non-classical logic.

Allowing for partial impossible worlds is what, for example, Nolan (1997) does. When allowing for partial worlds, one would have to provide a motivated account of what sets of sentences counts as a world.[18] One could opt for all sentences closed under some (possibly non-classical) logical entailment. However, it seems that *no* consequence relation will do (see for example Vander Laan 1997). This is because, for every consequence relation, it seems that one could come up with an example of an impossible world where this relation does not hold. As Nolan (1997, p. 547) says: "I think, however, that modifying one's account of logical consequence in order to accommodate impossible situations is a mistake. For if there is an impossible situation for every way we say that things cannot be, there will be impossible situations where even the principles of subclassical logics fail".

Accordingly, the most common account of partial worlds is letting each set of sentences describe a world (cf. Priest 1992; Vander Laan 1997; Nolan 1997). However, Bjerring (2014a) argues that such an account is not good enough for a semantics for counterpossibles. For, if it is the case that each set of sentences describes a world, then the account "allows that every counterpossible can be true, and that every counterpossible can be false—perhaps except for $A \mathrel{\Box\!\!\rightarrow} A$" (Bjerring, 2014a, p. 344). In other words, Bjerring has a general worry that such a framework is too trivial in that "there is nothing in the construction of worlds—nothing in the underlying world-ontology—that helps reflect the non-trivial semantic, metaphysical, and logical dependencies or relations that obtain between various sentences" (idem.). Though some might argue that this is a good thing, Bjerring concludes that this is a bad result. He believes that worlds need to play a more explanatory role in the semantics of counterpossibles and that in order to do so, they ought "to have more structure than arbitrary sets of sentences" (idem.).

This brings us to what Bjerring takes to be the most promising direction for an extended Lewis/Stalnaker-analysis of counterpossibles, which is the account that we then will compare to the ordering we have adopted from Jago (2009, 2014). We will see that the Jago-ordering might actually play the role of the similarity ordering that Bjerring describes.

As we said, one might opt for closing worlds under entailment of some logic. Bjerring (2014a) generalises this to all logics and provides the following definition of a world:

(**World**) A set Γ of sentences [...] is a world w iff Γ is closed under logical consequence in logic $\mathcal{L}_i$ (2014a, p. 345)

With this definition, we can now define spheres of worlds that are all closed under entailment of the same logic.[19] For example, "[t]o construct the space of intuitionistically possible worlds, we take the class of all sets of sentences and close each

[18] Note that I do not want to commit to the idea of worlds being sets of sentences, however, for the purposes of *semantics*, this is a convenient way of illustrating partial worlds.

[19] As Bjerring (2014a), I leave it open whether or not we only want to quantify over existing logics or over conceivable logics as well.

of them under logical consequence in intuitonistic [sic.] logic" (p. 346). Importantly, we can add an ultra-weak logic, $\mathcal{L}_x$, that has no, or almost no, principles that govern its consequence relation. This allows us to capture certain sentences or counterpossibles that are about something "in the absence of any logic, and [. . .] that we can entertain counterpossibles that do not respect any logical constraints" (Bjerring, 2014a, p. 350) (see also Nolan 1997, pp. 547-8).[20] Bjerring (2014a) goes on to provide a detailed account of what it means then for a sentence to be true in a world through a complex analysis of fuzzy set theory and truth-by-degree. As this is not important for argument here, we will leave this aside and stick with the conceptual picture (i.e., that of spheres of worlds that adhere to the same logic).

Let us briefly mention one important conclusion of Bjerring's account. In the literature on counterpossibles, the *Strangeness of Impossibility Condition* is often mentioned (cf. Nolan 1997; Vander Laan 2004; Bjerring 2014a). This condition states that, in general, a possible world is closer to the actual world than any impossible world.[21] Bjerring's stratified modal space, allows him to present an adaptation of this condition that intuitively seems to be very appealing. He dubs it the *Relative Closeness Condition* and it can be thought of as a Strangeness of *Relative* Impossibility. The definition is as follows:

(**Relative Closeness Condition**) For any counterfactual whose antecedent presupposes that some logic $\mathcal{L}_i$ is correct (true, adequate), a world in modal space $W_{\mathcal{L}_i}$ [i.e., the set of worlds governed by logic $\mathcal{L}_i$] is closer to the actual world than any world in modal space $W_{\mathcal{L}_j}$, where $W_{\mathcal{L}_i} \neq W_{\mathcal{L}_j}$, and where $i \geq 1$ and $j > 1$ (2014a, p. 348)

With this in mind, let us now turn to Jago's ordering in relation to Bjerring's account. There are several aspects of the picture that Bjerring paints us that I believe could be captured or explicated by the Jago-ordering.

Bjerring argues that simply adding partial worlds to the semantics of counterpossibles is not enough and that such partial worlds are in need of additional structure. This is exactly Jago's motivation for ordering worlds in epistemic space, that they are in need of additional structure. Additionally, Bjerring ends up structuring the *space* of worlds, he even says that, "[s]ince worlds are located within specific spaces or spheres of worlds in stratified modal space, closeness intuitions pertain no longer just to worlds *but also to spaces of worlds*" (2014a, p. 347, emphasis added). Similarly, Jago does not provide additional structure to the worlds themselves, but to the

[20] What is interesting is that Bjerring makes the following remark with regards to this ultra-weak logic:

> In some sense, of course, the reply above still assumes that counterpossible reasoning takes place in some logic. But since the logic $\mathcal{L}_x$ can be arbitrarily weak and formally unconstrained, there is no harm in making this assumption for the purpose of giving a *semantics* for counterpossibles—although there might be for the purpose of giving an account of impossible *reasoning* (2014a, p. 350, emphasis added)

This distinction between the semantics of counterfactuals and the objects of counterfactual reasoning seems to be in line with the instrumentalistic account of semantics provided in this dissertation.

[21] Nolan (1997) says that he, indeed, finds it an intuitive principle, but that he also believes there probably to be some counterexamples. I agree with Nolan.

entirety of epistemic space. “It is the entire space of worlds that thereby gets structured, not the worlds themselves” (Jago, 2014, p. 198). It seems, then, that even though Bjerring and Jago are concerned with different problems, the motivation and the general outcome are somewhat similar. This makes it intuitively appealing to look into the relation between Jago’s ordering and the account sketched by Bjerring.

Remember that, as it is, Jago’s ordering is fixed by $\llbracket \mathbf{x} \rrbracket$ and $\llbracket \mathfrak{R} \rrbracket$, which we both assumed to be classical. But then, it seems, that we do not yet get the right results. For example, consider the following sentence:

(5.11) If Intuitionistic Logic were correct and $\neg\varphi$ were not true, then it would be the case that φ

Given that the Law of Excluded Middle is not valid in intuitionistic logic, one cannot deduce φ from $\neg\varphi$ not being true. So, it seems that (5.11) should be false. However, given that $\llbracket \mathbf{x} \rrbracket$ and $\llbracket \mathfrak{R} \rrbracket$ are classical, a world where both the antecedent and the consequence are true is less obviously impossible (it is possible), then a world where the antecedent is true while the consequent is not. So, if we keep $\llbracket \mathbf{x} \rrbracket$ and $\llbracket \mathfrak{R} \rrbracket$ fixed, the theory seems to predict (5.11) to be true.

Note that ‘if intuitionistic logic were correct’ is a meta-statement about which logic is the correct one. Maybe, such a meta-statement should be reflected in a change of context, or something else. Bjerring (2014a) often talks about counterpossibles “whose antecedent presupposes that some logic [. . .] is correct”, which is of great importance when evaluating such counterpossibles. Intuitively, the Jago-ordering has a pretty straightforward way of dealing with counterpossibles whose antecedents presuppose a different (non-classical) logic. For example, one could alter the contents of $\llbracket \mathbf{x} \rrbracket$, $\llbracket \mathfrak{R} \rrbracket$, or both. That is, if one would allow $\neg(\varphi \vee \neg\varphi)$ *not* to be in $\llbracket \mathbf{x} \rrbracket$, then (5.11) might indeed come out false (or, at least, it is not obvious that it, falsely, comes out true). I think that there is no definitive, clear way of determining whether one has to alter $\llbracket \mathbf{x} \rrbracket$, $\llbracket \mathfrak{R} \rrbracket$, both, or neither. This is, probably, very much context dependent. In cases such as (5.5-5.6), where both the antecedent and the consequent concern meta-statements about the logic in question, it seems likely that both $\llbracket \mathbf{x} \rrbracket$ and $\llbracket \mathfrak{R} \rrbracket$ should be altered. However, in cases where only the antecedent presupposes a particular logic, it might be more likely to only alter the content of one of the two. That is, the phenomenon described here, and the shiftiness of $\llbracket \mathbf{x} \rrbracket$ and $\llbracket \mathfrak{R} \rrbracket$, seems to be very context depend, similar to the similarity ordering for counterfactuals in general.

Clearly, this is very speculative, and future research is needed in order to fully work out the details, but it seems to me that the Jago-ordering is able to capture Bjerring’s (2014a) idea behind the Relative Closeness Principle (p. 348). By allowing shiftiness in $\llbracket \mathbf{x} \rrbracket$ and $\llbracket \mathfrak{R} \rrbracket$, we can allow for sentences concerning meta-statements about the logic in question to be properly analysed. There is one interesting difference between Bjerring’s account and the one suggested above, namely that Bjerring never mentions the context dependency. It thus seems that for Bjerring everything is fixed in the model, whereas the shiftiness of $\llbracket \mathbf{x} \rrbracket$ and $\llbracket \mathfrak{R} \rrbracket$ allows for more flexibility to capture the context-sensitivity.

It seems then that Jago’s ordering is able to capture most of what Bjerring sketches to be a way “we might develop an extended Lewis-Stalnaker semantics for that can avoid the problems” that other extensions had (2014a, p. 343).

The fact that the content of $[\![\mathsf{x}]\!]$ and $[\![\mathfrak{R}]\!]$ can shift due to context is not at all a radical idea. As we have seen above, counterfactuals in general are very context sensitive and the similarity ordering for counterfactuals, in general, differs greatly with context. Letting the contents of $[\![\mathsf{x}]\!]$ and $[\![\mathfrak{R}]\!]$ shift merely seems to reflect the intuitive context-sensitivity of counterfactuals.

The shiftiness of $[\![\mathsf{x}]\!]$ and $[\![\mathfrak{R}]\!]$ is also highlighted in different scenarios. Consider, again, the following sentence:

(5.5) If Intuitionistic Logic were the one true logic, then the Law of Excluded Middle would fail

Now consider this sentence being uttered in two different occasions by two different speakers:[22]

In scenario (1), the sentence is uttered by an intuitionistic logician, who believes intuitionistic logic to be correct, when she tries to explain the relation between the Law of Excluded Middle and intuitionistic logic to a student. In that case, (5.5) seems to be true.

In scenario (2), the sentence is uttered by a classical logician, who believes, above anything else, that the Law of Excluded Middle is valid. She discusses the status of the Law of Excluded Middle and argues that it would still be true, even if intuitionistic logic were correct. In this case, (5.5) seems to be false

It seems to me that the difference in truth-value of (5.5) can be explained by the shiftiness of the contents of $[\![\mathsf{x}]\!]$ and $[\![\mathfrak{R}]\!]$. For example, if $\neg(\varphi \lor \neg\varphi)$ is included in $[\![\mathsf{x}]\!]$ in (2), then it seems that the right predictions can be given. But, if in accordance with intuitionistic logic, $\neg(\varphi \lor \neg\varphi)$ is not seen as a contradiction (i.e., not in $[\![\mathsf{x}]\!]$), as in (1), then the right predictions can be given for that scenario.

As I mentioned before, it seems to me that the contents of $[\![\mathsf{x}]\!]$ and $[\![\mathfrak{R}]\!]$ and their shiftiness is an indication of context dependence and a lot of work can still be done on working out the precise details of it. However, it seems that the Jago-ordering is a good candidate to explicate some of this context-dependence and fulfilling the job of the similarity ordering for counterpossibles that Bjerring (2014a) sketches.

Let us now turn to, what Goodman (2004), calls *the new problem of counterpossibles.*

5.3.3 The Jago-ordering and explicitly impossible antecedents

Goodman (2004) argues that even extended Lewis/Stalnaker-accounts run into trouble with certain counterpossibles. He calls this *the new problem of counterpossibles.*

Goodman's problem is a different version of Lewis' motivation for allowing ties for closeness in his semantics for counterfactuals. Lewis argues for ties based on the following sentence pair:

[22] This shiftiness, it seems to me, is similar to the context-dependence of counterfactuals such as:

(5.4) If Edinburgh were in Italy, it would rain less in Edinburgh

(5.1) If Bizet and Verdi were compatriots, then Bizet would be Italian

(5.2) If Bizet and Verdi were compatriots, then Bizet would be French

As we saw above, Lewis argues that these examples show that there need to be ties in the similarity ordering, for both consequences seem equally likely. Goodman (2004) extends this line of reasoning to extended semantics for counterpossibles. Consider the following sentence pair (example is from Goodman 2004):

(5.12) If snow would be white and snow would be non-white, then snow would be white.

(5.13) If snow would be white and snow would be non-white, then snow would be non-white.

The intuition here is that both of these sentences are true. Goodman then argues "that if the 'Bizet/Verdi' sentences show that relevant possible worlds are tied for closeness to actuality, then [(5.12)] and [(5.13)] show that we ought to admit ties for closeness to actuality among the impossible worlds" (p. 56). Thus, according to Goodman, if one accepts an extended account of counterpossibles that includes impossible worlds, she finds herself "on the horns of a dilemma" (p. 59). Either she accepts that there are ties possible for closeness amongst impossible worlds. If she does so, then both (5.12) and (5.13) are both deemed false, which, intuitively, seems wrong. Or she does not accept ties amongst impossible worlds for closeness, in which case she has to accept one of the sentences as true and the other one as false, which also seems wrong.

Goodman (2004) concludes from all this that "we ought to retain a closest possible or impossible worlds account of counterfactuals, but I also think that the new problem of counterfactuals shows us that our semantics should *ignore* (in a sense) the impossible worlds that judge explicit contradictions true" (2004, p. 63-64, original emphasis). The problem that Goodman addresses here points to a similar problem for the Jago-ordering.

Consider again the minimal elements of the Jago-ordering, the worlds that represent obvious impossibilities, $[\![\mathsf{x}]\!]$. These worlds are such that, for a world, $w \in [\![\mathsf{x}]\!]$, $w' \preceq w$, for any world w'. So, if w and w' are both in $[\![\mathsf{x}]\!]$, then both $w' \preceq w$ and $w \preceq w'$. Let us abbreviate this to $w \sim w'$. It follows, that all the worlds in $[\![\mathsf{x}]\!]$ stand in an equivalence-relation to each other. That is, they are all equally blatantly impossible.

The sentences that Goodman discusses, contain antecedents that consider 'explicit contradictions', which, gathering from his examples, always contain blatant impossibilities of the form $\ulcorner\varphi \wedge \neg\varphi\urcorner$. If this is the case, then these 'explicit contradictions' are very likely to be in $[\![\mathsf{x}]\!]$. So, on our semantics, these would come out (trivially) false, for all the worlds in $[\![\mathsf{x}]\!]$ are on a par. Hence, there will be worlds where the consequent is true, there will be worlds where the consequent is false, and these will be equally similar. This points to the issue that the Jago-ordering does not extend beyond the borders of the obviously impossible. Hence, it judges any counterpossible whose antecedent is in $[\![\mathsf{x}]\!]$ as trivially false.

One may, with Goodman (2004), argue that this is a good thing and that our semantics should "*ignore* [. . .] the impossible worlds that judge explicit contradictions true" (p. 63-64, original emphasis). However, it seems that we *do* want to say something about such counterpossibles, for consider the following sentences:

(5.14) $(\varphi \wedge \neg\varphi) \,\Box\!\!\rightarrow \varphi$

(5.15) $(\varphi \wedge \neg\varphi) \,\Box\!\!\rightarrow \neg\varphi$

(5.16) $(\varphi \wedge \neg\varphi) \,\Box\!\!\rightarrow \chi$, for any χ

Arguably, (5.14) and (5.15) are less trivial then (5.16). Hence, in order to capture this, we need to say something about the ordering beyond the boundaries of the obviously impossible. In order to do so, we turn to the context, as is, generally, done quite often with counterfactuals.

Arguably, sentences such as (5.14) and (5.15) are used to illustrate some features of conjunction. If that is indeed the case, then the context should reflect this by keeping features such as conjunction-elimination and conjunction-introduction fixed. If such features are fixed, then worlds in which $\varphi \wedge \neg\varphi$ and φ (or $\neg\varphi$) are true are closer than worlds where $\varphi \wedge \neg\varphi$ and χ are true. Something along these lines would need to be worked out, but seems to be on the right track. One might object that this response shifts too much of the responsibility to context, however, this objection could be made against any account of counterfactuals that employ only possible worlds. Therefore, this is, at least, not an argument against the Jago-ordering in itself.

I want to briefly mention another phenomenon that the Jago-ordering, in combination with context, might shed some light on, namely, a phenomenon that is related to Goodman's (2004) conclusion that we should ignore explicit contradictions.

Consider two logicians debating the status of intuitionism versus classical logic. The classical logician is so sure of her idea that classical logic is the one true logic, she utters the following sentence:

(5.17) If the Law of Excluded Middle were to fail, then I would be the Queen of France

Note that it seems strange to assign a truth-value to such a sentence. For the classical logician only seems to utter (5.17) in order to illustrate how ridiculous she finds the idea that the Law of Excluded Middle might fail.[23] It seems that in this case the addressee can infer that the speaker takes the antecedent to be obviously impossible, that is, in $[\![\mathbf{x}]\!]$, or at least, very close to it. To mount a full-scale defence of this type of explanation would lead us to far astray into the fields of pragmatics. I simply want to note, that it seems that the Jago-ordering plus some context-sensitivity and pragmatics might have something sensible to say about such cases.

[23] Consider a more natural, parallel, conditional such as:

(5.1) Right [said sarcastically], if Belgium were to win the European Soccer Championship, I would be the Easter Bunny

.

5.4 Concluding Remarks and Future Research

In this chapter, we have looked at counterpossibles: counterfactuals with an impossible antecedent. Given the argument from utility in Chapter 3 and the claim that counterpossibles are the most prominent area for the use of impossible worlds, strengthening the case for non-vacuous counterpossibles strengthens the case for the use of impossible worlds in semantics. In order to do so, we looked at providing a (small) step forwards with regards to a similarity ordering for impossible worlds. We did so by applying the order, which Jago (2009, 2014) proposes for the problem of epistemic content, to counterpossibles.

Even though the Jago-ordering only seems to be applicable to a limit number of cases, it does seem as if the order has some intuitive appeal. Especially when comparing it to the informal accounts of Goodman (2004) and Bjerring (2014a), it looks as if the Jago-order is able to explicate and explain some interesting phenomena related to counterpossibles. In explicating these phenomena, we often need to turn to the context to fully explain the phenomenon in question. However, this only seems to be appropriate, for in general, similarity orderings for counterfactual are famously context-sensitive.

In this concluding section, I briefly want to flag two possible avenues of future research concerning counterpossibles and especially the Jago-ordering. That is, studying the formal properties of the Jago-ordering, $\preceq$, and looking into a possible extension to *countermathematicals.*

Much of the literature on counterfactuals and counterpossibles notes the many properties and restrictions concerning the similarity ordering of the Lewis/Stalnaker analysis. Whether it is weakly centred, strongly centred, transitive, satisfies the limit assumption, etcetera. Nothing has yet been said about this concerning the Jago-ordering. We followed Jago (2009) in assuming that $\preceq$ is a total-order on the set of worlds, however, as Jago himself notes, nothing hinges on this and it might very well be that turns out that a partial order is more apt. Given the set up of the Jago-ordering (i.e., as an ordering based on proof-complexity) it seems that a total-order is apt and that, for example, transitivity holds. Yet, a full technical analysis is needed in order to properly evaluate these claims.

Remember that Jago (2009) argued that worlds that represent obvious impossibilities are in the set of $\llbracket \mathsf{x} \rrbracket$. That is, worlds that represent obvious *logical* impossibilities. In discussing Bjerring's (2014a) logical spheres, we already noted that the content of $\llbracket \mathsf{x} \rrbracket$ (and that of $\llbracket \mathfrak{R} \rrbracket$) might change. So, why not change the content of $\llbracket \mathsf{x} \rrbracket$ to include 'obvious *mathematical* impossibilities'?

It looks as if there is nothing that, *prima facie*, prohibits us in letting $\llbracket \mathsf{x} \rrbracket$ contain mathematical obvious impossibilities. For example, there may be contexts where it is without a doubt clear that Peano's Axioms are the most fundamental, most obvious truths of number-theory. In such a context, it seems that the negations of Peano's Axioms are 'obvious impossibilities' and thus in $\llbracket \mathsf{x} \rrbracket$. If we then let $\llbracket \mathfrak{R} \rrbracket$ contain some simple mathematical inference rules, we could then, arguably, construct an ordering similar to the one Jago provides based on logical structure.

Clearly, ordering mathematically impossible worlds is not so easy is depicted here and it might be very controversial to include the negations of Peano's Axioms in $\llbracket \mathsf{x} \rrbracket$.

However, this was only meant as a vivid illustration to point out that it might be possible to extend a Jago-style ordering of impossible worlds for counter*mathematicals*. And even though this is only a rough sketch of what such an ordering could look like, it is interesting that we can already note the following. Subtle logical impossibilities are often subtle through their complex syntax. That is, obvious logical impossibilities are often very simple, syntactically. Yet, the same does not seem to hold for mathematics. Mathematical statements that are, syntactically, very simple may be very hard to prove and *vice versa*.

Extending a Jago-style ordering of impossible worlds for counter*metaphysicals* (e.g., 'If Kripke would have been born from different parent, then he would not have written *Naming and Necessity*') seems very difficult.[24] It seems that a Jago-style ordering for countermetaphysicals is, at least, very complicate and will contain very complicated $[\![\mathsf{x}]\!]$ and $[\![\mathfrak{R}]\!]$. With such a prospect, one might object that extended Lewis/Stalnaker accounts of counterpossibles become very *ad hoc*, having three different types of similarity orderings. However, arguably, this is what is to be expected. For there are many ways in which a world can be impossible: a world can violate a metaphysical necessity (e.g., I can be born from different parents), a world can violate a logical necessity (e.g., the Law of Excluded Middle can fail), or a world can violate a mathematical necessity (e.g., there could be a number whose successor is 0). It might be argued that the fact that we need different similarity orderings for these is not a weakness of the system, but that it is a reflection of the fact that a world can be impossible in different respects.

Extending the Jago-ordering to countermathematicals is left for future work, as well as testing the exact formal properties of the Jago-ordering, $\preceq$. I do believe, however, that the Jago-order has proved itself to be a valuable addition to the semantics of counterpossibles, providing a better understanding of the workings of counterlogicals and some handles to explicate certain context-dependent phenomena.

All in all, it seems that impossible worlds and their ordering are a valuable instrument in providing a semantics for counterpossibles. If that is indeed the case, this would strengthen the case of impossible worlds and their place in the semanticist's toolkit.

[24] Though, possibly not impossible.

Conclusion: Impossibilities, only if you believe them

Nothing's Impossible

The Doorknob, Alice in Wonderland

This dissertation can be thought of as an elaborate argument for the use of impossible worlds in natural language semantics. That is, it is a philosophical work, arguing for the use of a certain formal tool. Obviously, such an argument cannot be made without engaging in some formalisms.

The argument for impossible worlds in semantics is based on two arguments: one on a methodological level and one on a utility level. First, I argued that the costs for allowing impossible worlds in one's semantics is no higher than allowing for possible worlds in semantics. In particular, I argue that there are *no* costs. This is due to a methodological view that I dubbed *Semantic Agnosticism.* The view advocates an instrumentalist stance, *when* doing semantics. That is, the semanticist, qua semanticist, should remain agnostic about the nature of worlds, she merely uses world *postulates* in her model. This does not rule out that there are interesting matters of fact concerning the nature of worlds that she might investigate when doing metaphysics. It is just that these are two orthogonal issues. Secondly, I aimed to strengthen the argument from utility (for which now only a weaker conclusion, the acceptance of world *postulates* into one's semantics toolkit, was needed). We did so by suggesting a formal similarity ordering for impossible worlds that could be used in the analysis of counterlogicals. Though there is still a lot of work to be done in this field, the ordering seems promising enough to be applicable to counterlogicals.

Some philosophers might object to the absurd entities that are impossible worlds, yet it is clearly a feature of our language that we can discuss scenarios beyond the bounds of the possible (think of the attitude ascription of Mary's believing that Fermat's Last Theorem to be false). The point of this dissertation is to argue that, independent of what scruples one might have of the objects 'impossible worlds', in natural language semantics we need to allow for impossible world *postulates* in order to properly capture this feature of language use. The semantics provided in this thesis aims to provide a very general starting point to capture such features. It might be that the semantics needs to be polished in order to capture more of the ‘intentional inferences’ and counterpossible intuitions or that the semantics is good as it is. Further research will tell.

Let us conclude with a final insight of the Mad Hatter:

Alice: *This is impossible!*
Mad Hatter: *Only if you believe it is.*

References

van Atten, M. (2015). Luitzen egbertus jan brouwer. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Stanford, CA.: CSLI Publications, winter ed.

Bach, K. (1997). Do Belief Reports Report Beliefs? *Pacific Philosophical Quarterly*, *78*, 215–241.

Balaguer, M. (1996). A fictionalist account of the indispensable applications of mathematics. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *83*(3), 291–314.

Ball, D. (2016). Semantics as Measurement. Unpublished draft.

Beall, J. (2005). Transparent Disquotationalism. In J. Beall, & B. Armour-Garb (Eds.) *Deflationism and Paradox*, (pp. 7–22). Oxford: Oxford University Press.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford: Clarendon Press.

Berto, F. (2007). Is Dialetheism an Idealism? The Russellian Fallacy and the Dialethist's Dilemma. *Dialectica*, *61*(2), 235–263.

——— (2010). Impossible Worlds and Propositions: Against the Partity Thesis. *The Philosophical Quarterly*, *60*(240), 471–486.

——— (2013). Impossible Worlds. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Stanford, CA.: CSLI Publications.

Berto, F., & Plebani, M. (2015). *Ontology and Metaontology: A Contemporary Guide*. London: Bloomsbury.

Bjerring, J. C. (2010). *Non-Ideal Epistemic Spaces*. Ph.D. thesis, The Australian National University.

——— (2013). Impossible worlds and logical omniscience: an impossibility result. *Synthese*, *190*(13), 2505–2524.

——— (2014a). On counterpossibles. *Philosophical Studies*, *168*(2), 327–353.

——— (2014b). Problems in Epistemic Space. *Journal of Philosophical Logic*, *43*, 153–170.

Bjerring, J. C., & Schwarz, W. (2016). Granularity problems. Draft compiled on 25 February 2016 retrieved from `http://www.umsu.de/papers/granularity.pdf` on 24 May 2016.

Brogaard, B., & Salerno, J. (2013). Remarks on counterpossibles. *Synthese*, *190*, 639–660.

Carnap, R. (1956). *Meaning and Necessity: A Study in Semantic and Modal Logic*. Chicago, IL.: University of Chicago Press, 2nd ed.

Carroll, L. (1865). *Alice's Adventures in Wonderland*.

Chakravartty, A. (2015). Scientific realism. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Stanford, CA.: CSLI Publications, fall ed.

Chalmers, D. J. (2011). Propositions and attitude ascriptions: A fregean account. *Noûs*, *45*, 595–639.

Chierchia, G., & McConnell-Ginet, S. (1990). *Meaning and Grammar: An Introduction to Semantics*. Cambridge, MA.: MIT Press.

Copeland, B. J. (1979). On When a Semantics is Not a Semantics: Some Reasons for Disliking the Routley-Meyer Semantics for Relevance Logic. *Journal of Philosophical Logic*, *8*(1), 399–413.

——— (1983). Pure Semantics and Applied Semantics. *Topoi*, *2*, 197–204.

Cresswell, M., & von Stechow, A. (1982). *De Re* belief generalized. *Linguistics and Philosophy*, *5*, 503–535.

Curd, M., Cover, J. A., & Pincock, C. (Eds.) (2013). *Philosophy of Science*. New York, NY.: W. W. Norton & Company, second ed.

Dekker, P. (2012). *Dynamic Semantics*, vol. 91 of *Studies in Linguistics and Philosophy*. New York, NY: Springer.

Divers, J. (2002). *Possible Worlds*. London: Routledge.

——— (2006). Possible-Worlds Semantics Without Possible Worlds: The Agnostic Approach. *Mind*, *115*(458), 187–225.

Dummett, M. (1973). *Frege: Philosophy of Language*. New York, NY.: Harper & Row Publishers.

——— (1974). *Truth and Other Enigmas*. Cambridge, MA.: Harvard University Press.

Edelberg, W. (1994). Propositions, Circumstances, Objects. *Journal of Philosophical Logic*, *23*(1), 1–34.

Edgington, D. (1995). On conditionals. *Mind*, *104*(414), 235–329.

Eklund, M. (2011). Fictionalism. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Stanford, CA.: CSLI Publications, fall ed.

Elbourne, P. (2010). Why Propositions Might be Sets of Truth-supporting Circumstances. *Journal of Philosophical Logic*, *39*, 101–111.

Fine, K. (2013). Constructing the Impossible. Retrieved on 19th of April, 2016 from `https://www.academia.edu/11339241/Constructing_the_Impossible`.

Frege, G. (1892). On Sense and Reference. In P. Geach, & M. Black (Eds.) *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Basil Blackwell, 2nd ed. (1960). Reprinted in A. P. Martinich and D. Sosa, (2013).

Gamut, L. (1991). *Logic, Language, and Meaning. Volume 2, Intensional Logic and Logical Grammar*. Chicago, IL.: The University of Chicago Press.

Goodman, J. (2004). An Extended Lews/Stalnaker Semantics an the New Problems of Counterpossibles. *Philosophical Papers*, *33*(1), 35–66.

Groenendijk, J., & Stokhof, M. (1984). *Studies of the Semantics of Questions and the Pragmatics of Answers*. Ph.D. thesis, University of Amsterdam, Amsterdam.

Heim, I., & Kratzer, A. (1998). *Semantics in Generative Grammar*. Malden, MA.: Blackwell Publishing.

Hintikka, J. (1962). *Knowledge and Belief*. Ithaca, NY.: Cornell University Press.

——— (1969). Semantics for Propositional Attitudes. In D. Davidson, J. Hintikka, G. Nuchelmans, & W. C. Salmon (Eds.) *Models for Modalities*, (pp. 87–111). Dordrecht, Holland: D. Reidel Publishing Company.

Jago, M. (2009). Logical information and epistemic space. *Synthese*, *167*, 327–341.

——— (2012). Constructing worlds. *Synthese*, *189*, 59–74.

——— (2013). Are Impossible Worlds Trivial? In V. Puncochar, & P. Svarny (Eds.) *The 2012 Logica Yearbook*. Filosophia.

——— (2014). *The Impossible*. Oxford: Oxford University Press.

——— (2015). Impossible Worlds. *Noûs*, *49*(4), 713–728.

Janssen, T. M. V. (2016). Montague Semantics. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Stanford, CA.: CSLI Publications, spring ed.

Kanger, S. (1957). The morning star paradox. *Theoria*, *23*(1), 1–11.

Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.) *Themes from Kaplan*, (pp. 481–563). Oxford: Oxford University Press.

King, J. C. (2007). *The nature and Structure of Content*. Oxford: Oxford University Press.

——— (2014). Structured propositions. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Stanford, CA.: CSLI Publications, spring ed.

Kiourti, I. G. (2009). *Real Impossible Worlds: The Bounds of Possibility*. Ph.D. thesis, University of St. Andrews.

Krakauer, B. (2013). What are impossible worlds? *Philosophical Studies*, *165*(3), 989–1007.

Kripke, S. (1959). A Completeness Theorem in Modal Logic. *The Journal of Symbolic Logic*, *24*(1), 1–14.

——— (1980). *Naming and Necessity*. Oxford: Blackwell Publishers.

Lenci, A. (2008). Distributional semantics in linguistics and cognitive research. *Rivista di Linguistica*, *20*(1), 1–31.

Lewis, D. K. (1968). Counterpart Theory and Quantified Modal Logic. *Journal of Philosophy*, *65*, 113–126.

——— (1970). General semantics. *Synthese*, *22*, 18–67.

——— (1973). *Counterfactuals*. Cambridge, MA.: Harvard University Press.

——— (1978). Truth in fiction. *American Philosophical Quarterly*, *15*(1), 37–46.

——— (1980). Index, context, and content. In S. Kanger, & S. hman (Eds.) *Philosophy and Grammar*, (pp. 79–100). Dordrecht, Holland: D. Reidel Publishing Company.

——— (1986). *On the Plurality of Worlds*. Oxford: Blackwell Publishers.

Montague, M. (2007). Against Propositionalism. *Noûs*, *41*(3), 503–518.

Montague, R. (1970a). English as a formal language. In *Linguaggi nella Societa et nella Technica*, (pp. 188–221). Edizioni di Communita. Reprinted in Thomason, R. H. (ed.) (1974), pp. 188-221.

——— (1970b). Universal grammar. *Theoria*, *36*, 373–398. Reprinted in Thomason, R. H. (ed.) (1974) pp. 222-246.

Nolan, D. P. (1997). Impossible Worlds: A Modest Approach. *Notre Dame Journal of Formal Logic*, *38*(4), 535–572.

——— (2013). Impossible Worlds. *Philosophy Compass*, *8*(4), 360–372.

——— (2016). Modal Fictionalism. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Stanford, CA.: CSLI Publications, spring ed.

Partee, B. H. (1989). Possible worlds in model-theoretic semantics: A linguistic perspective. In S. Allen (Ed.) *Possible Worlds in Humanities, Arts, and Sciences: Proceedings of Nobel Symposium 65*, (pp. 93–123). Berlin & New York: Walter de Gruyter.

Pickel, B. (2015). Structured Propositions in Generative Grammar. Draft.

Priest, G. (1992). What is a Non-Normal World? *Logique & Analyse*, *139-140*, 291–302.

——— (1997). Sylvan's Box. *Notre Dame Journal of Formal Logic*, *38*, 573–581.

——— (2005). *Towards Non-Being; the logic and metaphysics of intentionality*. Oxford: Oxford University Press.

——— (2006). *Doubt Truth to be a Liar*. Oxford: Oxford University Press.

——— (2008). *An Introduction to Non-Classical Logic; From If to Is*. Cambridge: Cambridge University Press, 2nd ed.

Priest, G., & Berto, F. (2013). Dialetheism. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Stanford, CA.: CSLI Publications, summer ed.

Quine, W. (1948). On What There Is. *The Review of Metaphysics*, *2*(5), 21–38.

——— (1953). *From a Logical Point of View*. Cambridge, MA.: Harvard University Press.

——— (1960). *Word & Object*. Cambridge, MA.: The MIT Press.

Quine, W. V. O. (1956). Quantifiers and Propositional Attitudes. *Journal of Philosophy*, *53*(5), 177–187.

Rabern, B. (2012a). Against the identification of assertoric content with compositional value. *Synthese*, *189*(1), 75–96.

——— (2012b). *Monsters and Communication: The semantics of concontext shifting and sensitivity*. Ph.D. thesis, Australian National University, Canberra.

——— (forthcoming). The history of the use of $[\![.]\!]$-notation in natural language semantics. *Semantics & Pragmatics*.

Rantala, V. (1982). Impossible Worlds Semantics and Logical Omniscience. *Acta Philosophica Fennica*, *35*, 106–115.

Rescher, N., & Brandom, R. (1980). *The Logic of Inconsistency*. Oxford: Blackwell Publishers.

Ripley, D. (2012). Structures and circumstances: two ways to fine-grain propositions. *Synthese*, *189*, 97–118.

Rosen, G. (1990). Modal Fictionalism. *Mind*, *99*(395), 327–354.

Salmon, N. (1986). *Frege's Puzzle*. Cambridge, MA.: The MIT Press.

Schoonen, T. (2014). *Beliefs and 'Believes'*. Master's thesis, University of Edinburgh, Edinburgh, United Kingdom.

Schwarz, W. (2016). Semantic possibility. Unpublished draft, version 4 April 2016, retrieved on 22th of April from `http://www.umsu.de/papers/metasemantics.pdf`.

Sider, T. (2010). *Logic for philosophy*. Oxford: Oxford University Press.

Soames, S. (1987). Direct Reference, Propositional Attitudes, and Semantic Content. *Philosophical Topics*, *15*, 47–87.

——— (1988). Substitutivity. In J. J. Thomson (Ed.) *On Being and Saying: Essays in Honor of Richard Cartwright*, (pp. 99–132). Cambridge, MA.: The MIT Press.

——— (1989). Direct Reference and Propositional Attitudes. In J. Almog, J. Perry, & H. Wettstein (Eds.) *Themes from Kaplan*, (pp. 393–420). New York, NY.: Oxford University Press.

——— (2008). Why Propositions Cannot be Sets of Truth-supporting Circumstances. *Journal of Philosophical Logic*, *37*, 267–276.

——— (2010). *What Is Meaning?*. Princeton, NJ.: Princeton University Press.

Stalnaker, R. C. (1968). A theory of conditionals. In H. et al. (Ed.) *Ifs: conditionals, belief, decision, chance, and time*, (pp. 41–55). Dordrecht, Holland: D. Reidel Publishing Company.

——— (1976a). Possible Worlds. *Noûs*, *10*(1), 65–75.

——— (1976b). Propositions. In A. F. Mackay, & D. Merrill (Eds.) *Issues in the Philosophy of Language*, (pp. 79–91). New Haven: Yale University Press.

——— (1978). Assertion. *Syntax and Semantics*, *9*, 315–332.

——— (1984a). *Inquiry*. Cambridge, MA.: The MIT Press.

——— (1984b). The problem of intentionality. In *Inquiry*, (pp. 1–26). The MIT Press.

——— (1996). Impossibilities. *Philosophical Topics*, *24*(1), 193–204.

Stanley, J. (2001). Hermeneutic Fictionalism. *Midwest Studies in Philosophy*, *XXV*, 36–71.

——— (2008). Philosophy of Language. In D. Moran (Ed.) *The Routledge Companion to 20th Century Philosophy*, (pp. 382–437). New York: Routledge.

Stokhof, M. (2013). Arguing about dynamic meaning. In A. Baltag, & S. Smets (Eds.) *Logical/Informational Dynamics*. Springer.

Thomason, R. H. (Ed.) (1974). *Formal Philosophy, Selected Papers of Richard Montague*. New Haven and London: Yale University Press.

Turney, P. D., & Pantel, P. (2010). From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188.

Van Inwagen, P. (1998). Meta-Ontology. *Erkenntnis*, *48*, 233–250.

Vander Laan, D. A. (1997). The Ontology of Impossible Worlds. *Notre Dame Journal of Formal Logic*, *38*(4), 597–620.

——— (2004). Counterpossibles and Similarity. In F. Jackson, & G. Priest (Eds.) *Lewisian Themes*, (pp. 258–276). Oxford: Clarendon Press.

Veltman, F. (2005). Making Counterfactual Assumptions. *Journal of Semantics*, *22*, 159–180.

Von Fintel, K., & Heim, I. (2011). Intensional semantics. Unpublished ms. from `http://web.mit.edu/fintel/fintel-heim-intensional.pdf`.

Walton, K. (1990). *Mimesis as Make-Believe*. Cambridge, MA.: Harvard University Press.

Williamson, T. (2007). *The Philosophy of Philosophy*. Oxford: Blackwell Publishing.

——— (forthcoming). Counterpossibles. In B. Armour-Garb, & F. Kroon (Eds.) *Philosophical Fictionalism*.

Yablo, S. (1993). Is Conceivability a Guide to Possibility? *Philosophy and Phenomenological Research*, *53*(1), 1–42.

——— (1998). Does Ontology Rest on a Mistake? *Proceedings of the Aristotelian Society, Supplementary Volumes*, *72*, 229–261.

——— (2001). Go Figure: A path through fictionalism. *Midwest studies in philosophy*, *25*(1), 72–102.

——— (2010). *Things: Papers on Objects, Events, and Properties*. Oxford: Oxford University Press.

Yagisawa, T. (1988). Beyond Possible Worlds. *Philosophical Studies*, *53*(2), 175–204.

——— (2010). *Worlds and Individuals, Possible and Otherwise*. Oxford: Oxford University Press.