Impossible Worlds

Aafke de Vos (3232026)

January 16, 2012

Bachelor's thesis in Philosophy

Supervised by Michael De

Contents

1 Introduction

For many years philosophers are discussing about the metaphysics of possible worlds.[1] Forty years ago, most people rejected the concept of possible world; they argued that this concept did not make sense or even declared it as useless. But possible worlds have proven useful for a large variety of purposes, and a lot of great philosophers have argued for possible worlds. Modal statements like 'there are many ways the world could have been' can be taken at face value, as talking about possible worlds other than the actual one. Furthermore, a lot of modal concepts can be analyzed in terms of possible worlds. Examples are physical necessity, *de re* modal statements and deontic obligation. However, there is no consensus about the metaphysical nature of possible worlds. The two main options in the metaphysics of possible worlds are modal realism and ersatzism (or actualist abstractionalism) in its various forms. Modal realism takes merely possible worlds as concrete entities. According to the most prominent supporter of modal realism, David Lewis, merely possible worlds are spatiotemporally-isolated, causally-disconnected pieces of real estate on an ontological par with the actual universe. Like the modal realist, the ersatzer is a realist about worlds, but the ersatz possible worlds differ in their nature. According to the ersatzer, certain bits of abstract entities (for example propositions or states of affairs) play the modal role in our theorizing about ways the world could have been. There are many forms of ersatzism. For example, the linguistic ersatzer takes worlds to be a kind of *book* because they represent by containing sentences of an interpreted language. On the contrary, Robert Stalnaker thinks worlds are world-natures or maximal properties, but they are not representations.

Nowadays we stand at the beginning of the metaphysical debate about impossible worlds. Reasoning about impossible situations is very useful for theoretical purposes, but the benefits of impossible worlds are still not as fully explored as those of possible worlds. As with possible worlds, the supporters of impossible worlds disagree on their metaphysical nature. Most impossible worlds theorists argue that there is no ontological difference between merely possible and impossible worlds; so one's favorite metaphysics of possible worlds is also ones favorite metaphysics of impossible worlds. Corresponding to the two main options in the metaphysics of possible worlds, there are two main positions in the metaphysics of impossible worlds. Extended modal realism takes the existing framework of concrete possible worlds and adds concrete impossible worlds. Takashi Yagisawa[2] is a prominent supporter of this view. The extended ersatzer treats impossible worlds as actually existing abstract entities on a par with ersatz possible worlds. As there are various forms of ersatzism, there are also many abstractionalist impossible world theories. David vander Laan[3] is a prominent actualist abstractionalist of impossible worlds; he argues that impos-

[1] Philosophers which are involved in the early history of possible worlds are Duns Scotus, William of Ockham, John Wallis and Leibniz. More on the origins of possible worlds semantics can be found in Copeland, 2002.

[2] Yagisawa, 1988.

[3] Vander Laan, 1997.

sible worlds constitute of impossible states of affairs which are maximal.

In this paper, I will motivate an abstractionalist account of impossible worlds. In section 2, I will explain the concept 'impossible world' and I will specify my abstractionalist account of impossible worlds. In section 3, I will discuss some general motivations behind impossible worlds. My aim is to approach the nature of impossible worlds from a neutral point of view. However, while expanding my arguments, it turns out that specifying the nature of impossible worlds is often very important. In section 4, I will consider several arguments against impossible worlds; first I will focus on general objections, and later I will discuss an objection to an abstractionalist account of impossible worlds.

2 What are 'impossible worlds'?

First of all it has to be clear what is meant by the concept 'impossible world'. Just as we intuitively consider possible worlds as 'ways things could be', impossible worlds can be seen as 'ways things could not be'.[4] But we cannot respond to the question 'what are impossible worlds?' with the simple answer 'worlds where things hold, which could not be'. We have to answer the question: what are things, which could not be? Or, to state it somewhat differently: what sorts of things are impossible?

We often speak of things being impossible in a relative sense. Suppose a student is half way through writing her thesis, and she has her final deadline in three days. She may say: "There is no way I can finish my thesis in time." What the student really means is that, giving the amount of work to do and the time left, it is impossible to finish her thesis in time. But this is not an absolute impossibility; the student may suddenly have a vision and write more in three days than she has ever written in three months. In this paper I will not consider relative impossibilities, because they are not absolutely impossible: they are still possible. My paper is about worlds that are not possible with respect to an unrestricted notion of possibility.

I take impossible worlds to be worlds where logical, metaphysical or mathematical truths fail. A world is logically impossible when it fails to be governed by a classical logical system. This means that each logically impossible world either fails to be classically closed or fails to be classically consistent. An example of a logically impossible world is a world that makes true a proposition and its negation. Logically impossible worlds can be construed as ungoverned by any logical system, or they can be seen as governed by some nonclassical logical system (like paraconsistent logic.) But impossibility should be wider than logical impossibility, because necessity is narrower than logical necessity; some things are not possible, while they are still logical possible. Examples are a world where Saul Kripke is an earthworm or a world where the square root of 9 is 2. These are worlds where respectively metaphysical and mathematical truths fail. I think mathematical necessity and metaphysical necessity are unrestricted, like logical necessity, so false mathematical or metaphysical theories are part of impossible worlds.

In the next section, I will approach arguments for impossible worlds regardless of what their nature is. However, while developing my view with regard to the arguments, it turns out that there is no getting away from the question about nature. Therefore I will now introduce my view about the metaphysics of impossible worlds, which is an abstractionalist account. As I have already noticed, there are as many abstractionalist impossible worlds theories as there are ersatz theories of possible worlds. My view of impossible worlds corresponds to Vander Laan's actualist abstractionism.[5] According to my view, everything which exists, exists actually. Possible and impossible worlds are *maximal* states

[4] This statement is controversial, because some philosophers (Stalnaker and Van Inwagen for example) think possible worlds are not ways.

[5] Vander Laan, 1997.

of affairs. A state of affairs is maximal if and only if for every proposition P either P or its negation is true in it. Possible worlds constitute of possible states of affairs, while impossible worlds constitute of impossible states of affairs. States of affairs are to be individuated according to what is true in them, by their *propositional content*.[6] I will call the set of propositions which are true in some world w, the *book* on w: B_w.[7] The modal status of a world is connected with the status of its book. A world is actual if and only if the conjunction of propositions true in it is true, possible if and only if this conjunction is possibly true and necessary if and only if this conjunction is necessarily true.

[6] This is a term used by Vander Laan, 1997, p. 602.

[7] I have borrowed the term 'book' from Vander Laan, 1997, p. 602. However, Vander Laan applies the term book on state of affairs, while I apply it on worlds directly.

3 Arguments for impossible worlds

Possible worlds have proven useful for different purposes. As I already mentioned, they allow us to give metaphysical analyses of modal expressions and concepts like physical necessity, *de re* modal statements and deontic obligation. Another important application of possible worlds lies in the analysis of counterfactual conditionals. A counterfactual conditional, sometimes abbreviated to 'counterfactual', is a conditional statement indicating what would or might be the case if its antecedent were true. An example of a counterfactual, is:

(A) If you were to swim in the sea today, you would get a cold.

Impossible worlds have been proposed on the basis of the same kind of benefits; they simply improve possible worlds analyses. In this paper I will consider four metaphysical motivations behind impossible worlds. Two of them are directly related to possible worlds theories: impossible worlds provide a solution to the coarse-grainedness problem, and impossible worlds help us give a better semantic analysis of counterpossible conditionals. The other two motivations are considered apart from possible worlds theories: impossible worlds are needed for hypothetical reasoning and for disagreement about what is logically valid.

3.1 Coarse-grainedness

A strong motivation behind impossible worlds arises from the fact that insofar as possible worlds help us to individuate distinct propositions as distinct sets of possible worlds, the distinction they offer is too coarse-grained to adequately deal with intuitively distinct impossible propositions and with necessary propositions. For example, the propositions 'snow is both white and non-white' and 'the square root of 9 is 2' both are false at any possible world and thus both are identified with the empty set. But we are tempted to say that these propositions are clearly distinct. Examples of necessary propositions are '$2 + 2 = 4$' and '$5 + 5 = 10$'. Those propositions both are true at any possible world and thus both are identified with the same set including all possible worlds. However, these propositions seem to be clearly distinct. I will call this problem for possible worlds theories the 'coarse-grainedness problem'

Impossible worlds can be used to distinguish between intuitively distinct impossible propositions and between intuitively distinct necessary propositions. The impossible worlds where snow is both white and non-white do not necessarily coincide with the impossible worlds where the square root of 9 is 2. Likewise, the impossible worlds where $2 + 2 = 4$ do not necessarily coincide with the impossible worlds where $5 + 5 = 10$. Furthermore there are distinct but necessarily coextensive properties such as triangularity and trilaterality. Impossible worlds can be used to distinguish between those properties; some impossibilia will have triangularity but not trilaterality, and other impossibilia will have trilaterality but not triangularity. So impossible worlds can be used to solve the 'coarse-grainedness problem'. In his paper *Beyond Possible Worlds*,

Takashi Yagisawa discusses the 'coarse-grainedness problem' and argues that his account of extended modal realism solves this problem.[8] However, in this situation the actualist abstractionalist is in a better position than the extended modal realist. As Yagisawa acknowledges, the extended modal realist faces the problem that according to this proposal no property is ever identical with any property whatever:

> For any property P and any property Q, either it is possible for P and Q not to be coextensive or it is impossible. If it is possible, there is a possible world where P and Q are not coextensive. If it is impossible, there is an impossible world where P and Q are not coextensive. Either way, the set of all possibilia and impossibilia having P is different than the set of all possibilia and impossibilia having Q. Therefore, according to the above proposal, P and Q are not the same property. This is true for any P and Q whatsoever, including P and P.[9]

I will call this problem the 'fine-grainedness problem'. It is a big problem for the concretist who identifies properties with their extensions. For if it is true that the extension of P is not identical with itself in some impossible worlds, then the concretist has to accept it is true *simpliciter* that the extension of P is not identical with itself. In his paper *The Ontology of Impossible Worlds*, David vander Laan argues that this objection does not weigh against an actualist abstractionalist theory of impossible worlds.[10] It simply does not apply to actualist theories at all, because according to these theories, the set of possibilia and impossibilia having P contains only elements that actually have P. I think the crucial point is that the actualist argues that there are no impossible or merely possible entities that fill out the extensions of properties as the concretist requires. Therefore I agree with Vander Laan that the 'fine-grainedness problem' is not a problem for an actualist abstractionalist theory of impossible worlds.

Vander Laan even claims that the actualist takes advantage of impossible worlds when faced with the fine-grainedness problem. He suggests that property P is identical to property Q if the following biconditional holds:

(IC) For every possible or impossible world W, an entity X has P in W if and only if X has Q in W.[11]

Is Vander Laan's suggestion right? We have to distinguish two cases. Whenever P and Q name the same property, 'X has P' and 'X has Q' express the same proposition, so in this case IC holds. Whenever P and Q name different properties, 'X has P' and 'X has Q' express different propositions, and there will be some world in which one of the two is true and the other is false. In this latter case IC does not hold. This is exactly what we want. For example,

[8] Yagisawa, 1988, pp. 190–197.
[9] Yagisawa, 1988, p. 195.
[10] Vander Laan, 1997, pp. 609–612.
[11] Vander Laan, 1997, p. 610.

suppose that P is the predicate 'being a vixen' and Q is the predicate 'being a female fox'. Because 'being a vixen' and 'being a female' fox are just names of the same property, worlds where X is a vixen but not a female fox will be worlds where X is female fox but not a female fox, and X is a vixen but not a vixen. The same goes for worlds where X is a female fox but not a vixen. So (IC) holds, because 'X is a vixen' and 'X is a female fox' express the same proposition.

But when do P and Q name the same property? I think P and Q name the same property, when they are necessarily coreferential. Necessarily coextensive properties like 'being triangular' and 'being trilateral' are not necessarily coreferential, so there is an impossible world where some X is triangular and not trilateral, but at which it is not the case that X is not triangular and X is trilateral. So (IC) does not hold for necessarily coextensive properties, but it does hold for necessarily coreferential properties, and this is exactly what we want. On the basis of this analysis I think Vander Laan's suggestion is a good one. The actualist abstractionalist of impossible worlds is able to distinguish between distinct but necessarily coextensive properties such as 'triangularity' and 'trilaterality', but does not distinguish between necessarily coreferential properties such as 'vixen' and 'female fox'.

3.2 Counterpossibles

At the beginning of this section, I introduced the notion of a counterfactual conditional, and gave an example of such a conditional. David Lewis and Robert Stalnaker independently proposed analyses of the truth conditions for counterfactuals based on possible worlds and a relation of closeness among such worlds. Although their views differ in certain details, the fundamentals are the same. They can be described as follows:

Standard Account: A counterfactual, i.e., a sentence R of the form "If A were true, then B would be true", is true at a world β iff either A is impossible, or in set of closest possible worlds α to β where A is true, B is also true; otherwise R is false at β.[12]

Before I will consider the relation of closeness according to Lewis's and Stalnaker's view, I have to introduce some standard terminology. When considering a conditional, I will call any world where the antecedent is true an A-world, and the world at which the conditional is to be evaluated the 'base world'. Without loss of generality I will be taking the actual world to be the base world in this section. Lewis and Stalnaker both claim that the closeness relation among the A-worlds and the base worlds is primitive. Lewis thinks that closeness can best be explicated in terms of comparative overall similarity of possible worlds, and Stalnaker thinks closeness is best explicated in terms of minimal change. Below, I will discuss the notion of closeness in detail.

[12] I have used the notation of Goodman, 2004, p. 38. However, I have made some small modifications to make the account acceptable for both Stalnaker *and* Lewis.

However, this standard account (SA) does not handle well counterpossible conditionals. Counterpossible conditionals, sometimes abbreviated to 'counterpossibles', are counterfactuals which have metaphysically impossible antecedents. Counterlogical conditionals, or 'counterlogicals', form a subclass of counterpossibles; their antecedents are explicitly contradictory. SA wrongly implies that counterpossibles are vacuously true, while there are intuitively false counterpossibles. Take for example the counterpossible:

(B1) If snow were both white and non-white, I would be happy.

This counterpossible is plainly false, but possible worlds analyses render it vacuously true. Now consider the following counterpossible:

(B2) If snow were both white and non-white, classical logic would fail.

In contrary to the previous example, this counterpossible seems true, but it is counterintuitive to say it is *vacuously* true. The truth of this counterpossible seems to follow from the way its antecedent is connected to its consequent. We can imagine a fairytale in which snow is both white and non white, and in which *consequently* classical logic fails. As regards the intuitively false counterpossible (B1) there seems to be no connection between the antecedent and the consequent.

I will call this problem for possible worlds theories the 'problem of counterpossibles'. There is a simple solution to this problem: incorporating impossible worlds into our analysis. Dropping the restriction on the worlds that they are (metaphysically or logically) possible, gives us a more fine-grained account of the truth conditions of counterpossible conditionals. Still the core of SA can be preserved, together with all its virtues. So the problem of counterpossibles gives us an important motivation behind impossible worlds. Without choosing for the Lewisian or the Stalnakerian view, we can extend SA with impossible worlds, as follows:

Extended Standard Account: A counterfactual, i.e., a sentence R of the form "If A were true, then B would be true", is true at a world β iff in the set of closest possible or impossible worlds α to β where A is true, B is also true; otherwise R is false at β.[13]

In words, the basic idea is that a conditional is true if and only if at the closest possible or impossible world where the antecedent is true, the consequent is true as well. But what does the notion of "closest" mean? There is a lot to say about closeness; therefore I will spell out this notion in detail.

[13] I have used the notation of Goodman, 2004, p. 44. However, I have made some small modifications to make the account acceptable for both Stalnaker *and* Lewis.

3.2.1 Closeness

A natural way to order impossible worlds is to extend the traditional closeness relations between possible worlds. Lewis and Stalnaker both claim there is an ordering of all possible worlds with respect to closeness; worlds stand in similarity relations and similarity comes in degrees. According to Lewis's view, all possible worlds can be arranged around the actual world, in a system of nested spheres that mirror the degree of similarity of possible worlds with respect to the actual world.[14] For example, a world which is equal to the actual world except that I am wearing a red sweater instead of the yellow one I am actually wearing, is intuitively closer to the actual world than a world which is equal to the actual world except that I am immortal *mutatis mutandis*. These possible worlds are represented by different spheres in Lewis's system, and the sweater-sphere is smaller than the immortal-sphere.

Lewis and Stalnaker agree on two formal constraints on the ordening of possible worlds with respect to closeness.[15] These are the following:

Centering: No world α is closer to any world β than β itself.

Nesting: If some world α is closer to some world β than some world γ is to β, then γ is not closer to β than α is to β.

I also accept these constraints; I think *Centering* is very plausible and *Nesting* even seems to be an analytic truth. To the extent of this paper it is not worth it to discuss these conditions. In addition to these assumptions, Stalnaker makes the following further assumption about closeness among possible worlds which Lewis rejects:

Uniqueness: For any base world β, there always is a uniquely closest A-world.

According to Stalnaker, two A-worlds that are 'tied for closeness' are simply one and the same world. In his book *Counterfactuals*, Lewis notes that this assumption corresponds to a principle of counterfactuals, which is called the 'The Law of the Conditional Excluded Middle' (LCEM): *If P were true then Q would be true OR if P were true then Q would not be true.*[16] Lewis notes that this law is plausible, because it explains why we do not distinguish between the external negation of a whole conditional $\neg(P\Box\!\!\rightarrow Q)$ and the internal negation of the consequent $P\Box\!\!\rightarrow\neg Q$. However, Lewis provides a counterexample to this principle, which has become famous.[17] Consider the following counterfactuals:

(C1) If Bizet and Verdi were compatriots, then Bizet would be Italian.

[14] Lewis, 1973. pp. 4–19.
[15] Goodman, 2004, p. 42.
[16] Lewis, 1973, p. 79, and Goodman, 2004, p. 48.
[17] Lewis, 1973, p. 80.

(C2) If Bizet and Verdi were compatriots, then Bizet would not be Italian.

Remember I assumed the actual world to be the base world. Because in our actual world Bizet is French and Verdi is Italian, there is simply no reason to suppose that one consequent would be true rather than the other were the antecedent true. But intuitively C1 or C2 should be false, because they cannot both be true. Therefore C1 and C2 are both false, and LCEM should be rejected. I take this to be a great counterexample to LCEM. Furthermore, this example shows that a possible worlds account should admit that the possible world in which Bizet and Verdi are Italian compatriots is tied for closeness to the actual world with the possible world in which Bizet and Verdi are French compatriots.[18] I think this example proves that ties for closeness among possible worlds are possible. A same sort example can be given to prove the possibility of ties for closeness among impossible worlds. Therefore I will add *Uniqueness* to my account of possible and impossible worlds.

We should have a look at the following formal constraint regarding the relative distances of impossible worlds.

Strangeness of Impossibility: Any possible world is closer to the actual world than any impossible world.

This condition seems intuitively plausible to me. It seems like physical reality will be turned upside down, before logical laws and mathematical or metaphysical truths will fail. Moreover, with this condition the treatment of counterpossibles comes in as a natural extension of the standard Lewis-Stalnaker approach SA. When we look at Lewis's system of spheres I described above, *Strangeness of Impossibility* allows us to add further, larger spheres where worlds outside possibility stand. However, as Daniel Nolan argues in his article "Impossible Worlds: A Modest Approach", there are counterexamples to *Strangeness of Impossibility*. Slightly impossible worlds are to be preferred to some particularly bizarre possible worlds. He considers the following conditional:

> If intuitionistic logic came to be thought a much more satisfactory basis for mathematics by the experts, and if intuitionistic investigations led to breakthroughs in many areas of inquiry, and if important technological advances were made by the best minds in the field, which they would not have come to if they had been stuck in the rut of nonintuitionistic logic, then intuitionistic logic would turn out to be correct after all.[19]

In case *Strangeness of Impossibility* holds, the classical logician dismisses this conditional, because the closest possible or impossible world where the antecedent is true, is still a possible world where intuitionistic logic is incorrect. This example shows us the weakness of *Strangeness of Impossibility*. Because

[18] Goodman, 2004, p. 49.
[19] Nolan, 1997, p. 550.

of this weakness, Nolan proposes a weaker condition, which he calls *Lesser Strangeness of Impossibility.*[20] I will state this condition in my own words:

Lesser Strangeness of Impossibility: No impossible world is closer to the actual world than any possible world.

I am tempted to think this condition can deal with examples like the one Nolan considers. It allows the impossible world in which intuitionistic logic is correct to be tied for closeness with at least one possible world. (Remember, I admit ties for closeness.) This may not be the most desirable situation, but maybe it is the 'happy medium', because it is also very important to have some form of *Strangeness of Impossibility.* Accepting such a condition, ensures much more formal predictability for conditionals with possible antecedents.[21] As I already said, with *Strangeness of Impossibility* the treatment of counterpossibles comes in as a natural extension of SA. However, despite of the benefits of the *Lesser Strangeness of Impossibility* I think it still is problematic. In subsection 3.4, I will explain why I have to reject the condition.

The following constraint I will introduce, considers the ordering of possible and impossible worlds. I did not find a condition like this in the literature on impossible worlds. However, I think it is a very plausible condition which should definitely be considered. The condition is based on my account of possible and impossible worlds, which I introduced in section 2. Remember that according to this account, the *book* on some world is the set of propositions which are true in that world.

Matching books: Take α, β and γ to be distinct worlds. World α is closer to world β than world γ is, if the next two conditions are satisfied:

1. The union of the books on α and β is the same or smaller than the union of the books on β and γ.

2. The intersection of the books on α and β is the same or bigger than the intersection of the books on β and γ.

Notice that because α, β and γ are distinct worlds, in at least one of those two conditions the inequality is strict.

In words, this constraint concerns the differences between the books on different worlds. It corresponds to my intuition that worlds which have books that have many differences are further away from each other than worlds which have books that have little differences. *Matching books* offers a very plausible systematic analysis of closeness between worlds. This constraint is particularly interesting when it comes to the internal ordering of impossible worlds.

[20] Nolan, 1997, p. 566.
[21] Nolan, 1997, p. 566.

I will use *Matching books* while evaluating some constraints about the ordering of impossible worlds which were suggested by Nolan.[22] First I have to introduce some new terminology. The Absurd World is the world in which every proposition whatsoever is true. Assuming there are no distinct, indiscernible impossible worlds, the Absurd World is the only impossible world that is classically closed. The following constraint is suggested by Nolan, and named by Goodman:

Extremity of the Absurd: The Absurd World is the furthest impossible world from the actual world.[23]

I will consider this suggestion together with another one, which is also suggested by Nolan and named by Goodman:

Logical Nearness: Impossible 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.[24]

I think these latter conditions, *Extremity of the Absurd* and *Logical Nearness*, are in tension, because the Absurd World is the only impossible world in which no violations to the logic of our actual world occur. Indeed, all impossible worlds but the Absurd World fail to be classically closed. So according to *Logical Nearness*, the Absurd World is closer to the actual world than any other impossible world, which contradicts *Extremity of the Absurd* immediately. So at least one of the two conditions should be rejected.

I doubt if the Absurd World is the furthest impossible world from the actual world. Consider the world in which:

1. Every proposition P is true exactly when $\neg P$ is true in the actual world, and every proposition $\neg P$ is true exactly when P is true in the actual world.

2. If P is true in the actual world, then P is not true and if $\neg P$ is true in the actual world, then $\neg P$ is not true.

Call this world the 'Converse World'. The Converse World is an impossible world, according to my account of impossible worlds proposed in section 2. At the Converse World all metaphysical and mathematical truths fail. For example, in the Converse World Saul Kripke is an earthworm and the square root of 9 is 2. But Saul Kripke is also a poached egg, and the square root of 9 is also 6 (and also 100, and also 46, etc.) Indeed, every proposition about Saul Kripke and every proposition about the square root of 9 is true, except the propositions which are true in our actual world.

If we now look at the Absurd World (α), the actual world (β) and the Converse World (γ) it follows from *Matching books* that the Absurd World is

[22] Nolan, 1997, pp. 545-548.
[23] Goodman, 2004, p. 54.
[24] Goodman, 2004, p. 54.

definitely closer to the actual world than the Converse World is. I will call the book on the Absurd World B_{ab}, the book on the actual world B_{ac} and the book on the Converse World B_{co}. The union of B_{ab} and B_{ac} is B_{ab}, and the union of B_{ac} and B_{co} is the set of every possible proposition P and its negation $\neg P$ (by definition of the Converse World). This latter set is exactly B_{ab} so both unions are the exactly the same. Thus condition 1 is satisfied. The intersection of B_{ab} and B_{ac} is the whole of B_{ac}, because every proposition P or negation $\neg P$ which is true in the actual world, is also true in the Absurd World (by definition of the Absurd World). On the contrary, the intersection of B_{ac} and B_{co} is the empty set (by definition of the Converse World). So the intersection of B_{ab} and B_{ac} is bigger than the intersection of B_{ac} and B_{co} and thus condition 2 is also satisfied. I conclude that the Absurd World is closer to the actual world than the Converse World is, therefore the Absurd World is not the impossible world which is the furthest away from the actual word and thus *Extremity of the Absurd* should be rejected.

We can even say that the Converse World is the *furthest* away from the actual world. Take α to be some possible or impossible world which is not the Converse World. The union B_α and B_{ac} can not be bigger than B_{ab}, because the Absurd World already contains every proposition. As I have shown above, the union of B_{ac} and B_{co} is exactly B_{ab} so both unions are at most same, thus condition 1 is satisfied. The intersection of B_α and B_{ac} can not be the empty set, because α must have a proposition in common with the actual world, otherwise it is the Converse World. And because the intersection of B_{ac} and B_{co} is the empty set, the intersection of B_α and B_{ac} is bigger than the intersection of B_{ac} and B_{co} and thus condition 2 is also satisfied. So α indeed comes out closer than the Converse World, thus the Converse world is the furthest away from the actual world. I will state this outcome in the following constraint.

Extremity of the Converse World: The Converse World is the furthest impossible world from the actual world.

However, there is also a problem with *Logical Nearness*, namely that the Absurd World turns out to be the closest impossible world (as I have already mentioned.) For all the other worlds, I think *Logical Nearness* is very plausible. I would like to propose a weaker version of the condition, as follows:

Restricted Logical Nearness: Take α and β to be distinct impossible worlds which are not classically closed. World α is closer to the actual world than world β, if there are less inconsistencies at world α than at world β.

However, it is questionable whether we can combine *Matching books* and *Restricted Logical Nearness* in one thorough account. It seems like we have to give priority to one of those constraints, or we should make a synthesis between the constraints in one form or another. For now, I think *Matching books* is the most promising constraint on closeness; it corresponds to my intuitions and to *Extremity of the Converse World*. But more research should be done on a

synthesis of *Matching books* and *Restricted Logical Nearness*.

3.3 Hypothetical reasoning

This way of analyzing counterpossible conditionals can be extended to hypothetical reasoning. I will start with an example of this kind of reasoning. Suppose a proponent of constructive mathematics denies the Axiom of Choice[25], but still wants to reason from the hypothesis that the Axiom of Choice is true. He reasons as follows:

1. If the Axiom of Choice were true, then it would be applicable on non-countable sets.

2. If the Axiom of Choice were applicable on non-countable sets, then it would be the case that a solid sphere can be dissected into a finite number of pieces that can be reassembled into two solid spheres with the same volume as the original.[26]

3. If it were the case that a solid sphere can be dissected into a finite number of pieces that can be reassembled into two solid spheres with the same volume as the original, physics would be turned upside down.

4. If the Axiom of Choice were true, physics would be turned upside down.

In this example, the mathematician starts with the hypothesis that the Axiom of Choice is true, and with counterpossible reasoning he gets new starting points (which are the antecedents of the counterpossible sentences.) More precisely: he accepts new conditionals on the basis of accepting old conditionals whose antecedents and consequents are related. He concludes that if the Axiom of Choice were true, physics would be turned upside down. This kind of reasoning is very important in philosophy; I think it is in the nature of philosophers to execute thought experiments about impossibilities, which is in fact equal to hypothetical reasoning.

In his paper "Impossible Worlds: A Modest Approach" Daniel Nolan argues that philosophers implicitly reason with counterpossibles when they reason hypothetically.[27] He distinguishes hypothetical reasoning from categorical reasoning, because of the following differences. In categorical reasoning, a starting point is introduced if it is known to be true. Hypothetical reasoning is less strict with regard to starting points, in that propositions can be taken as starting points whether they are thought to be true or not.[28] For example, if I am attempting to work out how things would have gone if Lewis has argued for an actualist abstractionalist theory of possible worlds, I can take as starting points claims about intentions of Lewis which I do not in fact believe. However,

[25] The version of the Axiom of Choice I am talking about only holds for classical logic.
[26] This is known as the Banach-Tarski paradox.
[27] Nolan, 1997, pp. 557–561.
[28] Nolan, 1997, p. 558.

hypothetical reasoning is more strict with regard to starting points, in that unrestricted importation of things known to be true can disrupt the exploration of the hypothesis, or even change the subject.[29] Some claims which I take to be true about Lewis contradict my assumption that Lewis did argue for abstract possible worlds; an example is the claim that Lewis believed the actual and the concrete do not coincide. Those claims are inadmissible into the "what if", while they are perfectly entertainable in categorical reasoning. So some claims known to be false are admissible, and some known facts are inadmissible as starting points in hypothetical reasoning.

When we are reasoning on the basis of a hypothesis which we do not accept, we consider the closest world where the starting points are true. Other truths at this world are admissible as new starting points in our hypothetical reasoning. Using this method we only get new starting points which do not change the subject. At the closest world in which Lewis has argued for an actualist abstractionalist theory, the statement that he believed the actual and the concrete do not coincide, is simply false.

Another virtue of hypothetical reasoning, is that it provides a way for all sorts of logicians to assess the commitments of theories they take to be impossible. If it can be shown that a commitment of a theory is unacceptable, this will serve as a better refutation of the theory than showing that a logical consequence is unacceptable.[30] The proponent of constructive mathematics, in my example above, can use the hypothetical reasoning to refute the Axiom of Choice. Indeed, the commitment 'physics would be turned upside down' can be judged as unacceptable.

3.4 Disagreement

It is a fact that not all logicians agree about which logic is correct; for example intuitionists and nonintuitionists disagree about whether the Law of the Excluded Middle and the Double Negation Elimination are axioms. However, nonintuitionists can understand intuitionism, and vice versa (this statement is somewhat controversial, see below for a discussion.) But intuitionism and nonintuitionism can not both be true, because they contradict each other. If intuitionism were true, nonintuitionism would be impossible, and vice versa. To state it more general: if there are disagreements in logic, and at most one of the parties to the various disputes can be correct, then it seems that the other parties are reasoning about, and believing in, impossibilities. And to understand implications of their rivals' view, logicians need to be able to consider and usefully reason about logical impossibilities. This argument is posed by Daniel Nolan in his paper "Impossible Worlds: A Modest Approach".[31] The argument is also applicable to mathematics and metaphysics; in these fields it is also important to have the ability to coherently discuss impossibilities.

[29] Nolan, 1997, p. 558.
[30] Nolan, 1997, p. 560.
[31] Nolan, 1997, pp. 536–541.

There are several responses to this argument. One way is to say that weaker logics are merely leaving out some of the principles which could have been added. But this is not a satisfactionary strategy, because different logicians can clearly believe opposite principles. Another strategy is to say that it is not true that nonintuitionists understand intuitionists because those intuitionists are talking incoherent nonsense, or vice versa. According to Nolan, this strategy is not common these days, because it makes no claims which are even assessable for plausibility. I agree with Nolan; intuitionistic logic is a consistent and coherent logical system. To label it as nonsense is to avoid the metaphysical debate. Furthermore, this objection is a stronger version of the Quinean reply, and therefore I will spend my time on considering this latter reply.

The Quinean reply can be abbreviated in his famous motto that "'to change the logic is to change the subject". Quine states that apparently disagreeing logicians are actually talking about different things. According to this strategy, when intuitionists deny that the Law of the Excluded Middle holds, they are actually talking about provability instead of truth. I think Quines statement makes nonsense of too many disputes between classical logicians, intuitionists, paraconsistentists, etc. It is much more fruitful to accept that there are genuine disputes in logic by assuming that each logical party understands the rival logical parties as comprehensible but impossible theories. When we assume that the true logic is the classical logic, we can still reason counterpossibly about situations in which a certain non-classical logic were correct. The sentence "If intuitionism were correct, the Law of the Excluded Middle would fail" seems correct to me and furthermore valuable. When considering this sentence, we take the assumptions of intuitionism to be true and then start to reason about the Law of the Excluded Middle. Our conclusion will be that this law does not hold in non-finitary contexts. This conclusion will confirm the truth of the sentence "If intuitionism were correct, the Law of the Excluded Middle would fail".

From these responses we have learned that my argument is this section is only correct when there are genuine, meaningful disagreements. Not all disagreements are of this kind, but we know there are these sort of disagreements, so this argument for impossible worlds seems valid to me. However, there is still something to say about the Quinean reply. If classical logic is the correct logic, then possible worlds should be closed under classical logic, and the rest will be impossible. But the intuitionist takes possible worlds to be closed under intuitionistic logic and thus he reasons differently about counterpossibles. So Quine could argue that the classical logician and the intuitionist are reasoning about counterpossibles from a different point of view. This is only a problem, if our account of closeness treats possible and impossible worlds differently. So the only condition which is problematic, is the *Lesser Strangeness of Impossibility* condition. Unfortunately we must yet reject this condition, despite of its benefits. It is the only way to face this Quinean objection.

The argument of disagreement is also applicable to mathematics and metaphysics; in these fields it is also important to have the ability to coherently discuss impossibilities. Different set theorist have different views on for exam-

ple the Axiom of Choice. Assuming that there is one true universe of sets, at most one of the different set-theoretical parties can be correct. Still mathematicians are able to work under the assumption that some rival set-theoretic principle holds and reason coherently from this assumption. False mathematical theories are just parts of impossible theories. In metaphysics, there are also a lot of disputes. Metaphysicians often evaluate metaphysical theories which they consider to be wrong. With the following example, I will stay close to actuality: in this paper I am evaluating modal realism which I consider to be wrong. Nevertheless, I have to consider situations in which modal realism is correct, and reason about such situations as if they were actual. So I am actually considering impossible worlds.

4 Arguments against impossible worlds

In this section I will consider arguments against impossible worlds. There are several difficulties of impossible worlds theories. First I will consider objections to the idea that counterlogicals, or even all counterpossibles, can be false. These are the most important objections to discuss, because I take the argument of analyzing counterpossibles to be a very strong argument for impossible worlds. However, this argument is based on intuitions about the falsehood of some counterpossibles, and appeals to intuitions can be dangerous in metaphysics. Therefore I will discuss the objections against the analysis with counterpossibles in depth. Furthermore, I will discuss general criticism of impossible worlds, repectively one objection from Lewis and one from Stalnaker. Finally I will discuss an objection which is specifically aimed at ersatzers.

4.1 Arguments against counterpossibles

Some philosophers are objecting to the idea that counterlogicals, or even counterpossibles, can be false or non-trivially true. Goodman argues in his paper "An extended Lewis/Stalnaker semantics and the new problem of counterpossibles"[32] that counterlogicals are always vacuously true. He has no real motivation for this:

> However, I do not think there are any false counterlogicals. I have found no sound arguments to support this claim. I simply find myself inclined to accept some version of David Lewis's 'shrug defense' when it comes to resisting (allegedly false) counterlogicals as counterexamples to SA.[33]

I think counterlogicals in which the antecedent is not connected to the consequent, are clear examples of false counterlogicals. The following counterlogical, which I considered earlier in my paper, is one of those clear examples:

(B1) If snow were both white and non-white, I would be happy.

There are also counterlogicals in which the antecedent has a connection with the consequent, but nevertheless the counterlogical is intuitively false. Consider for example the following counterlogical:

(D) If Barack Obama were both human and non human, then he would have lost the 2008 election.

Being a non human does not change something to Obama's ability to become a president, for he still is a human. I guess there exists a rule that to participate to the presidential election of the United States of America, you must be human.

[32] Goodman, 2004.
[33] Goodman, 2004, p. 52.

But there seem to be no rule that people who are non humans are excluded from participation to this election. Counterpossibles in which the antecedent is connected to the consequent, should be analyzed in the same way as we analyze other counterpossibles, even when their antecedents are explicitly contradictory.

My view is underpinned by Graham Priest's paper "Sylvan's Box: a Short Story and Ten Morals". This paper contains a short story which is essentially inconsistent, but perfectly intelligible. In the story, Priest visits the home of his deceased friend Richard Sylvan, to help Nick Griffin, Sylvan's literary executor, with organizing the finished and unfinished papers belonging to Sylvan's intellectual legacy. Priest discovers a mysterious box that once belonged to Sylvan, which is absolutely empty and simultaneously has an object inside it. Priest is stunned, and after a while he gives the box to Griffin. Griffin is also amazed, and the two men start thinking about this incredible finding. Priest and Griffin immediately agree they should make the box public, but when they start to think about the possible consequences of such an action, they change their minds. In the end of the story some incredible and impossible things happen, but even when the inconsistent is involved, people act in intelligible ways. Specifically, the story never says that Priest shoots the box off into space. Priest notes that to understand the story one has to draw inferences from what is explicitly presented and from the background information. If the deductive canons employed were those of classical logic, we could infer from the description of the box that it was shot off into space.[34] Indeed, according to classical logic, the following counterpossible is vacuously true:

(E) If the box Priest had found were both empty and non-empty, then he would have shot it off into space.

This is also a counterlogical in which the antecedent is connected to the consequent. However, it is intuitively false. My thoughts about the story and (E) are as follows. We should apply the analysis of counterpossibles as stated above on this example. So we are looking for the impossible world in which the box Priest had found were both empty and non-empty, which is the closest to our base world. (In this case the base world is the story.) For simplicity I give this closest world the name i. In this world i the Law of Explosion (or Ex Contradictione Quodlibet) fails to hold, but other laws of classical logic still hold. Because of the failure of the Law of Explosion in i, at i it is not the case that Priest shoots the box off into space. Therefore (E) should be rejected. I think this account of analyzing counterlogicals is very attractive, and Goodman has no good reason to reject it.

In his book *The Philosophy of Philosophy* Tim Williamson objects to the idea that counterpossibles can be false or non-trivially true.[35] He considers the following counterlogical:

[34] Priest, 1997, p. 580.
[35] Williamson, 2007.

(1) If $5 + 7$ where 13, then $5 + 6$ would be 12.

Intuitively this is a non-vacuously true counterpossible. Williamson argues that a consequence of this non-vacuously truth would be that $5 + 5 = 11$ and that $5 + 4 = 10$, and that $4 + 4 = 9$, and that $4 + 3 = 8$ and that ... and that $1 + 0 = 2$, and that $0 + 0 = 1$. So a consequence would be that $0 = 1$. Therefore the following counterfactual would be a consequence of (1):

(2) If the number of answers I gave to a given question were 0, then the number of answers I gave would be 1.

Note that (2) is not a counterpossible, because there is a possible world where I give no answers to a given question. This counterfactual is not only intuitively false, but according to a possible world theory based on closeness, it would also be considered false.

My answer to Williamson's objection, is as follows. It seems like Williamson does not hold the context fixed while reasoning. The context at which (1) comes out true is one at which the closest antecedent world (or closest antecedent worlds) is impossible. However, the context at which (2) comes out false is one at which the closest antecedent world (or closest antecedent worlds) is possible; of course there is no possible world at which 0 is equal to 1. It is not valid to change the context while reasoning, so Williamson should change (2) into the following counterpossible:

(2') If 0 were equal to 1 and the number of answers I gave to a given question were 0, then the number of answers I gave would be 1.

This counterpossible is intuitively true, and will also be deemed true according to the Extended Standard Account for evaluating counterpossibles. Of course at the closest impossible world in which 0 equals 1 and the number of answers I gave to a give question were 0, it is also the case that I gave 1 answer to that question.

4.2 Lewis's rejection of impossible worlds

Lewis rejects impossible worlds in a footnote in his book *On the Plurality of Worlds*.

> This discussion of restricting modifiers enables me to say why I have no use for impossible worlds, on a par with the possible worlds. For comparison, suppose travellers told of a space in this world - a marvellous mountain, far away in the bush - where contradictions are true. Allegedly we have truths of the form 'On the mountain both P and not P'. But if 'on the mountain' is a restricting modifier, which works by limiting domains of implicit and explicit quantification to a certain part of all that is, then it has no effect on the truth-

> functional connectives. Then the order of modifier and connective makes no difference...[T]he alledged truth 'On the mountain P and not P' is equivalent to the overt contradiction 'On the mountain P, and not: on the mountain P'... But there is no subject matter, however marvellous, about which you can tell the truth by contradicting yourself. Therefore there is no mountain where contradictions are true.[36]

Lewis's objection is based on the assumption that 'at so-and-so world' is a restricting modifier. If 'at so-and-so world' works as a restricting modifier, it restricts the quantifiers within its scope to parts of that world, and this implies that it can distribute through the truth-functional connectives. So the general assumption that 'at so-and-so world' works as a restricting modifier implies the following assumption:

(F) At world $w : (A \wedge \neg A)$ entails 'At world $w : A \wedge (\neg$ At world $w : A)$'

Any inconsistency at some impossible world entails an inconsistency at the actual world, which is indeed a big problem. But I think this is only a problem for a concretist theory of impossible worlds. A supporter of an actualist abstractionalist theory of impossible worlds is not confronted with this problem. Lewis himself seems to admit this:

> If worlds were like stories or story-tellers, there would indeed be room for worlds according to which contradictions are true. The sad truth about the prevarications of these worlds would not itself be contradictory.[37]

This quote shows exactly why Lewis's objection to impossible worlds is no problem for the extended (linguistic) ersatzer. When we take worlds to be maximal states of affairs, we should use the modifier 'at so-and-so world' like we use 'in such-and-such story'. Indeed, states of affairs represent things as having certain properties and standing in certain relations. Fortunately, in state of affairs the following holds:

(G) In state of affairs $S, \neg A$ does not entail 'Not: in state of affairs S, A'

So if impossible worlds were world-books, they can represent impossibilities without actually instantiating them.[38]

[36] Lewis, 1986, p. 7n.

[37] Lewis, 1986, p. 7n.

[38] My view corresponds to Vander Laan's reply, which can be found in Vander Laan, 1997, p. 606.

4.3 Stalnaker's objection to impossible worlds

In his paper "Impossibilities"[39] Stalnaker also objects against impossible worlds. His objection is connected to Lewis's argument against impossible worlds; it considers the semantics of negation.[40] The classical semantic clause for negation is that $\neg A$ is true if and only if A is not true. If we allow worlds at which both A and $\neg A$ are true, Stalnaker says, the semantics of negation may be violated. According to Stalnaker, revising the semantics of negation is not an option, because it is a very basic operator.

My reply to Stalnaker is that an impossible world may render true a contradiction, but this does not imply that 'not' means something different at this world. This world can simply behave differently when it comes to negation. But the way in which our actual world, or any possible world, behaves when it comes to negation, is not affected by the fact that at some impossible world some sentence can hold together with its negation. Furthermore I think Stalnaker's objection is not a specific objection against impossible worlds. The same sort of objection can be applied to possible world theories. For example, consider a possible world in which snow is black. This world will make some propositions about snow true, which our actual world does not make true, for example 'snow is black'. If Stalnaker's objection against impossible worlds were correct, we should also conclude that the word 'snow' means something different at the possible world which I just introduced. However, I think this is not the case; the world simply behaves differently when it comes to snow.

4.4 Primitive modality

The 'no primitive modality' objection is originally an objection of Lewis against ersatzism. It still is an important aspect of the debate between concretists and abstractionists of possible worlds. According to Lewis, ersatzers must invoke primitive modality in one form or another, which is problematic.[41] The rejection of primitive modality is a central tenet of Lewisian possible world approaches; it motivates taking possible worlds to be concrete. According to Lewis, our total theory about logical space should be stated without recourse to modal notions.

However, the 'no primitive modality' objection seem to be even stronger for extended ersatzism. All ersatzers agree on the following:

(H1) It is possible that P if and only if there is a world according to which P.

Once we admit impossible worlds to our metaphysics, this biconditional becomes false from the left side to the right side. To fix this falsehood, we have to adjust this clause, as follows:

[39] Stalnaker, 1996.
[40] Stalnaker, 1996.
[41] Lewis, 1986, pp. 13–14, 150–157, 167–170.

(H2) It is possible that P if and only if there is a possible world according to which P.

But now possibility is expressed in terms of possibility, and thus primitive modality is invoked in this clause.

I am wondering if it is such a terrible thing that ersatz possible and impossible worlds theories take modality as primitive. Some philosophers have stated that the modal is mysterious because we have no empirical access to it.[42] But Lewis's reduction of the modal to the non-modal also does not solve that problem. Lewis's concrete worlds are also not empirically accessible. There are even philosophers who think that Lewis too must use primitive modality in his theory.[43] Furthermore, why should the modal be grounded in the non modal? I am wondering what is wrong with the view that possibility, necessity and other modal notions form a circle of interrelated concepts, and the use of some modal notion should be coherent with regard to the other notions in the circle.

[42] Melia, 2008, p. 147.
[43] Shalkowski, 1994.

5 Conclusion

In my paper I have considered several arguments for impossible worlds. First, incorporating impossible worlds to our metaphysics solves the coarse-grainedness problem of possible world theories. Although extended modal realism is one of the theories that solves this coarse-grainedness problem, it faces a new problem, namely the problem of fine-grainedness. This is a very important reason to take the actualist abstractionalist theory of impossible worlds to be the most promising.

Second, impossible worlds help us give a better semantic analysis of counterpossible conditionals. The standard possible world account wrongly implies that counterpossibles are vacuously true, while there are obviously false counterpossibles. By extending the standard account to impossible worlds, this problem is solved. The basic idea is that a conditional is true if and only if at the closest possible or impossible world where the antecedent is true, the consequent is true as well. However, it is crucial to spell out the relation of 'closeness' in detail. In my paper I have analyzed different constraints with regard to closeness. Although a thorough account is not yet forthcoming, we should continue the debate on those conditions. I believe this to be one of the crucial points of the debate on impossible worlds.

The third argument I considered, concerns hypothetical reasoning. By allowing impossible worlds into our metaphysics we are able to give a fertile account of hypothetical reasoning. When we are reasoning on the basis of a hypothesis which we take to be impossible, we consider the closest impossible world where the starting points are true. Other truths at this world are admissible as new starting points in our hypothetical reasoning. The biggest virtue of this account, is that it provides a way for all sorts of logicians to assess the commitments of theories they take to be impossible.

Fourth, disagreeing logicians (and mathematicians, and metaphysicists) need impossible worlds. If there are genuine, meaningful disagreements in logic, and at most one of the parties to the various disputes can be correct, the other parties are reasoning about, and believing in, impossibilities. Furthermore, to understand implications of their rivals' view, logicians need to be able to consider and usefully reason about logical impossibilities. In this subsection I have given much attention to a reply made by Quine. Because of this objection, I have even changed my account of closeness.

I have considered objections to my view in section 4. First I discussed objections to the idea that counterlogicals, or even counterpossibles, can be false. Then I discussed two general objections against impossible worlds, respectively stated by Lewis and Stalnaker. Finally, I considered the 'no primitive modality' objection. I must admit ersatzism, and especially extended ersatzism, invokes primitive modality, but I think this does not have to be a problem.

6 References

- Copeland, "The Genesis of Possible World Semantics", *Journal of Philosophical Logic*, vol. 31, 2002, pp. 99-137.
- Goodman, Jeffrey. "An extended Lewis/Stalnaker semantics and the new problem of counterpossibles", *Philosophical Papers*, vol. 33, 2004, pp. 35–66.
- Lewis, David. *Counterfactuals*, Oxford: Blackwell, 1973.
- Lewis, David. *On the Plurality of Worlds*, Oxford: Blackwell, 1986.
- Melia, Joseph. "Ersatz Possible Worlds", *Contemporary Debates in Metaphysics*, ed. Matthias Steup and Ernest Sosa. Malden, MA [etc.]: Blackwell, 2008, pp. 135–151.
- Nolan, Daniel. "Impossible Worlds: A Modest Approach", *Notre Dame Journal of Formal Logic*, vol. 38 (issue 4), 1997, pp. 535–572.
- Priest, Graham. "Sylvan's Box: A Short Story and Ten Morals", *Notre Dame Journal of Formal Logic*, vol. 38 (issue 4), 1997, pp. 573–582.
- Vander Laan, David. "The Ontology of Impossible Worlds", *Notre Dame Journal of Formal Logic*, vol. 38 (issue 4), 1997, pp. 597–620.
- Shalkowski, Scott. "The Ontological Ground of the Alethic Modality". *The Philosophical Review*, vol. 103 (issue 4), 1994, pp. 669–688.
- Stalnaker, Robert. "Impossibilities", *Philosophical Topics*, vol. 24, 1996, pp. 193–204.
- Williamson, Tim. *The Philosophy of Philosophy*, Oxford: Blackwell, 2007.
- Yagisawa, Takashi. "Beyond possible worlds", *Philosophical Studies*, vol. 53, 1988, pp. 175–204.