South American Journal of Logic
Vol. X, n. X, pp. 1–20, forthcoming
ISSN: 2446-6719
Please quote published version!

Possible arguments against Impossible Worlds in Wittgenstein's *Tractatus*[1]

Til Eyinck

Abstract

At first glance, impossible world semantics appear to be useful adaptations of normal modal logic. Proponents of impossible worlds argue, e.g., that in the context of metaphysical disagreement, impossible worlds would provide a key to modelling the respective dispute situation. The same philosophers also argue that we need impossible worlds to model what they consider to be the conceivability of logical impossibilities. With the help of Wittgenstein's early philosophy, or, better said, with what I believe to be a visualisation of his ideas on nesting models for possibility, I develop some thoughts on the likely shared weaknesses of such arguments for non-normal world approaches. Said goes hand in hand with the supplementary proposal of rethinking our conventional nesting models for possibility.

Keywords: possibility nesting, normal modal logic, metaphysical disagreement, conceivability.

Introduction

Our motto might be: 'Let us not be bewitched'.
(L. Wittgenstein, Zettel, 199e)

Some authors believe that there are good reasons to include non-normal modal logic in our philosophical toolbox. Contrary to the view adopted here, they believe that there are strong reasons in favour of impossible world approaches

[1]This essay is essentially a transcription of my talk 'Impossible worlds? *Let us not be bewitched*' given at the SALOME 1 (UNSAAC, Cuzco, January 2024). The essay serves in part as a de-formalization, attempted sharpening, and re-embedding of the arguments of my paper 'Should we embrace impossible worlds due to the flaws of normal modal logic?' [9]. I would therefore like to apologize to the attentive reader that it was not possible to thoroughly indicate every detail that I borrow from the mentioned paper.

– which in their view represent something like a gateway to 'the impossible'. If one looks a little closer, however, it can be said that they assume impossible worlds for various reasons: because they believe that human conceivability is tied to the logically impossible; or that metaphysical debates require a notion of logical impossibility; or that the problem of logical omniscience within the framework of normal modal logic can only be solved adopting impossible world frameworks; or that impossible worlds are helpful to distinguish between intuitively different counterpossible conditionals [2, 3, 5, 19]; among other reasons.

Recently, I argued that some of such prominent arguments, often relying on alleged weaknesses of normal modal logic (that is, approaches relying on possible worlds only), *fail* to provide strong reasons for non-normal approaches (that is, approaches relying also on impossible worlds). I suggested that there are likely common problems with different arguments for impossible worlds [9]. The present essay can be understood as a continuation of this interest to search for general reasons behind the suspected failure of arguments for impossible worlds. For it is one thing to show that some of those arguments are not strong enough to introduce impossible worlds for their sake, and another to attempt to specify the philosophically problematic connection between their (likely) interrelated weaknesses. In addition to this methodical starting point described above, however, there is another reason why I am criticizing arguments for impossible worlds and not impossible worlds 'directly': If one does not believe, as I do, that 'the (logically) impossible' is conceivable or a necessary aspect of our philosophical argumentation, then one should not talk about what the impossible is or how it could be modelled. That is why I try to limit my work to the best of my ability to criticizing arguments for the necessity of impossible world semantics, understood as a set of logical theories – whatever 'the impossible' might or might not designate.

In what follows, I try to make use of some aspects of Wittgenstein's early philosophy[2] to develop what I believe to be the beginning of a path to a partial but potentially overarching explanation of why many canonical arguments for impossible worlds fail – especially those that work via the alleged deficiency of possible worlds-only approaches. Consequently, it is only reasonable to divide this essay into three main parts: (a) an introduction to the topic of possible and impossible worlds (Section 1), (b) an elaboration of some aspects of Wittgenstein's early philosophy that are useful for the present concerns (Section 2), and (c) reflections on possible connections between (a) and (b), hopefully leading to an approach of a more general criticism of different arguments for impossible worlds (Section 3). This essay is a work to be located in the history of

[2] I will mainly rely on ideas related to his *Tractatus Logico-Philosophicus* [25], which I will from now refer to as *Tractatus*. I will be referring to the German original. If I had to choose an English translation, it would be the Pears/McGuinness version and not the one by Ogden.

logic, insofar as an attempt is made to adopt a critically-appreciative attitude towards a historical theory, linking it with current issues. It is a work of logic, insofar as it thereby tries to contribute something to contemporary issues.[3]

1 Normal (and non-normal) modal logic

Normal modal logic can be understood as a particularly well-behaved fragment of first order predicate logic [cf. 10]. To define possibility ($\Diamond$) in a Kripke-style model of normal modal logic, a model *only* containing possible worlds, we usually proceed as follows: we give truth conditions for formulas of type $\Diamond\varphi$ by defining a model M, a non-empty set W containing some possible worlds w, an accessibility relation R, and a valuation function V (R defines the world-to-world relations of M and V assigns truth values to the atoms of M). Thus, if the whole model structure $M = \langle W, R, V\rangle$ is given, one can say that: $\Diamond\varphi$ (equivalent to $\neg\Box\neg\varphi$) iff there is at least one world w in W, where the assigned truth value for φ is TRUE [4, pp. 95–8]. Normal modal logic's triumphant success in contemporary philosophy comes quite indisputably from the fact that there are various very compelling interpretations of its operators.

We can interpret the operators $\Diamond$ and $\Box$ of normal modal logic in many philosophically helpful ways. In the epistemic reading, for example, we take the worlds to represent epistemic states of an agent who may or may not have access to them. Thus, the operator K – analogous to $\Box$ – stands for 'knowledge' and B – analogous to $\Diamond$ – stands for 'belief'. So if there is at least one epistemic state – an epistemically accessible world – for which φ is TRUE, then we are talking about the agent modelled in this way believing φ. In a deontic reading, we interpret $\Box$ as 'ought' and $\Diamond$ as 'can' – and so on. It is important to understand that although the operators *look* different in the context of such interpretations, formally speaking they do not *function* differently [cf. 12, sec. 8.2–8.3; 17, sec. 9.2]. Normal modal logic is versatile in this respect.

Impossible world-approaches generally adapt the normal modal logic that we have just sketched and add some extra formal structure to it. Thus, a (simplified) model M' for possible and impossible worlds could be $M' = \langle W, N, R, I\rangle$ [22, p. 490]. Here N is a proper subset of W, containing the normal (possible) worlds, whereas the non-normal (impossible) worlds are not in N. Non-normal worlds are such worlds where the most basic logical truths, like the law of non-contradiction, can fail to hold. R is a ternary world-to-world relation and I is a map from propositions[4] to sets of worlds. R at normal worlds is binary, whilst

[3] My approach of conducting a historical dialogue that is not too optimistic but nevertheless as helpful as possible could not be described any better than Perler et al. do in [18].

[4] Unlike proponents of such approaches, I do not believe that we are dealing with propositions here, as I do not believe that propositions can be mapped to worlds or sets of worlds.

it is a ternary relation in the case of non-normal worlds. This kind of setup allows logicians, for example, to interpret certain counterfactual conditionals (namely, counterpossible conditionals) by evaluating the antecedent and the consequent at respectively different worlds [22, p. 492–3]. Section 1.1 provides a more detailed explanation of this line of argument for the extension of normal modal logic with non-normal worlds.

1.1 Arguments for impossible worlds (on the grounds of alleged flaws of normal modal logic)

According to proponents of impossible worlds, one (at first glance) obvious way to make use of impossible world semantics is to analyse counterpossibles as false. Counterpossibles are a subset of counterfactual conditionals that would normally come out true by their logical form alone in normal modal logic:

> [T]he Lewis-Stalnaker semantics has it that, if there are no A-worlds, $A \mathrel{\Box\!\!\rightarrow} B$ comes out automatically true. The conditional with the same antecedent and opposite consequent, $A \mathrel{\Box\!\!\rightarrow} \neg B$, comes out true, too, for the same reason. In general, all counterpossibles are vacuously true. The standard treatment of counterfactuals implies vacuism about counterpossibles. [4, p. 267]

The following sentences are examples of such counterpossible conditionals:

1. If a leading mathematician would have squared the circle, we should rethink our maths.
2. If Wittgenstein would have squared the circle in his private language, a logician on the first First South American Logic Meeting in Cuzco would have proven that cats are lions.
3. If someone squares the circle, the circle can be squared.

As we have different intuitions towards the truth of conditionals like 1–3, proponents of impossible worlds argue that we should find a language (they suggest impossible world approaches) that allows us to capture what our intuitions are said to suggest (that such counterpossibles should not always come out true). The aforementioned ternary relation allows us to evaluate the antecedent and the consequent of such sentences at respectively different worlds. At first glance, this form of truth assessment solves the problem; some counterpossibles can come out false, some true. In this way, proponents of impossible worlds argue, we can compensate the flaws of normal modal logic.

Such arguments that, first, model a generalizable dispute situation and, then, try to show how possible worlds-only approaches are not enough to do the job, are the motivation of this paper: that is, arguments for impossible worlds based on the alleged flaws of normal modal logic. In Section 3, I will discuss two of these arguments in more detail, connecting their critique to what I will say about Wittgenstein's early philosophy of language.

2 Wittgenstein's architecture of meaning

What follows is a possible reading of Wittgenstein's *Tractatus*. With 'architecture of meaning' I would like to introduce a metaphor for what I believe to be a visualisation of his views on possibility according to his early work. The choice of language, however, is not mine. I have adopted the expression from Stokhof [20], who uses it to frame his ideas on Wittgenstein's notion of meaning. So let us say a few words about the conception of meaning in the early Wittgenstein and, above all, about how it can be *represented* and used for our own needs in dealing with non-normal modal logic. As you read, keep in mind that labels $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{A}^\star, \mathbf{B}^\star, \mathbf{C}^\star, \mathbf{L}, \mathbf{L}^\star$ used below refer to Figure 1. Hence, it is recommended to juxtapose Figure 1 while reading Section 2.1.

2.1 Wittgenstein's notions of meaning and possibility

In an attempt to define what can be said (and thought) at all, Wittgenstein assumes an isomorphic structural relationship – a mapping – between a specific area of reality ($\mathbf{A}$), meaningful sentences ($\mathbf{B}$), and thoughts ($\mathbf{C}$).[5]

According to Wittgenstein, there are essentially two ways in which a non-composite sentence can be meaningful: either it is meaningful insofar as it corresponds to a thought *as well as* to a part of reality, namely, to an occurring state of affairs called 'a fact'. Or it is meaningful insofar it corresponds to a thought *as well as* to another part of reality – namely to a state of affairs that does not occur.[6] According to Wittgenstein, we form meaningful compound expressions by combining meaningful non-composite expressions (to which a meaningful thought also corresponds[7]) with each other, implicitly or explicitly following the rules of Boolean algebra. The underlying idea corresponds (except for the thought part) to what is called the principle of compositionality:

> For every complex expression e in L, the meaning of e in L is determined by the structure of e in L and the meanings of the constituents of e in L. [21, sec. 1.1; cf. 14, pp. 90–1]

[5] Cf. Tractatus, 2.021, 2.0211, 2.0231, 2.024, and [13, especially p. 219].
[6] Cf. Tractatus, 2, 2.06, and [6, sec. 2.2].
[7] Cf. Tractatus, 2, 3.01, 3.02, 3.2, 4, 4.014.

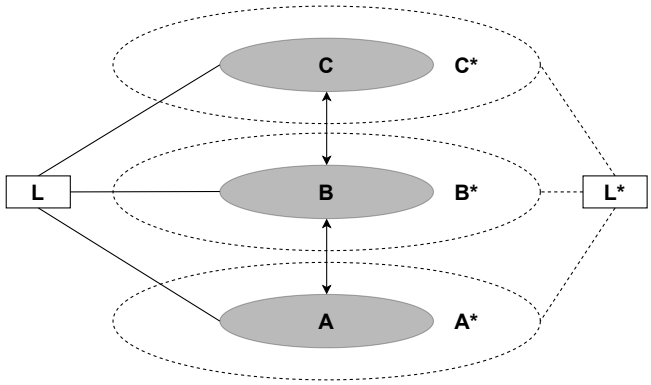

Figure 1: (A reading of) the conception of meaning in the *Tractatus*.

According to the *Tractatus*, compound sentences are meaningful, albeit complex, inasmuch as they correspond to one of the just mentioned areas of reality and with thoughts in terms of their atomic elements and their essential structuring. Suppose, for example, that a conjunction composed according to the rules of algebra consists of a conjunct that is meaningful insofar as it corresponds to an element of the world as the world actually is, i.e., a pertinent state of affairs, and a further conjunct that corresponds to a non-actualised state of affairs. Then the sentence would be a meaningful one according to Wittgenstein. All the ways the world *can* be according to these ideas corresponds to what I labelled **A** in Figure 1 above. How meaningful thoughts *can* be according to such ideas is what I labelled **C**, and how meaningful language *can* be is what I labelled **B**. (One could say, perhaps a little imprudently, that **A** denotes something like a forerunner of our present-day conception of modality [cf. 11; 14, pp. 87–115]. And one could further say that **B** might mean logical possibility, whereas **C** might stand for something like epistemic possibility.)

According to Wittgenstein, *only* those statements that, roughly speaking, 'correspond to the rules of this above schema', express something meaningful that can be *said*. All statements, complex or simple, that do *not* fulfil the tripartite relationship of reality (**A**), language (**B**), and thought (**C**), i.e., that do not 'satisfy the above rules', are either *nonsensical* or *senseless*.[8] According to Wittgenstein, the contradiction, a genuinely senseless logical structures, constitutes the limit (**L**) of what can be meaningfully said. Beyond this limit lies the unsayable nonsensical ($\mathbf{B}^\star$). The contradiction constitutes the outer limit of what is meaningful because it can be evaluated *independently* of whether a state of affairs holds or not, independently of whether a thought describes the world or a possible state of the world; in contrast to other sentences, contradic-

[8]Cf. Tractatus, 4.003, 4.4611, 6.41–6.421, and [23, pp. 12–4, 98–131].

tions are false *solely on the basis of their logical form.*[9] For further illustration, consider Figure 1, which is a possible representation of what I believe to be the general concept of meaning of the *Tractatus* that I just retraced.

Considering what has been said so far, it seems relatively uncontroversial to say that the 'nucleus of meaning' described in the *Tractatus* – i.e., the composition and interrelation of the meaningful thought, the meaningful sentence, and the meaningful way reality can be – can be described (and, through the above graphic, represented) by a transitive mapping relation[10] from $\mathbf{A}$ onto $\mathbf{B}$ onto $\mathbf{C}$. How the world cannot be ($\mathbf{A}^\star, \mathbf{L}$), what does not 'map' onto how the world could be or is ($\mathbf{B}^\star, \mathbf{C}^\star, \mathbf{L}$), what is not a meaningful thought but could still be described as some kind of epistemic state ($\mathbf{C}^\star, \mathbf{L}$)[11], is either meaningless in a strict sense ($\mathbf{L}$) or nonsensical ($\mathbf{A}^\star, \mathbf{B}^\star, \mathbf{C}^\star$). According to Wittgenstein, the nonsensical ($\mathbf{A}^\star, \mathbf{B}^\star, \mathbf{C}^\star$), which can probably not be subordinated to a clear outer demarcation $\mathbf{L}^\star$ modelled on $\mathbf{L}$[12], cannot be said or expressed in meaningful, sensical language, but it can perhaps *show itself* [6, sec. 2.2].

This raises intriguing follow-up questions, which, for all I know, have so far only been partially settled: Does the early Wittgenstein, just as he clearly speaks about nonsensical statements whose content allegedly shows itself, also assume nonsensical thoughts that epistemic subjects can have? Or nonsensical, relevantly impossible states of the world? If yes, how can we assure that $\mathbf{A}^\star, \mathbf{B}^\star, \mathbf{C}^\star$ map onto one another as $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ do? Must we not be able to conceive the nonsensical ($\mathbf{C}^\star$) in order to be able to understand a text like the *Tractatus*, since it declares some of its own sentences to be nonsensical with the famous ladder-sentence?[13] Is it possible to draw a limiting line at the nonsensical ($\mathbf{L}^\star$), just as Wittgenstein tries with the realm of the meaninguful, sensical?[14] (The dotted line in the diagram above is intended to indicate what

[9] Cf. Tractatus, 5.141, 5.142, 5.143, 4.114, 5.61.

[10] By transitive I mean that if $\mathbf{A}$ maps onto $\mathbf{B}$ and $\mathbf{B}$ onto $\mathbf{C}$, then $\mathbf{A}$ also maps onto $\mathbf{C}$.

[11] It is interesting that parallels may emerge here with the ideas of thinkers such as Priest, who assume that we can conceive logical impossibilities (cf. Section 3.2).

[12] An argument about the external delimitability of the nonsensical could roughly proceed as follows: someone might say 'if we are dealing with the claim of the possibility of drawing a boundary to the nonsensical on a linguistic level, then we simply have no unambiguous linguistic form available that could be suitable for the task, since the repertoire of available criteria that can be formulated in formal language (tautology and contradiction) is ultimately already 'used up' with the boundary of the sensible, insofar as the nonsensical is underdetermined in terms of its correspondence and logical form'. This might be followed by an objection such as: 'But where does the postulate of drawing boundaries via definable linguistic criteria come from? Can the boundary of the nonsensical not be one of 'not showing itself'? What cannot be shown and cannot be said is neither meaningful, nor nonsensical, nor meaningless'. One could reply that showing is underdetermined. And so on and so forth.

[13] Cf. Tractatus, 6.54.

[14] The idea behind this question concerning the possibility of delimiting the nonsensical is roughly as follows: For Wittgenstein, the space of the possible, which coincides with the

I believe, that, if it is possible to draw $\mathbf{L}^\star$, it must be different in kind than $\mathbf{L}$.) But let us stop here.[15] For it would certainly be necessary to go into many of the historical and systematic difficulties of the *Tractatus* and its reception in detail to attempt to give an answer to such questions, which cannot be done here. What interests me about the above outlined 'architecture of meaning' with regard to our concerns here (arguments against impossible worlds) is a different, but possibly related matter which I discuss in Section 2.2 below.

2.2 What we might learn from the *Tractatus* regarding nesting models for possibility

Why do I want to ask such a radical, and in some respects probably mistaken, theory[16] about its contribution to a current debate that deals with logical impossibility? The main reason is that false or inconsistent theories can also lead to knowledge if we only refer to certain components of them. If we take David Lewis's modal realism as example, a theory that probably few philosophers consider to be true, we can observe how it persists because it may nevertheless challenge our philosophical reasoning again and again [16, pp. 303–4]. A similar case is perhaps Wittgenstein's logical essentialism in the *Tractatus*.

Even if we do not know exactly what logic Wittgenstein has in mind when he speaks of logical truths and contradictions, what exactly he means with 'facts', or what atomic statements are truly supposed to be, we may still be able to gain something from his 'architecture of meaning' for our concerns here. So-called nestings of different kinds of possibilities are usually represented in

meaningful, is on the one hand delimited by contradiction in so far as the contradiction provides the criterion for what belongs to it, i.e. what is possible (since something contradictory cannot belong to the space of the possible). On the other hand, however, the contradiction also delimits the space of the possible via its demarcation from the nonsensical, which is impossible, since Wittgenstein specifically considers the criterion of contradiction to be violated here. Given this role of the contradiction for the boundary between sense and nonsense, the following question arises: is there something that defines the outer edge of nonsense, as contradiction does with sense? Is there even an outer edge of nonsense or of nonsensical thought and language?

[15] Here are a few more follow-up questions for Wittgenstein aficionados: Does not Wittgenstein's attention to the nonsensical – which has now been sufficiently proven to underlie the project of the *Tractatus* [cf. 23, pp. 12 ff.] – coupled with the dichotomy of saying and showing – retained in his late philosophy – show that although he understands reference and truth as a question of meaningful language, he also assigns some strong semiotic significance to the nonsensical already in his early philosophy? And what if we think of the nature of meaning as a narrowing down of certain, and not other, possibilities of the realm of possibility, as it has been suggested by some? Can we then not also understand his ambition of exploring the 'core' of meaning ($\mathbf{A}, \mathbf{B}, \mathbf{C}$) as such a reduction of the 'space of possibility' in an even broader sense (in the sense of an explicit limitation of the possible in order, as he says in the preface of the *Tractatus*, to find out something about the beloved unsayable)?

[16] That is, my above reading of the conception of meaning in the early Wittgenstein.

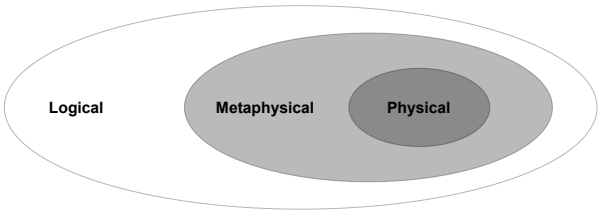

Figure 2: A (possible) nesting model for *possibility*.

the form of Venn diagrams. We humans seem to be particularly good at visualizing how different types of possibility depend on or relate to each other when we represent them two-dimensionally, in Euclidean plane, so to speak. Figure 2 provides an example, which I borrow from *The Stanford Encyclopedia of Philosophy* [15, sec. 1].[17] Given – or with – such conceptual representations, philosophers express things like the following: 'Everything that is physically possible is metaphysically possible, and what is physically or metaphysically possible is logically possible, but that does not necessarily mean that everything that is logically possible is also metaphysically or physically possible'. One could even say that this information can be reliably extracted from the conventional use of this form of representation.

In philosophy, we use representations of nested possibilities like the one above for illustrative purposes, in order to show how we believe things to be, depending on our arguments and assumption. If our opinion of metaphysics, for example, changes, the representation obviously should change as well:

> Inflationists, such as David Chalmers [believe that] there is only one modal notion or primitive, such that metaphysical and logical modality coincide ... Deflationists, such as Sydney Shoemaker ... argue that metaphysical modality coincides with physical modality. Skeptics, such as Graham Priest [a proponent of impossible worlds], question whether there is a notion of metaphysical necessity that is distinct from both, analytic necessity (which corresponds to conceptual necessity) and physical necessity. [15, sec. 1]

(I myself will argue here, as an inflationist, against the idea of the inevitability of impossible world semantics.)

With the wording *architecture of meaning* referring to Wittgenstein's early philosophy, I intended to introduce a *metaphor* for his thoughts on possibilities in the *Tractatus* (and their nesting). These thoughts, as I believe and as I have

[17] The two illustrations are alike in all decisive respects.

done above, should be *represented* as three dimensional structure, in a three-dimensional Euclidean space, so to speak. For, as we have seen, they are based, on the one hand, on different (namely, three) modes of possibility, which then project onto one another. On the other hand, each plane represents its own kind of possibility, including its boundaries (one epistemic, one linguistic, one of physical possibility in terms of Wittgenstein's essentialism) as well as what, in each case, may lie behind it.

What I intend to employ in the following is *not* Wittgenstein's assumption of an isomorphic relation of the three mentioned modal layers (and his related logical essentialism), but mainly the idea of a mapping of different spaces of possibility onto each other, representing ideas on modality as layered planes. For I think that this way of representing our philosophical thoughts should help us to question the reliability of arguments advocating the need for impossible world semantics. One could object that this involves too much historical work just to ultimately criticise arguments that could also be criticised without such effort.[18] However, it will become clear, I hope, that the idea of tracing back layered nesting models of possibility to the architecture of meaning in the *Tractatus* can enrich both the systematic and the historical discourse on modality.

3 *Layered* nesting models for possibility

Based on some aspects of what was introduced as the 'architecture of meaning' in Wittgenstein's *Tractatus*, two arguments for impossible worlds will be criticized in the following on the basis of layered nesting models for possibility. In each case, these are representations of possibility relations that do not take the form of a single plane, but of a projection of planes.

3.1 Arguing for impossible worlds via metaphysical dispute

The first argument I will reconstruct is supposed to show that possible world semantics are not sufficient to describe what happens when two rational agents engage in a deep metaphysical disagreement. Francesco Berto and Mark Jago, proponents of impossible worlds, argue that the following dispute situation shows a way in which possible world-only approaches fail [4, pp. 23–4]:

Someone, let us call him Harry, discusses with someone called Lilith about the nature of properties. Both agree on the fact that propositions *are* sets of

[18] I would like to take this opportunity to thank Elia Zardini, who asked me at the SALOME 1 conference whether the recourse to Wittgenstein was at all necessary for the project I was pursuing. In any case, I would probably not have produced the essay, and the talk in the first place, without the recourse to Wittgenstein.

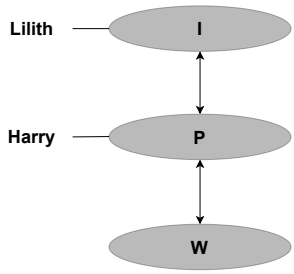

Figure 3: (Alleged) equivalence.

possible worlds. Harry believes that there are transcendent platonic universals (P) and Lilith that universals are immanent (I). (Note that the argument is not about the content of the individual beliefs, i.e., about universals. So if the reader does not think much of the discussion around universals, he or she can 'substitute' the two stances with any set of contradictory metaphysical claims.)

i Both believe that propositions *are* sets of possible worlds.[19]

ii Harry believes that P necessarily true and I is necessarily false. Thus, he believes that the set of all possible worlds W is equivalent to P. He believes too that $W = (P \text{ or } I)$, as, by disjunction-introduction, the claim 'P or I' remains true for all possible worlds and therefore (as propositions are sets of possible worlds) the disjunction 'P or I' is identical to P.

iii The same is valid for Lilith: she believes that I is necessarily true and P necessarily false, therefore, she believes that $W = I$ and she believes too that $W = (I \text{ or } P)$, as, by disjunction-introduction, the claim remains true for all possible worlds and, thus, the disjunction is identical to I.

c Therefore, from **i-iii**, Harry and Lilith should agree upon the fact that their claims are equivalent and, to avoid this, they should refute normal modal logic-only approaches – unless they want to give up dispute in metaphysical matters instead.

Berto and Jago suggest that Harry and Lilith should reject the first premise, that propositions are sets of possible worlds. For then they could continue to rationally discuss, no longer being forced to subscribe that their respective beliefs are equivalent as represented in Figure 3. We should introduce impossible worlds, Berto and Jago argue, so that we can say something like 'Lilith's claim

[19]Therefore, if 'two' propositions correspond to the same set of possible worlds, the two sets actually 'are' one proposition.

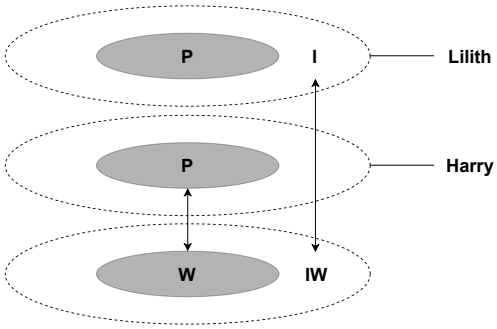

Figure 4: Dispute relying on impossible worlds.

does not map to the set of all possible worlds, but it still maps to impossible worlds (abbreviated with **IW** in Figure 4); Harry might be right about how reality is, but it does not follow that Lilith necessarily holds an equivalent claim to be true'. If Berto and Jago are right, then metaphysical dispute – as perhaps paradigmatically exemplified by this case – might really require that some or all propositions are not just sets of *possible* worlds [4, p. 24].

However, as I have shown elsewhere before [9, sec. 5], the argument has an issue, because, if one assumes that Harry's view is the right metaphysical claim, I think that Harry would still accept premise **ii**, but he would then reject what Lilith is said to accept in premise **iii**. The reason is that, holding fixed his own epistemic perspective (scenario **A** in Figure 5), he would have to accuse Lilith of carrying out her disjunction introduction based on a false belief (I). The same holds conversely if one takes Lilith's assumption to be the correct one (scenario **B**, again, in Figure 5).

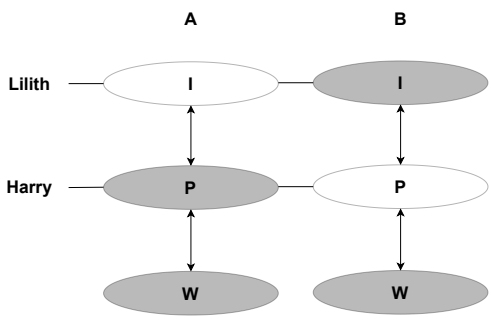

Figure 5: Dispute without relying on impossible worlds.

Therefore, this argument is not strong in showing that impossible worlds are necessary, since the conclusion does not follow from **i-iii**. In scenario **A**, I does

not map to W and P does not in scenario **B**. (Note that the point here is not that the disjunctions 'P or I' and 'I or P' are not equivalent – they are – but rather that one agent does not believe the conviction of the other to be true. So even without impossible worlds, there is a way to keep the two positions apart. But perhaps in such under-determined situations of metaphysical discourse we should actually withhold our judgment anyway until further reasons speak for or against one of the positions.)

3.2 Arguing for impossible worlds via conceivability

The second argument is supposed to show that possible world semantics are not sufficient to model what we conceive when we allegedly conceive impossibilities. Graham Priest, a proponent of impossible worlds, argues that, on the basis of approaches that do *not* make use of impossible worlds:

> Some things that are epistemically possible would seem to be logically impossible. Thus, before Wiles' proof of the truth of Fermat's last theorem, its negation was epistemically possible, though logically impossible. [19, p. 2652]

Therefore, according to Priest, the dispute around Fermat's last theorem (FLT) is a case of dispute in which a logical impossibility ($\neg FLT$) must be regarded as epistemically possible, conceivable[20], that means, must be understood as being the content of some kind of epistemic state. The state of conceiving, according to Priest, further involves conceivable objects of some general kind. Priest believes that we can conceive of everything that is expressed in terms that we can understand: "I can conceive of and imagine anything that can be described in terms that I understand" [19, p. 2659].

According to Priest, 'conceiving FLT' and 'conceiving FLT not being a logical truth' both are viable epistemic states. How else, he argues, could we describe, for example, what drives epistemic agents when they sometimes conceive a mathematical theorem, even though the respective theorem might later turn out false? How else could we describe what makes agents irrational, which do believe that FLT is not a logical truth? The now following thought experiment is my reconstruction of one of Priest's arguments on why we should abandon approaches relying *only* on possible worlds [19]. After the presentation of the thought experiment, I will refer to the labels **I, I⋆, M, M⋆, H, H⋆, L, L⋆** of Figure 6 to further illustrate it.

i Harry and Lilith (both mathematical laypersons) as well as Matteo (an excellent mathematician) all believe that, if an x can be conceived, then

[20] Priest uses 'conceivable' as roughly synonymous with 'imaginable' [19, p. 2658].

x must be logically possible. They believe that if one conceives x, then x or what is expressed by x necessarily corresponds to some quantification over the set of all logically possible worlds.

ii Harry conceives FLT to be a truth before Matteo proves that FLT is indeed a logical truth.

iii Lilith conceives $\neg FLT$ to be a truth before Matteo proves it to be false.

iv FLT is shown to be a logical truth by Matteo. He conceives FLT. Harry and Lilith are informed about the truth of FLT as soon as the proof is given and they are in a position to understand the language of Matteo.[21]

v If FLT is shown to be a truth at some moment, then it was a truth also before. If FLT is conceivable at some moment, then it was also conceivable at any moment before.[22]

c On the grounds of **i**-**v**, Harry, Lilith, and Matteo should agree to introduce impossible worlds and therefore deny **i**. For, if FLT is shown to be a logical truth by Matteo (**iv**), then FLT was a truth before (by **iv**-**v**). But if only logical possibilities can be conceived (**i**), then conceiving FLT not to be a truth after the proof and, by **v**, conceiving FLT not to be a truth before the proof would not hold for any agent. But this would contradict Lilith's state of conceiving before the proof. The only permissible state of conceiving for any reasoner concerning FLT based on this reasoning (accepting **i**), according to Priest, would therefore be to conceive it as true [19]. But how can we then explain what Lilith conceives before she is informed about Matteo's proof?

On these grounds, Priest seems to argue that we should not give up **iii**, but instead modify **i**; we should weaken our notion of conceivability and accept impossible worlds as a tool to model our conceiving of impossibilities. Because, according to Priest, we are interested in describing what Harry and Lilith conceive before the proof, even if one of them might have conceived a logical impossibility, in order to understand how humans reason in general. Therefore, he argues, we should accept a more general notion of what is conceivable by adding impossible worlds as a tool to model such states of conceiving;

[21]That they must be in a position to understand the terms, again, follows from Priest's postulate cited above: 'I can conceive of and imagine anything that can be described in terms that I understand' [19, p. 2659].

[22]Priest does not explicitly express this assumption. However, it seems that a condition of possibility of his thought experiment on impossibility is that we require the regularities governing knowledge and conceivability, as well as the continuity of the relevant abilities of the epistemic agents involved, to be non-variable over time. Thus, I list these conditions here.

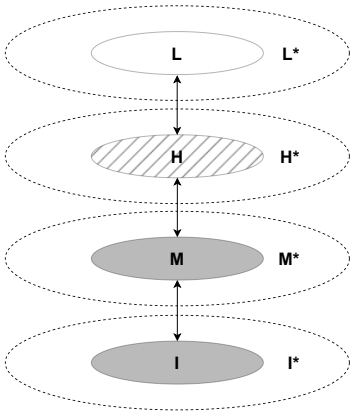

Figure 6: Mapping of epistemic possibilities.

'If something is conceivable for an epistemic agent, then it is either a logical possibility or a logical impossibility'. This corresponds to Priest's conviction, since, as we have seen above, he believes that we can also imagine inconsistent, i.e. logically impossible, matters. Priest thus believes that, provided we can imagine something epistemically, this something either is an impossibility or, as commonly and less controversially assumed, a logically possible matter. Of course, there may be things that we cannot conceive of. But, according to Priest, from the fact that we can conceive something it does not follow that this something is necessarily logically possible, and it also does not follow from the fact that something is logically impossible that we cannot conceive it. Priest thus ultimately wants to concede to the epistemic agents involved that they can not only conceive what is possible in principle ($\mathbf{M}, \mathbf{H}, \mathbf{L}$) but also what is logically impossible ($\mathbf{M}^\star, \mathbf{H}^\star, \mathbf{L}^\star$). (The labels correspond to the respective first letters of the names of the fictitious agents of the thought experiment.) Let us criticise this argument by visualising it in Figure 6 as a layered mapping of the respective interrelated fields of possibility.

In what follows, as I have tried to make plausible elsewhere [9] in greater detail reconstructing the work of Duc [8], I want to defend the view that there is another way of describing the dispute situation without relying on impossible worlds – without therefore assuming what is represented in Figure 6 by $\mathbf{M}^\star, \mathbf{H}^\star$ and $\mathbf{L}^\star$. For, following Chalmers, one can assume that what was called Lilith's 'conceiving' before the proof is in fact a case of *prima-facie conceiving* [7, sec. 1], but that Matteo's state of conceiving after the proof is a case corresponding to *ideal conceiving*. According to Chalmers, one way we can define prima facie and ideal conceivability is the following:

S will be prima facie conceivable for a subject when that subject

> cannot (after consideration) detect any contradiction in the hypothesis expressed by S. S is ideally conceivable when S is conceivable on ideal rational reflection. [7, p. 147]

Building on this, it can be argued that there is a difference in kind between Matteo's understanding of FLT and the understanding of the others. For to the extent that it can be shown that a mathematical proof of the truth of FLT is *indistinguishable in relevant respects from* or *equivalent to* or even *an event of* 'ideal rational reflection', FLT is ideally conceivable by Matteo according to the above definition and ideally conceived given his proof.[23] Under this interpretation of the relation between conceivability and rationality, FLT would be conceived by Matteo (after the proof) just as an ideal agent would (**I**). Matteo's ideal conceiving **M** would then map onto **I** under this interpretation of his epistemic disposition.

Things are different for Lilith. She is not in a position to know if FLT is a logical truth, she does not detect all the implications of FLT that Matteo 'sees', given his proof. For if she would have 'seen' all the logical implications of FLT relevant for the proof, she would not prima-facie conceive *what she believes to be* the negation of FLT according to the above definition given by Chalmers. (Because ideally conceiving the truth of FLT implies knowing that at relevant depth FLT implies no contradiction. But if she knew that FLT implies no contradiction, then she would know that $\neg FLT$ does, and, therefore, she would not be able to prima-facie-conceive $\neg FLT$ according to Chalmers' definition.) But it would be a non-sequitur to infer that Lilith can conceive $\neg FLT$. Her 'conceiving' concerns implications of what *she holds to be* FLT or what she holds to be its negation. We can, again, under the above assumptions about the relation between rationality and conceivability, say that she does not conceive in the same way an ideal reasoner or a potentially ideal but finite reasoner like Matteo, given his proof, conceive (**I, M**).

[23] Duc has shown that we can model finite epistemic agents whose epistemic constitution may nevertheless be interpreted, under certain conditions, as being equivalent to 'ideal rational reflection'. For the following applies to finite agents modelled in his logic: 'If A is a truth, then the agent c can get to know that A, on condition that the agent puts enough effort $\langle \mathsf{F} \rangle$ in the process of getting to know A': $\vdash A \Rightarrow \vdash \langle \mathsf{F} \rangle \Box A$' [1, p. 11]. The new operator $\langle \mathsf{F} \rangle$ introduced in his dynamic epistemic modal logic has the special feature that it can be interpreted as synonymous with the classical knowledge operator of normal modal logic once a sufficient computational depth has been reached with regard to a specific task. On this basis, it may also be possible to argue that a mathematical proof of FLT being a logical truth is synonymous with 'investing enough effort' into the respective task and thus corresponds or amounts to ideal rational reflection in terms of the K operator. Actually, at this point one would have to discuss the phenomenon of logical omniscience in relation to dynamic epistemic modal logic, which is a possible solution in the tradition of Hintikka to circumvent logical omniscience and still maintain the principle of rationality in the sense of the knowledge operator of normal modal logic. There is no room for a more detailed discussion here [cf. 1, 8, 9].

That does not mean we cannot understand what Lilith and Harry conceive (**L**, **H**) if we do not allow the conceiving of impossibilities (**I***, **M***, **H***, **L***) to be part of our explanation. It just means that we cannot understand what Lilith, Harry, or Matteo would conceive if they conceived a logical impossibility. But should this bother us? We surely still can talk about what agents like Lilith might conceive, erroneously *believing* to conceive something they actually do not conceive. We can therefore describe the relevant steps of the above situation by sharpening our concept of conceiving. Lilith conceives *what she holds to be* non-contradictory implications of *what she holds to be* FLT; and Harry conceives *what he holds to be* non-contradictory implications of *what he holds to be* FLT. By luck or intuition Harry might actually conceive FLT or implications of FLT, but neither Harry nor Lilith conceive in the way Matteo does. (That is why **H** in Figure 6, contrary to **L**, is shown neither blank, nor filled, but dashed: in order to make it clear that, without further information, it might still be uncertain whether **H** maps to **I** and **M**.)

4 Conclusion (and a thoroughly possible outlook)

The elaboration of (what I believe to be) a layered representation of the general conception of meaning in the *Tractatus* (cf. Section 2.1) leads to question whether two-dimensional representations of our philosophical ideas on modality might not sometimes be better captured by three-dimensional representations (cf. Section 2.2 ff.). For in the case of the discussion concerning the necessity of non-normal modal logic for philosophical theorizing and modelling, it has been shown through layered visualisations that the arguments that are supposed to speak in favour of impossible worlds show relevant non-mappings of modal realms. However, according to these very arguments, these realms would actually be required to map to each other. This was demonstrated for the dispute argument (cf. Section 3.1) and for the argument that works via conceivability (cf. Section 3.2). It can thus likely be inferred that different arguments for impossible worlds suffer from similar structural problems, which manifest in the form of relevant non-mappings as soon as layered representations/visualisations are employed.

Of course, it would have to be shown that the here-forwarded mode of representation is good at showing how different arguments for impossible worlds fail, and not merely because Homo sapiens is good at processing such visualizations. In other words, one would have to show what philosophical reasons against arguments for impossible worlds these visualizations can allow us to systematically stumble upon. I believe that if one were to further explore the structural similarity (hinted here via the use of visualizations) between specific aspects of Wittgenstein's early philosophy and our representation of modal

spaces, these philosophical reasons could gradually become apparent. What can already be said at this point in the form of only partially substantiated assumptions about these philosophical reasons, however, is the following.

Modal visualisations in three-dimensional Euclidean space are likely valuable because they seem to enable us to grasp how different arguments for impossible worlds share a certain implicit or explicit conception of logical space, namely a logical space in some sense 'expanded' – visually speaking, in a Euclidean plane – by logical impossibility. This, in turn, coupled with the problematic non-mapping of modal realms at crucial points of arguments for impossible worlds, raises the assumption that there might be a good reason why Wittgenstein concedes to the space of the meaningful a (layered) two-dimensionality, whereas he seems to think of logical (and epistemic and physical) impossibility as a one-dimensional limitation of each plane in this three-dimensional layered nesting (cf. Figure 1).

One way to make this idea of Wittgenstein valuable to today's modal logic discourse might be, for example, to ask whether the behaviour of the knowledge operator in normal modal logic is not particularly desirable precisely *because* it sets the limits of our thinking through contradictions, in a similar way to what Wittgenstein seems to be pursuing: for knowledge of P is only given if there is no accessible possible world (i.e., epistemic state) where non-P holds (cf. Section 1). Knowledge in terms of the knowledge operator of normal modal logic therefore goes, one might say with Wittgenstein, 'to the *point* of contradiction and no further'.[24]

To venture one last speculative remark, I would like to conclude by transferring Wittgenstein's dichotomy of saying and showing to the discussed task we are currently facing in the field of modal logic: it seems as if in today's philosophical practice we have to decide, in Wittgenstein's words, whether we want to admit that we can *show* or even *say* something through or with impossible world semantics. So are we willing to accept impossible world semantics in our 'philosophical toolbox' based on structurally unconvincing arguments, or should we perhaps acknowledge that we can be 'bewitched' not only by normal language, but also by formal language – that is not to be confused with ideal language?

Acknowledgment

Out of attachment I thank the METRONOM. I would also like to thank the anonymous referee of the SAJL and all participants of SALOME1 for their helpful contributions.

[24]Due to the connection between knowledge and ideal conceivability, these considerations would also have an impact on the considerations made in Section 3.2.

References

[1] S. Artemov and R. Kuznets, Logical omniscience as infeasibility, *Ann. Pure Appl. Log.* 165:6-25, 2014.

[2] F. Berto, Impossible Worlds and the Logic of Imagination, *Erkenntnis* 82:1277-1297, 2017.

[3] F. Berto, On Conceiving the Inconsistent, *Proc. Aristot. Soc.* 114:103–121, 2014 .

[4] F. Berto & M. Jago, *Impossible Worlds*, Oxford University Press, 2019.

[5] F. Berto & R. French & G. Priest & D. Ripley, Williamson on Counter-possibles, *J. Philos. Logic* 47: 693-713, 2018.

[6] A. Biletzki and A. Matar, Ludwig Wittgenstein, *The Stanford Encyclopedia of Philosophy*, fall 2023 Edition.

[7] D. Chalmers, Does Conceivability Entail Possibility?, In *Conceivability and Possibility*, Oxford University Press, pp. 145-200, 2019.

[8] HO NGOC DUC, Reasoning about Rational, but not Logically Omniscient, Agents, *J. Log. Comput.* 7:633-648, 1997.

[9] T. Eyinck, Should We Embrace Impossible Worlds Due to the Flaws of Normal Modal Logic?, *Log. Univ.*, 2024.

[10] V. Goranko and O. Martino, 5 Model theory of modal logic, In *Studies in Logic and Practical Reasoning* 3, pp. 249-329, 2007.

[11] N. Griffin, Modality and the Tractatus, *Dialogue*, 34(4):807-814, 1995.

[12] R. Hilpinen, Deontic Logic, In *The Blackwell Guide to Philosophical Logic*, pp. 159-182, Blackwell, 2017.

[13] J. Hintikka, Quantification and the picture theory of language, *The Monist*, 53(2):204-230, 1969.

[14] M.B. Hintikka and J. Hintikka, *Investigating Wittgenstein*, Basil Blackwell, 1986.

[15] A. Mallozzi, A. Vaidya, and M. Wallner, The Epistemology of Modality, *The Stanford Encyclopedia of Philosophy*, fall 2023 Edition.

[16] K. McDaniel, Modal Realisms, *Philosophical Perspectives*, 20:303-331, 2006.

[17] J.-J. Meyer, Epistemic Logic, *The Blackwell Guide to Philosophical Logic*, pp. 183-202, Blackwell, 2017.

[18] D. Perler, M. van Ackeren, and L. Klein, The Alienation Effect in the Historiography of Philosophy, In *Philosophy and the Historical Perspective,* pp. 140–154, Proceedings of the British Academy, 2018 .

[19] G. Priest, Thinking the impossible, *Philos. Stud.* 173:2649–2662, 2016.

[20] M. Stokhof, The Architecture of Meaning: Wittgenstein's Tractatus and Formal Semantics, In *Wittgenstein's Enduring Arguments*, pp. 211-244, Routledge, 2008.

[21] Z.G. Szabó, Compositionality, *The Stanford Encyclopedia of Philosophy*, fall 2022 Edition.

[22] K. Tanaka, Logically Impossible Worlds, *Australas. J. Philos.* 15:489-497, 2018.

[23] H. Tetens, *Wittgensteins "Tractatus"*. Philipp Reclam Jun., 2009.

[24] H. Wansing, Bemerkungen Zur Semantik Nicht-Normaler Möglicher Welten, *Math. Log. Q.* 35:551-557, 1989.

[25] L. Wittgenstein. *Tractatus logico-philosophicus.* (Cited using the German study edition: W., L.: Werke, Suhrkamp, 1984. Vol. 1/8.)

[26] L. Wittgenstein, *Zettel*, Basil Blackwell, 1967.

Til Eyinck
Petrarch Institute, University of Cologne
Richard-Strauss-Str. 2, 50931, Cologne, Germany
E-mail: `teyinck1@uni-koeln.de`