Log. Univers.

https://doi.org/10.1007/s11787-024-00344-4

Logica Universalis

Should We Embrace Impossible Worlds Due to the Flaws of Normal Modal Logic?

Til Eyinck

Abstract. Some philosophers advance the claim that the phenomena of logical omniscience and of the indiscernibility of metaphysical statements, which arise in (certain) interpretations of normal modal logic, provide strong reasons in favour of *impossible* world approaches. These two specific lines of argument will be presented and discussed in this paper. Contrary to the recent much-held view that the characteristics of these two phenomena provide us with strong reasons to adopt impossible world approaches, the view defended here is that no such 'knock-down arguments' do emanate on those grounds. This is not to rule out that there cannot be any other good reasons for assuming impossible world semantics. However, the discussion of a further argument for impossible worlds will suggest that different attempts to argue for them likely present intertwined problems.

Mathematics Subject Classification. Primary 03B45; Secondary 03B42, 03B60, 03B65, 68T30.

Keywords. Normal modal logic, Impossible world semantics, Logical omniscience, Disagreement in metaphysics.

1. Introduction

The main goal of this paper is to discuss some pros and cons of impossible world (IW)[1] semantics, not to answer what IW essentially are or could be. Therefore, IW theories are discussed here, in the broadest sense, as an existing set of logical theories mainly arisen from the work on modal logics and presented as useful tools in recent publications [2,4,5,13]. To simplify the research landscape of modal-logics, I assume that there are two general positions regarding IW semantics:

[1] I will use IW in the following as an abbreviation for impossible world as well as for the plural form, impossible world*s*.

Published online: 31 January 2024

IW-sceptics, who do not want to rely on any kind of reasoning grounded on IW; the sceptics do not want to base their inferences on assumptions developed using IW. There are plenty of worries about IW-concepts and, following the sceptics, most of them relate to the worry that normal modal logics (NML) lose their useful qualities when IW enter the game.

The second position is, by contrast, open towards the idea of IW. Mainly, the supporters of IW are driven by the goal of avoiding the problems connected to the phenomenon of hyperintensionality, which is an umbrella term Berto and Jago suggest for issues that NML is allegedly struggling to capture [4, pp. 21–22].

First I will define some general formal concepts. Then I critically discuss why some authors hold IW to be needed; it will through this be easier to see how one can commit to either of the two positions.

2. Aren't Possible Worlds Already Enough?

Just to give a basic idea of what NML can look like, I briefly define possibility ($\Diamond$) in a Kripke-style model of NML, a model *only* containing possible worlds (PW). To give truth conditions for formulas of type $\Diamond\varphi$ we use 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). So that, if the whole structure (M = $\langle$ W,R,V$\rangle$) is given, one can say: $\Diamond\varphi$ (equivalent to $\neg\Box\neg\varphi$), if there is at least one world w in W, where the assigned truth value for φ is $\top$ [4, pp. 95–98)].

Now, having at hand such concepts of NML, one might ask oneself: why should I need IW? Isn't the setup described above in its main features already powerful enough? Some authors say that it might be too powerful in some sense and too weak in another: Berto and Jago state in their book *Impossible Worlds*, "possible worlds are a success story of philosophical theorizing"[4, p. 21] because they have been applied in almost every area of philosophy and even outside of it. Still they (the NML) face some serious problems [Ibid.].

After giving a brief example on how IW formally look like in the discussion, I will try to trace and debate two of these main problems of NML (the phenomenon of the indiscernibility of metaphysical statements and the phenomenon of logical omniscience), mainly by rebuilding thoughts from Berto and Jago.

3. Commonality Between Possible and Impossible Worlds

Although it is not the aim of this paper to explain what IW are, it would be wrong to enter the debate without even scratching the surface of what they are or could be.[2] Therefore, I limit myself to give a short and generic sketch. A

[2]If one is interested in a deeper understanding of the different kinds of impossibilities that can be distinguished when talking about impossible worlds in modal logics, it might be interesting examining the following paper: [13].

good word for what IW are is framework, as IW-approaches generally consist in adapting the NML that we have seen above and adding some extra formal structure to it. A (simplified) model M' for possible and impossible worlds therefore can look like this:

$$\mathrm{M}' = \langle \mathrm{W,N,R,I} \rangle \qquad [13, \mathrm{p.}\,490].$$

Here N is a proper subset of W, containing the normal (possible) worlds, whereas the non-normal (impossible) worlds are not in N. R is a ternary world-to-world relation $\mathrm{R} \subset \mathrm{W} \times \mathrm{W} \times \mathrm{W}$ and I is a map from propositions to sets of worlds. R satisfies the normality condition:

$$\langle \mathrm{w,w', w''} \rangle \in \mathrm{R} \text{ iff } \mathrm{w'} = \mathrm{w''} \text{ for every } \mathrm{w} \in \mathrm{N} \text{ and every } \mathrm{w', w''} \in \mathrm{W}.[13, \mathrm{p.}\,491]$$

This means exactly that R at normal worlds is binary, whilst it is ternary at non-normal worlds. In short, this[3] kind of setup allows logicians to interpret certain counterfactual conditionals (namely counterpossible conditionals) by evaluating the antecedent and the consequent at respectively different worlds [13, pp. 492–493]. In the following section I would like to make a brief remark on such interpretations of counterpossibles giving some examples. I will however not discuss the subject matter in detail, for which this is not the place.

4. A Remark on Counterpossibles

One can consider a metaphysically impossible sentence P, which *cannot possibly* be true at any possible world of any possible NML model. A candidate for such a P might be "Wittgenstein squared the circle in his private language". If such a sentence constitutes a (necessarily) false antecedent, this leads—due to the Lewis-Stalnaker semantics of NML[4]—to the necessary truth of any consequence. This makes sentences like "If Wittgenstein would have squared the circle in his private language, Popper would have written the Holy Bible" come out true for NML-Frameworks. Some philosophers consider this way of interpreting counterpossibles unintuitive as well as problematic enough in its subject matter to provide cause for questioning PW-only approaches. These thinkers therefore adopt IW-setups like the one above to—broadly speaking—make sense of impossibilities, so that the corresponding counterpossible conditional *can* come out false [4, p. 267 ff]. An impossible world is consequently something like a 'safe-space' of a logic L, defined within the logic L, where logical truths fail in a controllable manner, or, as Tanaka puts it,

[3] One should keep in mind that this is still a simplified formalisation for the sake of clarity.

[4] "[T]he Lewis-Stalnaker semantics has it that, if there are no A-worlds, A $\Box\!\!\rightarrow$ B comes out automatically true. The conditional with the same antecedent and opposite consequent, A $\Box\!\!\rightarrow \neg$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].

> [...] [it] is a part of the structure that defines the logical truths of the system but is a world where those very logical truths fail to hold. [13, p. 493]

Quite similarly, most other IW-approaches try to proceed and thereby describe how human imagination might work [2] or enrich the project of logic, thinking about how logic could be [14].

Counterpossible conditionals, which are a subset of counterfactual conditionals, are usually problematised within NML analysis because our intuitions about how they should be analyzed do not match the semantic interpretation provided by the NML. For example, Berto et al. argue in their joint paper "Williamson on counterpossibles" that the fact that some specific counterfactual conditionals, when analysed in NML, make the conditional true is so counterintuitive that we should look for other semantic interpretations of counterpossibles (IW semantics) [5]. The following cases of counterpossibles can be used to illustrate their point:

- P1: If a leading mathematician would have squared the circle, we should rethink our maths.
- P2: If Wittgenstein would have squared the circle in his private language, a mathematician will prove that garlic is green.
- P3: If I would square the circle, I would probably be mad.

Berto et al. draw attention to the fact that philosophers have different intuitions regarding the truth of P1, P2 and P3. Therefore, following the authors there seems to be a need for IW approaches here, based upon our intuitions [5, p. 695].

The word "philosophers" used in the previous sentence means, one could paraphrase Berto et al.'s opinion, a set of rational agents whose elements to a large extent share the intuition that we should extend NML by IW in order to analyze P1, P2 and P3 correctly. Given, however, that a determination of the elements of that set is not the ambition of this paper and given the fact that "one can have an intuition that P even if P is false" [7, p. 70], two further lines of argument from the 'pro IW faction' are examined below for their robustness, which are less based on intuitions.

5. Metaphysical Debate—Digging Deep to Disagree

One main line of thought against PW-semantics, often advanced by the IW fraction, is this one: Having at hand a NML-model as the one above, one could assume a scenario in which someone, let us call him Harry, discusses with someone called Lilith about the nature of properties and that—as some philosophers would claim too—both agree on the fact that propositions *are* sets of possible worlds. Harry thinks that there are transcendent platonic universals (P) and Lilith thinks that there are immanent universals (I). Berto and Jago argue this to be a generalizable situation of metaphysical dispute that results in what follows [4, pp. 23–24]:

i Both believe that propositions *are* sets of possible worlds.[5]

ii Harry believes that P is necessarily true and I is necessarily false, therefore he believes that W = P. He believes too that W = (P or I), as, by disjunction-introduction, the claim (P or I) remains true for all PW and therefore (as propositions are set of PW) 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 is necessarily false, therefore she believes that W = I and she believes too that W = (I or P), as, by disjunction-introduction, the claim remains true for all PW and therefore the disjunction is identical to I.

c Therefore, from **i-iii**, both should agree upon the fact that their claims are equivalent and to avoid this they should refute NML, if they do not want to give up the project of metaphysics instead.

Berto and Jago suggest that Harry and Lilith should reject the first premise, that propositions are sets of PW, because then both could continue to discuss the nature of proprieties, as they would no longer be forced to subscribe that their respective beliefs are in fact equivalent. If Berto and Jago are right, consequently 'deep' metaphysical debate seems to request the idea that some or all propositions are *not* just sets of *possible* worlds, therefore perhaps sets of impossible worlds [4, p. 24].

However, I believe that the argument, as interesting as it is, has an issue, because, if one assumes that Harry's view is in fact 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, from his epistemic perspective, 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. Therefore, this argument did not turn out to be as powerful as it first appeared in showing the urge of non-PW conceptions, since the conclusion does not follow from **i-iii** (at least, in the way the argument has been [re-]phrased here). An argument that is similar in approach and also similarly problematic will concern us in Sect. 7 of this paper.

6. Logical Omniscience

The pro IW-fraction often advances the claim that the problem of logical omniscience gives us reasons to assume the existence of IW, as they regard it as a problem inherent to epistemic NML-interpretations. However, the unifying feature of accounts that try to avoid logical omniscience by modelling NML is

[5]If 'two' propositions correspond to the same set of possible worlds, the two sets actually are one proposition. One additional thing that I want to note about the first premise of their argument is that it does not consider the assumption of possible worlds being sets of propositions and therefore a maximally extended proposition, which may be a valid claim to make.

their qualitative change of NML (therefore, IW accounts constitute a qualitative change to NML). Qualitative changes to NML imply severe complications, as for instance a relevant weakening of the so obtained logic L'. The weaker L' does not count as normal anymore; it is not a NML [1, p. 9] [4, p. 99 ff]. Is there no other way to avoid the problem of logical omniscience?

In Hintikka's tradition, I defend the view that the 'problem' of logical omniscience—in many of its manifestations—should rather be called a phenomenon than a problem, as

> [...] the semantical solution of the problem of logical omniscience [...] coincides with the syntactical (proof-theoretical) solution [...] [9, p. 483].

To this end, the problem of logical omniscience must first be outlined. One way to phrase logical omniscience, which in itself follows directly from the axioms of NML, is: whenever an agent c knows all of the formulas in a set Γ and A follows logically from Γ, then c also knows A [12, ch. 4]. The following implications do result directly from interpreting knowledge (K) as quantifying over PW within NML [4, p. 24]:

i If the agent c KA and A entails B, then c KB.
ii If A is valid, then c KA.
iii It is not the case that c KA and c K¬A.

One might ask why these implications should be regarded as problematic. They ought to be understood as problematic, because they mean that any so modelled agent that knows what natural numbers are, a fortiori needs to know all mathematical truths related to natural numbers and moreover necessarily has only consistent beliefs [12, ch.4 ff]. Isn't that enough to raise doubts about the usefulness of accounts that do not make use of IW? Maybe not. To defend Hintikka's position, I assume the existence of a computable, normal epistemic modal logic CML, designed to describe cognitive agents. One should think of it as one logic out of the group of all possible NML, which all share the basic structure of NML described above. Some authors in Hintikka's tradition propose a method to define finite agents and contemporarily leave the 'NML-core' of CML untouched:

They acknowledge the aforementioned "whenever an agent c knows all of the formulas in a set Γ and A follows logically from Γ, then c also knows A". They do though modify it by the intuitive condition that c can get to know A, *if* she has enough time and computational power at her disposal, but that this does not necessarily entail that c actually knows A [8]. This is what Hintikka means when he says, that

> Just because people [which in this case stands for finite agents] [...] may fail to follow the logical consequences of what they know ad infinitum, they may have to keep a logical eye on options which only look possible but contain hidden contradictions [9, p. 476].

Duc suggests to interpret K with something like 'is in a position to know' for the cases in which c is not meant to be an ideal agent [8, p. 635 ff]. Probably, I do not have the *resources and the time* to even try to present the problem/phenomenon of logical omniscience in its adequate *depth* here; I will therefore focus on the reconstruction of some results from Artemov and Kuznets, which, by enriching the project of Duc, argue, in Hintikka's fashion, in favour of the NML-approach.

Duc introduces a knowledge operator $\langle F \rangle$. The operator should be interpreted as 'If A is a truth, then the agent c can get to know that A, if she puts enough effort in the process of getting to know A'. Formally, this looks as follows [1, p. 11]:

$$\vdash A \Rightarrow \vdash \langle F \rangle \Box A.$$

The implementation of such a second knowledge operator has two (philosophically) pleasant consequences [1, Ibid.] [4, p. 118 ff]:

1 The usual operator $\Box$ preserves logical omniscience.
2 $\langle F \rangle$ allows modelling a non-omniscient agent within CML.

So, firstly, through the usual modal operator, the philosophically relevant notion of ideal rationality, namely rationality in the sense of the K-operator of NML, is preserved. Weaker 'derivatives' of NML, e.g. IW semantics, are not able to achieve this. Secondly, it is possible[6] to model fallible agents, which is of philosophical interest not only—but already—due to the fact that only fallible agents are known to write philosophy papers. The operator $\langle F \rangle$ is a step towards a *quantitative* mathematical representation of the effort a finite agent requires to overcome the necessary computational tasks leading to the potential solution of a computable task. The effort is what the resulting[7] (potentially but not actually omniscient) agent c:finite (the CML-'counterpart' to the ideal agent c) needs to invest to compute—(potentially all) the computable formulas in CML.

At first glance, one might find that Duc's approach constitutes a qualitative change to NML too, as it adds a completely new formal structure to NML (namely, at least the new operator). Nevertheless, this is not the whole truth:

The crux of the matter is that the operator $\langle F \rangle$, at its maximal depth, might be interpreted as equivalent to the original operator $\Box$, respectively K. At maximal depth, the *meaning of* $\Box$ and the *meaning of* $\langle F \rangle \Box$ could be identical. If this turns out to be a good interpretation of the operators, logical omniscience would then be avoided within CML for c:finite and maintained by the notion of what constitutes the state of maximal computational depth, which varies from task to task. The fact that it might be a huge difficulty to determine the state of maximal depth for some tasks is not a deficit of NML but rather a 'promethean' ambition: an ambition that consists in wanting to make (epistemically) one's own what may be qualitatively tempting but is ontologically distinct and therefore difficult—if not impossible—to achieve.

[6] Within the same, well-behaved framework.
[7] The agent c consists i.a. in the way the operator $\langle F \rangle \Box$ behaves.

What I state is that—what in the literature corresponds to—implicit knowledge is possible for c:finite, whilst explicit knowledge is what c:finite[8] already has 'computed', whereas knowledge in the sense of the classical NML operator K does only provide a meaningful interpretation of c:finite if the depth is maximal.[9] From this perspective, logical omniscience only becomes a serious problem in epistemic logic when one tries to define a finite and still deeply consistent agent, thereby mixing up what may constitute the 'ontologically detached' side of logical investigation with the goal of speaking about something finite.[10] There is, from this perspective, no obvious reason for which NML should 'abolish itself' assuming IW.

This does however not imply that, besides of *theoretically* avoiding logical omniscience, there can be no other reasons in favour to develop IW semantics. Since, for example, the arithmetic modelling of a computational depth is not at all an unproblematic undertaking [10][4, p. 119 ff], IW semantics are of use for the aim of *practically* avoiding the 'problem' of omniscience.

7. Prima Facie Conceivability of Ideal But Finite Agents

Pro-IW philosophers are confident that IW approaches should finally be fully included in the canon of our philosophical toolbox. Some of them see the insistence on PW alone as something akin to intellectual western-centrism; something to be overcome in favour of the philosophical benefits that would come along with an understanding of the "structure of the impossible"[11, p. 2661], therefore, along with IW approaches. Why should one not, despite what was said, reap the fruits of IW approaches untroubled?

So far, we have examined the extent to which there are reasons to introduce IW *given* NML as tried logical tools—we have looked at the discussion of counterpossibles (1), the metaphysical dispute argument (2) and the problems around logical omniscience (3). Especially (2) and (3) have not proven convincing. Let's now look at a fourth reason why proponents of IW argue to adopt IW given PW: that PW would not be sufficient to adequately describe what some call conceivability or imaginability (4). Via (4) we will hopefully shed some light onto why, in the defence of PW, a focus was put on (1), (2) and (3): there is a connection between them.

Priest uses 'conceivable' as roughly synonymous with 'imaginable'[11, p. 2658]. He observes that, on the basis of NML-only approaches,

> 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. [11, p. 2652]

Therefore, according to him, the dispute around "Fermat's last theorem" (FLT) is a case of dispute in which a logical impossibility ($\neg$FLT) must be

---

[8] At the moment of the evaluation of the content of its knowledge.

[9] With regard to these thoughts, compare the OK operator in: [1, p. 13 ff].

[10] Or infinite but physical, if such a thing is possible.

regarded as conceivable, that means, must be understood as being the content of an epistemic state $\mathbb{S}$, a state that should not be confused with mere figurative/pictorial representations in our minds. $\mathbb{S}$, further, would involve objects appropriate to its nature—according to Priest states of conceiving involve conceivable objects φ of some general kind: $\mathbb{S}(\varphi)$.[11] If we now want to say that an agent a conceives some φ at a time tn, we will write '$a\mathbb{S}(\varphi)\mathrm{t}n$', which shall be read as 'a conceives φ at time n'.

According to Priest, $\mathbb{S}(\mathrm{FLT})\mathrm{t}n$ and $\mathbb{S}(\neg\mathrm{FLT})\mathrm{t}n$ both describe 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, for which $\mathbb{S}(\neg\mathrm{FLT})\mathrm{t}n$ might hold? Similar positions have been forwarded by other pro-IW philosophers [3].

Let us now partly recycle the structure of the dispute argument out of Sect. 5 and reassume our fictive epistemic agents Lilith and Harry. Through this we will discuss Priest's thoughts on what he holds to be conceiving impossibilities [11]:

- **i** Harry and Lilith believe that if a φ can be conceived, then φ must be logically possible. Therefore they believe that if $\mathbb{S}(\varphi)$, φ or what is expressed by φ necessarily corresponds to some quantification over W.
- **ii** Harry$\mathbb{S}(\mathrm{FLT})\mathrm{t}^1$ holds (although Harry does not know if FLT is a logical truth by then—he might have some intuition about it).
- **iii** Lilith$\mathbb{S}(\neg\mathrm{FLT})\mathrm{t}^1$ holds (although Lilith does not know if ¬FLT is a logical truth by then—she might have some intuition about it).
- **iv** FLT is shown to be expressing a logical truth $\top$ at t^2 by a mathematician. The mathematician conceives FLT. Harry and Lilith are informed about the truth of FLT at t^2 and they are in a position to understand the language of the mathematician.[12]
- **v** If FLT is shown to be true at t^2, then it was true at t^1. If FLT is conceivable at t^2, then it is also conceivable at t^1.[13]

c: Following Priest, Harry, Lilith and the mathematician, on the grounds of i-v, should agree to introduce IW and therefore deny i:

Because, if FLT is shown to be a logical truth at t^2 by the mathematician (iv), then FLT is a logical truth at t^1 (by iv-v). But if only logical possibilities

[11]Although the use of parentheses in $\mathbb{S}(\varphi)$ might suggest it, a classical predicative structure is by no means meant here. The notation is intended to signify a conceivable (or allegedly conceivable) object φ as the content of the state of conceiving ($\mathbb{S}$), seeking a way to formally express what Priest probably imagines conceiving to be.

[12]That they must be in a position to understand the terms according to Priest follows from his postulate: "I can conceive of and imagine anything that can be described in terms that I understand" [11, p. 2659].

[13]We should be talking about ideal conceivability here. But more concerning this notion of conceivability will be said further down the text. Moreover, while Priest does not explicitly express this assumption v, it seems to be a condition of the possibility of his thought experiments on impossibility that we require the regularities governing knowledge and conceivability to be non-variable over time as well as the continuity of the relevant abilities of the epistemic agents involved, and so we list these conditions here.

can be conceived (i), then $\mathbb{S}(\neg\text{FLT})t^2$ and, by v, $\mathbb{S}(\neg\text{FLT})t^1$ would not hold for any agent. But this would contradict Lilith$\mathbb{S}(\neg\text{FLT})t^1$, as Lilith$\mathbb{S}(\neg\text{FLT})t^1$ implies that there is an agent (Lilith) that conceived ¬FLT (according to iii). The only permissible state of conceiving for any reasoner at t^1 concerning FLT based on this reasoning (accepting i) would therefore be $\mathbb{S}(\text{FLT})t^1$ (ii) [11].

From this Priest seems to argue that we should not give up iii, but instead modify i; we should weaken our notion of conceivability and accept IW as a tool to model impossibilities Λ that can be conceived just as possibilities; $\mathbb{S}(\Lambda)$. Because, according to Priest, we are interested in describing what Harry and Lilith conceive at t^1 in order to understand how humans reason in general. Therefore, he argues, we should accept a most general notion of what is conceivable by introducing IW; If something is conceivable for an agent, then it is either a logical possibility or not ("I can conceive of and imagine anything that can be described in terms that I understand"[11, p. 2659]). But great caution is required here, because a lot is happening.

In what follows we want to assume that for Lilith, Harry, the mathematician, *and every other epistemic agent*, $\mathbb{S}(\neg\text{FLT})$ might well be unrealizable at any given time. We also want to argue that Harry$\mathbb{S}(\text{FLT})t^1$ might in fact only hold, if Harry at t^1 conceives the way the mathematician conceives at t^2. It will be suggested that under these new assumptions, there will be nevertheless a viable way of explaining what Lilith (and Harry) might conceive even though they do not seem to conceive in the way the mathematician does. For if, following Chalmers, one assumes that Lilith$\mathbb{S}(\neg\text{FLT})t^1$ is a case of *prima-facie conceivability*[6, ch.1], but the state of the mathematician at time t^2 is a case of *ideal conceivability*, one can also explain what is going on in the above situation—without relying on IW:

Suppose Harry, Lilith and the mathematician are all finite but potentially ideal agents according to what we said in Sect. 6. Prima facie and ideal conceivability according to such a notion of ideal rationality can be defined as follows:

> 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. [6, ch.1]

According to this definition, the mathematician truly *conceives* FLT at t^2. This is the case because, given FLT is a logical truth, according to Sect. 6 answering if FLT is a logical truth is possible for a potentially ideal but finite agent when some adequate depth of reasoning is reached: $\vdash\text{FLT} \Rightarrow \vdash\langle F\rangle\Box\text{FLT}$. Therefore, if $\vdash\text{FLT} \Rightarrow \vdash\langle F\rangle\Box\text{FLT}$, and the adequate depth of reasoning is reached, then the mathematician conceives FLT ideally, because, as we have seen, $\langle F\rangle\Box\text{FLT}$ then actually might corresponds to $\Box\text{FLT}$, and $\Box$ respectively to K. FLT is therefore conceived by the mathematician, given FLT "is conceivable on ideal rational reflection" and given sufficient effort is being made (see ibid.).

For Lilith, things are different; although we assume her also to be an ideal but finite reasoner, at time t^1, as she does not put enough effort into the task, she does not detect all the implications of FLT the mathematician saw at time t^2. For if she would have, she would not prima-facie conceive what she believes to be the negation of FLT according to the above definition. (As 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 at relevant depth at t^1, then she would know that ¬FLT does, and therefore she would not be able to prima-facie-conceive ¬FLT according to the definition.)

But it would be off to infer that Lilith$\mathbb{S}(\neg\text{FLT})t^1$ holds, since her 'conceiving' does not actually refer to FLT the way the conceiving of the mathematician refers to FLT at t^2. Her conceiving concerns implications of what *she holds to be* FLT or what she holds to be its negation. She doesn't conceive ¬FLT in the same way an ideal infinite reasoner and the mathematician (at adequate depth) conceive.

That doesn't mean we can't understand what Lilith and Harry conceive, if we do not deny premise i. It just means we can't understand what Lilith and Harry would conceive, if they conceived a logical impossibility—but why should we care? We can also describe the relevant steps of the above dispute situation in this way by sharpening our concept of conceiving; Lilith conceives *what she holds to be* non-contradictory implications of FLT and Harry conceives *what he holds to be* non-contradictory implications of FLT. By luck or intuition Harry might actually conceive FLT or implications of FLT at t^1, but neither Harry nor Lilith conceive in the way the mathematician conceives FLT. In this sense, we can still learn a lot from Priest's thoughts, namely that, given a complex undecided theorem φ, perhaps we should talk less in terms of conceiving φ and more about conceiving what we believe φ to be (until we understand φ at an adequate depth and therefore possibly conceive φ the way an ideal agent does or could). By the way, this also conforms more to the conventional handling in mathematics of the non-synonymous terms theorem and conjecture.[14]

It would now have to be shown in detail how the dispute argument of this section relates to the dispute argument about metaphysical discourse of Sect. 5 of this paper. That would go beyond the scope of this article. Hopefully, however, it has turned out that both arguments, besides sharing part of their form, rest on the peculiar assumption that in order to further discuss what is metaphysically/logically possible or necessary, we allegedly must be able to understand the structure of what is logically impossible.

Intuition, from this perspective, can possibly put us on the right track, insofar as they sometimes might save ourselves the work of what is associated with erroneously believing to conceive impossibilities. However, they can also lead to the opposite.

This should very likely have implications for how we deal with counterpossibles. Because, as already mentioned half-jokingly, only fallible (finite) agents are known to write philosophy papers: if counterpossibles become the focus of the ratio of such fallible agents, then it is important to weigh up

[14]The term 'theorem' in the case of FLT (before the proof) was an unusual choice of word in mathematics. Statements of which one does not yet know whether they are capable of being evaluated are in fact normally called conjectures and not theorems.

whether things are really necessitated on the basis of impossibilities in the case of counterpossibles. For, insofar as our interpretation of what an impossible antecedent expresses is tied to what we can conceive, it is unclear whether the discussion of counterpossibles is fruitful, as it is often unclear or implicit what we conceive when we prima-facie conceive. In any case, the discussion of counterpossibles would still be valuable, simply for the sake of understanding what we prima-facie conceive when we discuss what we *believe* the antecedents to be expressing.

8. Conclusion

Among the examined lines of argumentation, no thoroughly convincing argument as to why impossible world semantics should be adopted could be found:

On the one hand, the argument that aimed at the indistinguishability of certain metaphysical statements interpreted within the framework of normal modal logic did not turn out to be as strong as expected. It overlooks, as seen above, a crucial detail in the course of evaluating scenarios of conflicting beliefs.

On the other hand, arguing in Hintikka's tradition, logical omniscience is not fundamentally a problem; to preserve the notion of ideal rationality in the form of the classical knowledge operator of normal modal logic is probably something (philosophically) desirable.

Impossible world semantics may be quite helpful in solving problems related to the modelling of finite epistemic agents or in evaluating an *intuitively* problematic subset of counterpossible conditionals, in itself a subset of counterfactual conditionals—it is questionable, however, whether this is worth its weakness.

The discussion of a third argument (for impossible worlds) on the part of proponents of impossible worlds—the argument of modelability of conceivability via these worlds—has additionally reinforced the suspicion that there are no strong reasons, given our canonical logical repertoire, for impossible worlds approaches. There seems, furthermore, to be a connection in disfavor of impossible world accounts between the respective argument lines for impossible worlds: not only do they seem to frequently appeal to similarly problematic dispute-structures, but they seem to implicitly appeal to the idea that our intuitions, which often seem to suggest that we are imagining x, would guarantee for the ideal conceivability of the alleged x. But even without pursuing the alleged structures of 'impossibilities', we may be able to explain quite a bit on the basis of our tried tools. From this perspective, the introduction of impossible worlds appears a bit like trying to open a nut with a sledgehammer.

Acknowledgements

I would like to thank Luis Rosa, Jasper Lohmar and Nicola Thomaschewski for their comments on earlier drafts of this paper. I would also like to thank the members of Metronom for their always refreshing company, beyond the scope of this essay.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

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

[2] Berto, F.: Impossible worlds and the logic of imagination. Erkenntnis **82**, 1277–1297 (2017)

[3] Berto, F.: On conceiving the inconsistent. Proc. Aristot. Soc. **114**, 103–121 (2014)

[4] Berto, F., Jago, M.: Impossible Worlds. Oxford University Press, Oxford (2019)

[5] Berto, F., French, R., Priest, G., Ripley, D.: Williamson on counterpossibles. J. Philos. Logic **47**, 693–713 (2018)

[6] Chalmers, D.: Does conceivability entail possibility? In: Conceivability and Possibility, pp. 145–200. Oxford University Press, Oxford (2019)

[7] Climenhaga, N.: Intuitions are used as evidence in philosophy. Mind **127**, 69–104 (2018)

[8] HO Ngoc DUC: Reasoning about rational, but not logically omniscient, agents. J. Log. Comput. **7**, 633–648 (1997)

[9] Hintikka, J.: Impossible possible worlds vindicated. J. Philos. Logic **4**, 475–484 (1975)

[10] Jago, M.: Hintikka and cresswell on logical omniscience. Log. Log. Philos. **15**, 325–354 (2006)

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

[12] Rendsvig, R., Symons, J.: Epistemic Logic. The Stanford Encyclopedia of Philosophy, Summer Edition (2021), plato.stanford.edu/archives/sum2021/entries/logic-epistemic/

[13] Tanaka, K.: Logically impossible worlds. Austral. J. Philos. **15**, 489–497 (2018)

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

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

Received: July 8, 2023.
Accepted: December 27, 2023.