Mission Impossible

Graham Priest

Abstract Saul Kripke's work on the semantics of non-normal modal logics introduced the idea of non-normal worlds, worlds here where certain connectives behave differently from the way in which they behave in the worlds of normal modal logics. Such worlds may be thought of as impossible worlds, though Kripke did not, himself, talk of them in this way. Since Kripke's invention, the notion of an impossible world has undergone much fruitful development and application. Impossible worlds may be of different kinds—or maybe different degrees of impossibility; and these worlds have found application in many areas where hyperintensionality appears to play a significant role: intentional mental states, counterfactuals, meaning, property theory, to name but a few areas. But what, exactly, is an impossible world? How is it best to characterise the notion? To date, the notion is used more by example than by definition. In this paper I will investigate the question, and propose a general characterisation, suitable for all standard purposes and tastes.

Dedication: It gives me great pleasure to dedicate this essay to Saul. Without his work, contemporary studies of non-classical logic would be unthinkable. It has been an inspiration for all who work in logic.

1 Introduction

Saul Kripke's work on the semantics of non-normal modal logics introduced the idea of non-normal worlds, worlds where logically impossible things may hold. Such worlds can naturally be thought of as impossible worlds. Since Kripke's invention in (1965), the notion of an impossible world has undergone much fruitful development

Graham Priest
Departments of Philosophy, the CUNY Graduate Center, the University of Melbourne, and the Ruhr University of Bochum, e-mail: priest.graham@gmail.com

and application. Impossible worlds may be of different kinds—or maybe different degrees of impossibility (for example, they may be closed under conditions of different degrees of strength, or none at all); and these worlds have found application in many areas where hyperintensionality appears to play a significant role: intentional mental states, counterfactuals, meaning, property theory, to name but a few areas.

But what, exactly, is an impossible world? How is it best to characterise the notion? To date, the notion is used more by example than by definition. In this paper I will investigate the question and propose a general characterisation, suitable for all standard purposes and tastes. It will be:

- independent of all applications
- neutral on the ontological question of what worlds, as such, are
- neutral on the question of what inferences are valid
- neutral on the question of what it is that makes them so

In particular, it can be deployed whatever one takes the correct logic to be.

The guiding thought of what follows is this:[1]

- a possible world is one that is closed under logical consequence, and an impossible world is one that is not.

However, this thought needs to be articulated in different ways, depending on whether one is dealing with the worlds of a formal logical semantics or worlds in reality themselves—whatever this means. In the first case, we have a logical consequence relation to hand, given by the formal semantics. In the second, we do not: we have to extract it from the worlds themselves, as it were.

In their (2018) and at greater length in their (2019, §1.4), Berto and Jago suggest four possible definitions of an impossible world: (1) a world where something impossible happens; (2) a world where the laws of logic fail; (3) a world where the logic is non-classical; (4) a world where contradictions hold. The account I shall provide is a version of (2). I reject (1) for reasons that we will come to in §5.1; (3) presupposes the truth of classical logic, whereas I seek an account which does not prejudge that issue. A similar comment applies to (4), even if it were not already problematic because a failure of $A \vee \neg A$ is just as impossible in classical logic as a holding of $A \wedge \neg A$.[2]

The paper has the following structure. In the next section I provide a little historical background on worlds, possible and impossible. Given what I have to say, it behoves me to explain what I take a logic to be. This is spelled out in the next section. In §4, I show how the guiding thought of the paper is to be articulated in the case of the worlds of a formal semantics. Then in §5, I turn to the much harder question of how it is to be articulated in the case of worlds themselves. In §6, I note a technical

---

[1] This definition is suggested, in effect, in Priest (1997a, p. 482).

[2] Zalta (1997) gives a different definition of worlds, possible and impossible; but it, too, is not logic-neutral. For him, all worlds (possible and impossible) have to be complete; but in paracomplete logics even possible worlds do not have to be complete. And for him, inconsistent worlds are impossible; but in paraconsistent logics, inconsistent worlds can be possible.

iteration of the material in §5. Then in the final section, §7, I tie up a number of the loose ends which have appeared in the discussion.

2 Background: "Kripke Semantics"

2.1 Possible Worlds

First, then, some historical background.

Kripke's work on the semantics of modal logic was a game-changer. Modal logic was studied in the West in both Ancient and Medieval logic, but the topic had largely fallen into oblivion by the start of the 20th Century. It was rescued from this state in the early part of the century by the work of C. I. Lewis, who proposed—famously—his five axiomatic systems of modal logic, $S1$-$S5$. Modal and other intentional notions came under heavy attack by Quine and others around the middle of the century, and I think it fair to say that by the 1960s they were under a deep cloud.

At about this time, Carnap, Prior, and others, were working with ideas that could provide a semantics for modal logic;[3] but it was Kripke's papers (1959) and (1963), deploying explicitly the notion of a possible world, which really established such semantics as logically kosher. It is no accident that this genre of semantics is now termed 'Kripke Semantics'. The Quinean skepticism did not go away immediately, and certainly the metaphysics of possible worlds caused many brows to furrow. However, the world semantics found so many applications, not only concerning modal notions, but also other intensional notions—such as belief and knowledge, counterfactuals, meaning, property theory, and a number of other areas—that in due course Quinean scruples quietly disappeared. The techniques were just too powerful and versatile not to be used—though philosophers could still, and do still, argue about what to make of possible worlds metaphysically.[4]

2.2 Impossible Worlds

The Lewis systems $S4$ and $S5$ are now termed *normal modal logics.* $S4$ and $S5$ are but two of a large family of such logics uncovered by Kripke semantics, and not the most fundamental. However, the systems $S1$, $S2$, and $S3$ are not normal systems. Something that distinguishes them syntactically is that the Rule of Necessitation, $\vdash A \Rightarrow \vdash \Box A$, fails. In yet another of his papers, Kripke (1965) provided a semantics

[3] There was also a slightly earlier algebraic semantics for some modal logics in the work of McKinsey an Tarski (1944, 1946); but I think it fair to say that this was seen as just a bit of pure mathematics.

[4] References: On the history of modal logic, see Knuuttila (2012). On the genesis of contemporary modal logic, see Ballarin (2007). On the Quinean attack on modal logic, see Barcan Marcus (1995). On possible worlds, see Menzel (2016).

for these.[5] To provide these semantics, Kripke introduced the idea of a different kind of world, which he called *non-normal*. Hence, these systems are now usually known as non-normal modal logics. At non-normal worlds, the truth conditions of formulas of the form $\Box A$ and $\Diamond A$ were not given in terms of a binary accessibility relation, as for normal worlds. Rather, if w is a non-normal world, then every formula of the form $\Box A$ is false, and every formula of the form $\Diamond A$ is true at w.[6] This means that logical truths, such as $\Box(A \vee \neg A)$, may fail at such worlds; and logical falsities, such as $\Diamond(A \wedge \neg A)$, may hold. This makes it natural to think of such worlds as logically impossible worlds—though Kripke did not refer to them as such.

Kripke's impossible worlds were of a very simple kind, but many more applications of the notion of an impossible world were soon on the market. In particular, they play a central role in the semantics of relevant logic[7]—though again, they were not conceptualised as impossible worlds till some time after their introduction.[8] In these semantics, if A is *any* logical truth, it may fail at some impossible worlds; and if A is *any* logical falsity, it may hold at some impossible worlds. The impossible worlds of relevant logic are still fairly orderly, however. In particular, they are closed under *modus ponens* for logical truths of the form $A \rightarrow B$. A use for much more anarchic impossible worlds came a little later in the semantics of intentional notions. Here, virtually *any* combination of formulas may hold or may fail.[9]

Indeed, there are now many applications of semantics with impossible worlds.[10] And I think it fair to say that if worlds are to be used for the analysis of many notions, especially hyperintensional notions, some of those worlds must be impossible worlds.[11]

Of course, the addition of impossible worlds of various kinds to the menagerie of worlds has complicated debates about the metaphysics of worlds. But this is not my concern here.[12] And of course, people have objected to impossible worlds. This, also, is not my concern here.[13] My concern is much more basic than these matters,

---

[5] This is a slight exaggeration. The semantics were for $S2$ and $S3$. $S1$ proved to be more recalcitrant, and semantics were not found till later. See Cresswell (1972).

[6] See Priest (2008, ch. 4).

[7] See Routley, Plumwood, Meyer, and Brady (1982).

[8] As far as I know, the first place where this terminology appears in print in this context is Priest (1992).

[9] See, e. g, Priest (2005), Bjerring (2013), Berto (2017).

[10] See Priest (1997a), the other essays in Priest (1997b), Berto and Jago (2018), and Parts II and III of Berto and Jago (2019). See also Berto (2017), Berto, French, Priest, and Ripley (2018), Bjerring (2013, 2014), Brogard and Salerno (2013), Jago (2013), Krakaur (2013), Nolan (2013), Sillari (2008), Stuart, McLoone and Grützner (201+), Weber and Omori (2019), Yagisawa (2010).

[11] See, e.g., Nolan (2013, 2014).

[12] The matter is discussed in Berto and Jago (2018), and Part I of Berto and Jago (2019). For what it's worth, I take all worlds other than the actual world to be non-existent objects. See Priest (2005, §7.3.)

[13] The matter is also discussed in Berto and Jago (2018). Again for what it is worth, Berto and Jago note six objections that have been made. I have nothing new to say about the first five of these. But I do about the sixth objection (which is discussed at greater length in Berto and Jago (2019, esp. §8.5)). This is to the effect that at impossible worlds the truth value(s) of a formula do not depend

and is this: What, exactly, is an impossible world? What is it about a world that makes it impossible? And let me emphasize right at the start, that I am talking about logical impossibility, not some other kind of impossibility.

3 What is a Logic?

Let us now turn to matters of philosophical substance, starting with the question of what, for the present purposes, we may take a logic to be—and in this essay, I will mean by 'logic' *deductive logic* (since the debate about possible and impossible worlds arises in the context of such logics).

We may start by fixing a language, $\mathfrak{L}$. For present, we may assume that this is some standard propositional language. We may now identify a logic with a consequence relation on $\mathfrak{L}$, $\vdash$. That is, it is a relation, $\Pi \vdash \Gamma$, where Π and Γ are sets of formulas of $\mathfrak{L}$.[14] Intuitively, one may think of this as meaning something like: if all of the members of Π hold, then some of the members of Γ hold. In the most general case, Π and Γ may be infinite, but provided we are dealing with a compact consequence relation, the infinite case may be defined from the finite case in the familiar way.[15] Since non-compact logics have played little role in modal logics, we will restrict our attention to finite Π and Γ here. I will make a brief comment on the more general case in §7.2.

It is standard to put further constraints on $\vdash$. Thus, the "Tarski conditions" are often imposed. $\vdash$ is Reflexive:

- $\Sigma \cup \{A\} \vdash \{A\} \cup \Delta$

Monotonic:

- if $\Sigma \vdash \Delta$, $\Theta \supseteq \Sigma$, and $\Delta \subseteq \Xi$, then $\Theta \vdash \Xi$

and Transitive:

- if $\Sigma \vdash \Delta \cup \{A\}$ and $\{A\} \cup \Theta \vdash \Xi$ then $\Sigma \cup \Theta \vdash \Delta \cup \Xi$

---

on the truth value(s) of its sub-formulas (maybe at other worlds). In that sense, the semantics is not compositional. And many have felt that a semantics should be compositional to account for learnability. I take this to be a mistaken thought. What learnability requires is that the semantics be recursive, or at least recursively enumerable, which the semantics with impossible worlds are, or may be. It is true that in the most general case the specification of the value of each formula at an impossible world must be fixed independently; but the same is true, of course, of propositional parameters. It might be suggested that in any learnable language there should be only a finite number of propositional parameters; if so, in the modal case, one might just restrict oneself to a language with formulas of some finite complexity (which is required, in any case, if the semantics are to be humanly learnable). Alternatively, one may simply restrict oneself to specifications of truth values of formulas at impossible worlds which are recursively enumerable.

[14] Those who envisage a world-semantics for inferences whose premises/conclusions are multisets (or some other type of data-structure), are free to make appropriate changes. As is standard, I will omit the set brackets when the sense is clear without them.

[15] $\Pi \vdash \Gamma$ iff for some finite subsets, $\Pi' \subseteq \Pi$ and $\Gamma' \subseteq \Gamma$, $\Pi' \vdash \Gamma'$.

Standard definitions of validity using world semantics deliver all these conditions. So I will assume them here. Again, I will make a brief comment on relaxing these conditions in §7.1.

Another constraint often imposed on a consequence relation is closure under substitution. That is, let s be any function from propositional parameters to the formulas of $\mathfrak{L}$. If A is any formula of the language, let A_s be A with every propositional parameter, p, replaced by $s(p)$. Let $\Sigma_s = \{A_s : A \in \Sigma\}$. Then closure under substitution is the condition:

- if $\Pi \vdash \Gamma$ then $\Pi_s \vdash \Gamma_s$

This closure condition will surface briefly in what follows, but I will not pack it into the definition of a consequence relation.

Finally, if the language has conjunction and disjunction operators which work in the usual fashion, one may reduce the finite non-empty case to the singleton case as follows: $\bigwedge\{\pi \in \Pi\} \vdash \bigvee\{\gamma \in \Gamma\}$.[16] However, we will not assume this here, since such conditions can fail in logics with world-semantics. Thus, if supervaluation is employed, one may have $A \vee B \vdash A \vee B$, but not $A \vee B \vdash A, B$; and if subvaluation is employed, one may have $A \wedge B \vdash A \wedge B$, but not $A, B \vdash A \wedge B$.[17]

4 Formal Semantics

We may now turn to our question: What is an impossible world? We need to distinguish here between the worlds of a formal semantics, and whatever it is that these are supposed to represent in reality. The answer concerning the first is a lot easier than the answer concerning the second. So let us take this first.

We suppose that we have a world-semantics for some logic, $\vdash$, such that validity is defined as truth—however that is cashed out—preservation over subset (possibly improper) of worlds. Worlds are simply mathematical entities of a certain kind—perhaps numbers or sets. Which are the possible ones? The obvious answer is that they are the ones over which validity is defined. But how might these be characterised intrinsically? We can answer this by considering the relation $w \Vdash A$, expressing the claim that A is true at w.

Given this, we can define a possible world (with respect to $\vdash$) as one which is closed under $\Vdash$. That is, if w is one of the worlds at issue, w is possible iff:

- if $\Pi \vdash \Gamma$ and $w \Vdash A$ for all $A \in \Pi$, then $w \Vdash B$ for some $B \in \Gamma$

A world is impossible if it is not possible.

Note that the definition makes no assumptions about what $\vdash$ is, and in that sense is logic-neutral. However, as one would expect, given that we are dealing with a pre-given semantics, the possible and impossible worlds at issue are not, since these

---

[16] And one can reduce the empty case to the singleton case if the language contains the constants $\bot$ and $\top$, the first entailing everything, and the second being entailed by everything.

[17] See Priest (2008, §7.10.)

will depend on whatever logic is at issue. The collection of worlds itself will also be determined by the formal semantics.

This definition of 'impossible world' is entirely unproblematic in most cases; however, it can give a few surprising results. For a start, in some semantics (for a logic, $\vdash$), a world that is officially classed as impossible (non-normal), may actually be possible according to this definition. For example, if what is true at such a world is entirely arbitrary, it can be closed under $\vdash$ by accident, as it were.[18] In that case, one can just move the world into the official class of possible worlds without changing the consequence relation.

Perhaps more interestingly, there are logics given by the Kripke semantics for $S2$ and $S3$, but where validity is defined over all worlds (not just non-normal ones).[19] Contrary to what one might have expected, as defined, all the worlds are possible. There are just two classes of these, normal and non-normal. (Modal formulas simply have disjunctive truth conditions.[20]) $S2$ and $S3$ notwithstanding, in these logics, since validity is preserved at all worlds, there are none at which logically aberrant things happen.

One might think that the opposite can also happen: that some officially possible worlds are impossible. Thus, in some semantics there is a trivial world, where everything is true, and/or (if there are no logical truths) an empty world where nothing is true.[21] These are possible according to this definition, though one might think them impossible. Actually, I think that in logics delivered by the semantics, these worlds *are* possible: nothing happens which is ruled out by the logic. However, if one does not like this, one could always treat these worlds as a special cases, as impossible by definition.

Finally, I note also that although an impossible world is not closed under $\vdash$, it may be closed under some weaker condition. Thus, in relevant logics, all worlds (possible and impossible) are closed under *modus ponens* for logical truths of the form $A \to B$.[22] However, in a general semantics for intentional operators, if w is an impossible world, the set $\{A : w \Vdash A\}$ may be quite arbitrary.[23] The amount of anarchy of an impossible world may therefore come by degrees.

5 Worlds Themselves

We now turn to the much tricker problem of defining what an impossible world is when we are dealing with worlds themselves. That is, we are not dealing simply

---

[18] See Priest (2005, 2nd edn, ch. 9).

[19] These are the Lemmon "E Systems". See Priest (2008, 4.4.5).

[20] See, further, Priest (2005, 2nd edn, 13.3.2).

[21] See, for example, Weber and Omori (2019).

[22] See Priest (2008, ch. 10).

[23] See Priest (2005, ch. 1).

with the mathematical machinery of some formal semantics, but the worlds *in re*—whatever that means.

We have an actual world, @, and a bunch of other worlds. $\mathfrak{L}$ is a language adequate for describing what holds at words, with—perhaps—the exception of the facts about logical consequence. However, now that we have no model theory, and so no notion of an interpretation to deploy, $\mathfrak{L}$ has to be taken as an interpreted language.

To say which worlds are possible/impossible, we will apply again the guiding thought of the essay: that a possible world is one which is closed under logical consequence. When dealing with the worlds of a formal semantics, we have a consequence relation ready to hand; but now this luxury is no longer available to us. And let me emphasise again that we are talking about *logical* impossibility—impossibility according to the laws of logic. There are, of course, other kinds. Indeed, we will meet one in a moment.[24]

I think that the beginning of an answer is to be found by analogy with physical impossibility. A physically impossible world is one where the laws of nature are different from those of the actual world. In such a world, some physical event or state of affairs which cannot happen at our world, can happen; or which can happen at our world, cannot happen. By analogy, a logically impossible world is one where the laws of logic are different from the laws of logic at the actual world. [25] The definition is neutral as to what that logic—"the correct logic"—is. The definite description does, however, presume that a world has a unique logic. So be it for the moment. I will come back to the matter in §7.3.

Of course, how one understands what it means to say that such and such are the laws of nature of a world is a somewhat contentious question.[26] However, it is not necessary to plunge into those murky depths here. Clearly, we *do* have to engage with the more unfamiliar question of what it means for such and such to be the logical laws of a world. To the extent that we need to answer the question, we can do this in terms of the relation $w \Vdash A$, for the worlds, w, in question. However, now that we are not dealing with a formal semantics, we cannot presuppose a definition provided by such. On how, in fact, the notion is to be analysed we need take no position, and so is another matter on which the account that follows is neutral.

5.1 A World's Extrinsic Logics

Since logic is at least something that is truth preserving, for a consequence relation, $\vdash$, to be the logic of a world, what is true at the world should be closed under $\vdash$. That is, for $\vdash$ to be the logic of w we require that:

[24] Some hold that there is a distinctive notion of metaphysical impossibility—that for example, it is metaphysically impossible that Hesperus is Phosphorus, though this is not a logical impossibility. Actually, I am skeptical of this. (See Priest (2021).) However, in any case, that notion of necessity is not what is at issue here.

[25] The definition is mooted in Priest (1992, §2).

[26] On how one might understand the notion of a law of nature, see Carroll (2020).

- if $\Pi \vdash \Gamma$ and $w \Vdash A$ for all $A \in \Pi$, then $w \Vdash B$ for some $B \in \Gamma$

If this condition is satisfied, let us then say that $\vdash$ is an *extrinsic logic* of w.

Clearly, an extrinsic logic of a world is not unique, since if $\vdash' \subseteq \vdash$ and $\vdash$ is an extrinsic logic of w, so is $\vdash'$. The obvious thought then is to define the logic of the world as one of the extreme cases: either the weakest truth preserving relation, or the strongest. But the weakest is the null relation, which relates nothing to anything; and the strongest is the one that is materially truth preserving. These have no claim to be a logic in any substantial sense.

Even if we set these extreme cases aside, there are inferences that are necessarily truth preserving but are not usually taken to be valid. Thus, the inference '*a* is red; so *a* is coloured' is truth preserving at possible worlds. It is not normally taken to be a deductively valid inference. One may rule out this sort of situation by insisting that in an extrinsic logic, not only are the inferences truth preserving, but their substitution instances are truth preserving as well. This also fails, however: such inferences may still be truth-preserving "by accident". Compare the case with that of a law of nature. Suppose that the laws of motion are those of Special Relativity, so that nothing accelerates through the speed of light. Now consider a world where the laws of motion are Newtonian, but which doesn't last very long, and where everything moves pretty slowly, so that nothing has time to reach the velocity of light. At such a world it is also true that nothing accelerates through the speed of light. However, this is simply a contingent feature of the world, and it does not make the laws of motion of the world those of Special Relativity. That is, the universal generalisation is contingent and not lawlike.

Similar things may also happen in the case of logic.[27] If $\vdash$ is the logic of @, it should be truth preserving not just at @, but at all possible worlds—whatever they are. Now, suppose that there are possible worlds where contradictions are true, but that @ is a consistent world. Then any inference of the form: '$p \wedge \neg p$; so q' is truth preserving at @, but not at all possible worlds. A similar sort of situation can arise in classical logic, as was pointed out by Etchemendy.[28] Suppose that it is a contingent fact that the actual world has only one object, but that some possible worlds are larger. Then 'p; so $\exists x \forall y\, x = y$' is truth preserving at @, but not at all possible worlds. A third example of the same phenomenon arises if one has a world semantics with an actuality operator, $\hat{@}$, in the language, such that for any (possible) world, w, $w \Vdash \hat{@}A$ iff $@ \Vdash A$.[29] For then 'A; so $\hat{@}A$' is truth preserving at the actual world, but not at all possible worlds.

---

[27] So there is a difference between a world where the laws of logic are different (from the actual laws), and a world where the logically impossible happens. I take the former to be the criterion of an impossible world, for the reasons just given. If I understand his account correctly, Tanaka (2018) endorses the latter.

[28] Etchemendy (1990, pp. 111 ff).

[29] See Crossley and Humberstone (1977).

5.2 The Intrinsic Logic of a World

In general, then, we have a collection of consequence relations which, for all we have seen so far, might be the logic of a world. Given that we are taking each world to have a unique logic, more needs to be said. A way to do so is by making reference to what each world "thinks" its own logic to be. In other words, one needs the language to contain the means to express the laws of validity.

In the case of physical laws, we can express the fact that a universal generalisation is a law of nature (lawlike), and not a contingent generalisation, by using modal operators. Thus, in a world where the laws of motion are Newtonian, an object *can* (physically) accelerate through the speed of light. In a world where the laws are those of Special Relativity, it *cannot*.

A similar trick will not work in the case of laws of logic, since modal distinctions will not suffice for the task. Thus, in many logics (e.g., LP, FDE), models may have a trivial world, where everything holds. So if the modal accessibility relation accesses this, $\models \Diamond A$ is true for all A, even though these are different logics.

The trouble is that logical consequence is essentially relational, and cannot be reduced to something non-relational. In particular, even in just the one-premise one-conclusion case, one cannot characterise $A \vdash B$ by something like $\neg\Diamond(A \wedge \neg B)$. In relevant logics, for example, $p \wedge (p \rightarrow q) \vdash q$, but there may be (consistent) modally accessible worlds, not closed under *modus ponens*, where $p \wedge (p \rightarrow q)$ and $\neg q$ both hold.

Hence, to express logical consequence, we need a notion that is essentially relational. To this end, let us extend our language with a new connective, $\Rightarrow$, such that $A_1, \ldots, A_m \Rightarrow B_1, \ldots, B_n$ is a formula, where the left and right hand sides are finite enumerations (possibly empty) of formulas of $\mathfrak{L}$.[30] Call the extended language $\mathfrak{L}^{\Rightarrow}$. If you want to read $\Rightarrow$ in English, it's something like: given that A_1, ..., and that A_n then it follows that that B_1, ..., or that B_m.[31] Notice that the connective is multigrade, in that it may have a variable number of arguments both left and right of the $\Rightarrow$.[32] (As we shall note in due course, if we assume that the language contains vocabulary for talking about sets, we might take what we need to be expressed by a relation between two sets, not a connective between formulas. However, I make no such assumption here.)

---

[30] As already noted, I make no assumption about the behaviour of conjunction and disjunction if they are in the language. So this cannot be reduced to the single premise/conclusion case.

[31] I note that even if we assume that the commas can be coded as conjunctions and disjunctions, in no standard logic with a world semantics is there a connective, $\rightarrow$, such that we are guaranteed that for a possible world, w, $w \Vdash (A_1 \wedge \ldots \wedge A_n) \rightarrow (B_1 \vee \ldots \vee B_m)$ iff $A_1 \wedge \ldots \wedge A_n \models B \vee \ldots \vee B_m$. If the logic has the "deduction theorem" for $\rightarrow$ we will have the right to left direction, but even then, nothing guarantees the left to right direction.

[32] One usually meets the characteristic of being multigrade in connection with predicates, not connectives. See Grandy (1976) and Taylor and Hazen (1992). However, multigrade connectives are known. See Humberstone (2011, pp. 630, 783) for discussions and references. See also Humberstone (2021). Taylor and Hazen operate with predicates of a fixed arity, each of whose places can be filled by an arbitrary number of arguments. This is essentially how the connective $\Rightarrow$ functions. One might call such predicates/connective 'polyunsaturated'.

$\Rightarrow$-statements are true or otherwise at a world, just as are the statements of $\mathfrak{L}$. As to what makes them so, we may take no stand here. But let us say that the relation between two sets of formulas, $\vdash_w$, defined by:

- $A_1, \ldots, A_m \vdash_w B_1, \ldots, B_n$ iff $w \Vdash A_1, \ldots, A_m \Rightarrow B_1, \ldots, B_n$

is the *intrinsic* logic of the world. *Warning*: there is no guarantee that $\vdash_w$ *is* a logic as we have defined it. Indeed, there is no reason that it must even be truth preserving at w. Let us call a world where it is, *coherent*. That is, a world, w, is coherent if $\vdash_w$ is an extrinsic logic of w.

Since truth preservation is a necessary constraint on the (actually) true theory of validity, the actual world, @, is coherent. (Though, of course, which $\Rightarrow$-sentences *are* true is a contentious matter of the correct logic, on which we also take no stand here.)

And now, at last, we can provide the required definitions. w is a possible world if it is coherent and has the same intrinsic logic as @. (So, in particular, @ is a possible world.) It is impossible otherwise. Thus, the actual world and its intrinsic logic determine which of the other words are possible/impossible.[33]

Let me finish this section with three observations concerning the definition.

First, assuming that $\Box A$ is true iff A is true at all possible worlds. Then, as one would expect, if $@ \Vdash A_1, \ldots, A_m \Rightarrow B_1, \ldots, B_n$, $@ \Vdash \Box(A_1, \ldots, A_m \Rightarrow B_1, \ldots, B_n)$.

Secondly, impossible worlds may be of different kinds. There are coherent ones, where truth is closed under the intrinsic logic of the world; it is just that that logic is not the same as that of @. But there are also incoherent worlds, where true $\Rightarrow$-statements do not preserve truth. One might already have expected there to be such worlds. After all, at an impossible world, anything might happen![34]

Finally, it is quite possible that two coherent worlds have all the same extrinsic logics, but different intrinsic logics. Analogously for physical necessity, two worlds might have all the same true universal generalisations, but which ones are lawlike may be different in different worlds.[35] Thus, two worlds may agree on what holds in the language $\mathfrak{L}$, whilst differing in what holds in the language $\mathfrak{L}^{\Rightarrow}$. In the same way, the description of a finite mathematical structure might be the same at both a world whose logic is classical and a world whose logic is intuitionist. However, in one world $\Rightarrow A \vee \neg A$ may hold, but not at the other.

[33] Clearly, any other coherent world could play the same role in this definition. So we could define a notion of possible-with-respect-to w, for each coherent world, w. Hence there will be sets of classical possible worlds, intuitionistic possible worlds, etc. However, real possibility is defined with respect to the actual world. And for better or for worse, there is only one of these, even if actuality is indexical, and every world takes itself to be actual.

[34] Of course, incoherence may come by degrees. Suppose that the world structure is given by that of some standard non-normal modal logic. Then the impossible worlds are closed under classical propositional logic, but not modal validities (e.g., $A \Rightarrow \Box(B \vee \neg B)$.)

[35] As even "neo-humeans" about physical necessity like David Lewis agree. (See Carroll (2020, §2).)

6 Higher Orders

The preceding considerations are sufficient to answer the central question of this essay. But a question will naturally occur to any logician at this point.

In §5, the language $\mathfrak{L}^{\Rightarrow}$ expresses a consequence relation between formulas in the language $\mathfrak{L}$. But we have extended $\mathfrak{L}$ to do this. So there is now the matter of the consequence relation between formulas of the extended language.[36]

To get a sense of what is at issue here, note that two of the Tarski closure conditions (Monotonicity and Transitivity) are themselves naturally framed as such inferences. Thus, Transitivity—or Cut, as it is more usually called in this context—can be expressed by the inference:

$$\frac{\Sigma \Rightarrow \Delta \cup \{A\} \quad \{A\} \cup \Theta \Rightarrow \Xi}{\Sigma \cup \Theta \Rightarrow \Delta \cup \Xi}$$

The facts about $\Rightarrow$ do not determine the facts about these meta-inferences. To see this, just note that the Cut Theorem for classical logic shows us that the same inferences are valid both with and without Cut.

Bearing these things in mind, we may now extend the construction used for $\mathfrak{L}$ to sentences of the language $\mathfrak{L}^{\Rightarrow}$. We can, in fact, proceed exactly as before at this higher level. Say that a consequence relation at this level, $\vdash$, is an extrinsic logic of world w if:

- if $\Pi \vdash \Gamma$ and $w \Vdash A$ for all $A \in \Pi$, then $w \Vdash B$ for some $B \in \Gamma$

where all of the formulas in Π and Γ contain $\Rightarrow$.[37] We now have to extend the language to express the intrinsic logic at one level up.

To keep track of things in what follows, let us write what we have hitherto written as $\Rightarrow$, as $\Rightarrow^0$, and call what has gone before, level 0. We now add a new connective to the language, $\Rightarrow^1$, such that $A_1, \ldots, A_m \Rightarrow^1 B_1, \ldots, B_n$ is a formula, where all of the As and Bs contain $\Rightarrow^0$. An instance of Cut might then be expressed by:[38]

$$(\Sigma \Rightarrow^0 \Delta \cup \{A\}, \{A\} \cup \Theta \Rightarrow^0 \Xi) \Rightarrow^1 (\Sigma \cup \Theta \Rightarrow^0 \Delta \cup \Xi)$$

Call this language $\mathfrak{L}^{\Rightarrow^1}$. Let us say that the relation $\vdash^1_w$ such that:

- $A_1, \ldots, A_m \vdash^1_w B_1, \ldots, B_n$ iff $w \Vdash A_1, \ldots, A_m \Rightarrow^1 B_1, \ldots, B_n$

is the intrinsic logic of w at level 1. A world is coherent at this level if the intrinsic logic is an extrinsic logic. @ is coherent at this level for the same reason that it was

[36] In truth, there is absolutely nothing in the construction which assumes that $\Rightarrow$ is not a part of $\mathfrak{L}$. However, I have not assumed this, since the connective $\Rightarrow$ is not of a kind that is familiar in logic. But see §7.2.

[37] Note that this construction does not allow for (the expression of) inferences of "cross type". With a bit of fiddling, we could do so by allowing for a more liberal syntax. Similar comments apply to what follows.

[38] I insert brackets for ease of readability. The set terms are to be thought of as the appropriate enumerations.

coherent at level 0. A world is now possible if it is coherent at levels 0 and 1, and its intrinsic logics are the same as those of @ at both levels.[39]

Of course, we now face the same situation with respect to inferences between sentences of $\mathfrak{L}^{\Rightarrow^1}$, and so on for higher levels. As I noted, the facts about $\Rightarrow^0$ do not uniquely determine the facts about $\Rightarrow^1$. Exactly the same is true at higher levels.[40] We may then just iterate our construction. The general procedure is this.

Let $\mathfrak{L}^0$ be $\mathfrak{L}$, and let $\mathfrak{L}^{i+1}$ be $\mathfrak{L}^i$ augmented with the connective $\Rightarrow^i$. For all $i > 1$, say that a consequence relation, $\vdash$, is an extrinsic consequence relation at level i for world w if:

- if $\Pi \vdash \Gamma$ and $w \Vdash A$ for all $A \in \Pi$, then $w \Vdash B$ for some $B \in \Gamma$

where all of the formulas of Π and Γ contain $\Rightarrow^{i-1}$. Let us say that the relation $\vdash^i_w$ such that:

- $A_1, \ldots, A_m \vdash^i_w B_1, \ldots, B_n$ iff $w \Vdash A_1, \ldots, A_m \Rightarrow^i B_1, \ldots, B_n$

is the intrinsic logic of w at level i. A world is coherent at this level if the intrinsic logic is an extrinsic logic. @ is coherent at this level as before.

A world is possible if it is coherent at all levels, and its intrinsic logics are the same as those of @ at all levels. If one wishes, one could express all the facts about intrinsic inferences in a single language. We simply take the union of all the languages $\mathfrak{L}^i$, for all finite i.

One could iterate the procedure higher up the ordinals if wished, but this is not necessary to express all the individual facts about inferences of finite order. Since premises and conclusions are always finite, if there is a valid inference between formulas of the language in the hierarchy, it will already be expressed by something of the form $A_1, \ldots, A_m \Rightarrow^i B_1, \ldots, B_n$, for finite i.[41]

7 *Et Cetera*

7.1 The Tarski Rules

It remains to tie up a few loose ends that have arisen.

Until now, I have assumed that a consequence relation satisfies the familiar Tarski rules of Reflexivity, Monotonicity and Transitivity, since standard logics with world semantics do satisfy these conditions. To a certain extent, these conditions may be relaxed. Let me explain how.

---

[39] In fact, we could take the intrinsic logic of a world (*sans* level) to be the union of its intrinsic logics at levels 0 and 1. The superscripts on $\Rightarrow$ could then be dropped. (Again, an analogous observation applies to what follows.) If the level 0 consequence relation is closed under uniform substitution, it would be natural (though by no means mandatory) to suppose that this extends to substitutions of the extended language.

[40] As shown by Barrio, Pailos, and Szmuc (2020).

[41] On iterating levels of meta-inferences into the transfinite, see Scambler (2020).

Montonicity is a defining condition of a deductive logic, and so is not negotiable in the present context. However, there are well known ways of defining deductive logics in which the other two fail.

Essentially, one has to understand what it is to be true at a world in different senses for premises and conclusions. One can do this quite naturally in a many-valued logic. In such logics, some of the values are designated, D, and validity is normally defined in terms of the preservation of having values in D. However, one may employ different sets of designated values for the premises and conclusions, D_Π and D_Γ. An inference is then valid if whenever all the premises have values in D_Π (at a world) some of the consequences have values in D_Γ (at that world). If $D_\Pi \subsetneq D_\Gamma$, Transitivity may fail, and if $D_\Gamma \subsetneq D_\Pi$, Reflexivity may fail.

Thus, suppose we deploy the semantics of K_3/LP, with values t, i, and f. The matrices for these logics are the same, but in K_3 the designated value is just t, whilst in LP the designated values are t and i.[42] Now, if we take $D_\Pi = \{t\}$, as in K_3, and $D_\Gamma = \{t, i\}$, as in LP, this gives the logic ST ("Strict/Tolerant"). In this, Transitivity fails. Reversing matters gives the logic TS ("Tolerant/Strict"). In this, Reflexivity fails.[43]

With the understanding that truth in the premises and truth in the conclusions are thus understood in different ways, everything above carries over in a straightforward fashion.

7.2 Non-Compact Logics

I have assumed that we are dealing with compact logics, so that we do not have to deal with infinite premise/conclusion sets in any essential way. We could dispense with this assumption. A logic then becomes a relation, $\Pi \vdash \Gamma$, where Π and Γ are arbitrary sets of formulas of $\mathfrak{L}$. This affects our ability to define the intrinsic logic of a world. One possibility here is to move to an infinitary language, where $\Rightarrow$ can have any number of sentences, finite or infinite, on its left and its right.

A more familiar approach would be obtained by moving to a first-order construction and, in particular, to a language which can refer to sentences, sets thereof, and relations between these things. One can then express the intrinsic logic of a world in the straightforward way, by the condition $w \Vdash \langle\Pi\rangle \Rightarrow \langle\Gamma\rangle$, where angle brackets are a name-forming functor, and $\Rightarrow$ is a binary predicate of the familiar kind.[44]

The details of executing this sort of approach are familiar enough, though I shall not pursue them here. The extra complexity would be large, and the gain in understanding for the question at hand virtually nil.

---

[42] On K_3 and LP, see Priest (2008, ch. 7, esp. §§7.3, 7.4).

[43] See, e.g., Cobreros, Égré, Ripley, and van Rooij (2013), and Ripley (2013)

[44] And note the following. Suppose that $\mathfrak{L}$ is the language of set theory. Then a model-theoretic definition of validity can be given in that language. Hence, the binary predicate $\Rightarrow$ is already, in effect, present in $\mathfrak{L}$, and does not need to be added to the language.

7.3 Logical Pluralism

Finally, let us turn to the matter of logical pluralism. Logical pluralism is the view that there is, in some sense, more than one correct logic. However, this idea can be cashed out in many different ways, and this is not the place to discuss the matter at length.[45]

The issue is, in fact, irrelevant to the material of §4, since, by assumption, we are dealing with one consequence relation. If there is a plurality of these, each of which is correct, each will determine its own collection of possible/impossible worlds.

Matters are different thereafter, since there I have *assumed* that every world (including the actual world) has one intrinsic logic. However, it is not difficult to dispense with this assumption if one is so inclined.

If one is a pluralist, one will have a bunch of possible logics, $\vdash_L$, in play. For each of these, one may have a corresponding connective, $\Rightarrow_L$. We may take each condition of the form $w \Vdash A_1, \ldots, A_m \Rightarrow_L B_1, \ldots, B_n$ to define an intrinsic logic of w. A world is coherent if every intrinsic logic is an extrinsic logic. @ is a coherent world, as before; and now a world is possible if it is coherent and has all the same intrinsic logics as @; it is impossible otherwise.

8 Conclusion

The considerations of the previous section indicate how the construction of the earlier sections can be enriched and elaborated in various directions. However, such things are not necessary for the main problem of the paper. This was to provide an answer to the question of what an impossible world is. The guiding thought of this essay was that a possible world is one that is closed under logical consequence, and an impossible world is one that is not. As have seen, however, this thought has to be articulated in different ways depending on whether we are dealing with a formal world-semantics for a particular consequence relation, or whether we are dealing with worlds *an sich*, where we had to work to dig out the relevant consequence relation. But in both cases the answer given to the question was independent any assumptions about the correct logic—and many other philosophical matters.

Doubtless, many matters concerning impossible worlds will continue to be debated. I hope that the present paper at least gives us an understanding of what it is we are talking about.[46]

---

[45] For some discussion see Priest (2006, ch. 12).

[46] For very helpful written comments on earlier drafts of this paper, I'm grateful to Franz Berto, Allen Hazen, Lloyd Humberstone, Mark Jago, Daniel Nolan, Koji Tanaka, Yale Weiss, Ed Zalta, and two anonymous referees. Talks based on this paper were given in (virtual) seminars at the universities of Utrecht, Melbourne, Padua, Amsterdam, and New York. I'm also grateful to the members of the audiences there for their helpful comments.

References

1. Ballarin, R. 2017. Modern Origins of Modal Logic. In E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/logic-modal-origins/.
2. Barcan Marcus, R. 1995. A Backwards Look at Quine's Animadversions on Modalities. In *Modalities: Philosophical Essays*, pp. 215–232. Oxford: Oxford University Press.
3. Barrio, E., Pailos, F., and Szmuc, D. 2020. A Hierarchy of Classical and Paraconsistent Logics. *Journal of Philosophical Logic* 49: 93–120.
4. Berto, F. 2017. Impossible Worlds and the Logic of Imagination. *Erkenntnis* 82: 1277–1297.
5. Berto, F., French, R., Priest, G., and Ripley, D. 2018. Williamson on Counterpossibles. *Journal of Philosophical Logic* 47: 693–713.
6. Berto, F., and Jago, M. 2018. Impossible Worlds. In E. Zalta (Ed.), S*tanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/impossible-worlds/.
7. Berto, F., and Jago, M. 2019. *Impossible Worlds*. Oxford: Oxford University Press.
8. Bjerring, J. C. 2013. Impossible Worlds and Logical Omniscience: an Impossibility Result. *Synthese* 190: 2505–2524.
9. Bjerring, J. C. 2014. On Counterpossibles. *Philosophical Studies* 168: 327–353.
10. Brogaard, B., and Salerno, J. 2013. Remarks on Counterpossibles. *Synthese* 190: 639–660.
11. Carroll, J. W. 2020. Laws of Nature. In E. Zalta (Ed.), S*tanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/laws-of-nature/.
12. Cobreros, P., Égré, P., Ripley, D., and van Rooij, R. 2013. Reaching Transparent Truth. *Mind* 122: 841–866.
13. Cresswell, M. 1972. The Completeness of $S1$ and Some Related Systems. *Notre Dame Journal of Formal Logic* 4: 485–496.
14. Crossley, J., and Humberstone, L. 1977. The Logic of 'Actually'. *Reports on Mathematical Logic* 8: 11–29.
15. Etchemendy, J. 1990. *The Concept of Logical Consequence*. Cambridge, MA: Harvard University Press.
16. Grandy, R. 1976. Anadic Logic and English. *Synthese* 32: 393–402.
17. Humberstone, L. 2011. *The Connectives*, Cambridge, MA: MIT Press.
18. Humberstone, L. 2021. Propositional Variables Occurring Exactly Once in Candidate Modal Axioms. *Filosofiska Studier* 8: 27–73.
19. Jago, M. 2013. Impossible Worlds. *Noûs* 49: 713–728.
20. Knuuttila, S. 2012. A History of Modal Traditions. In Gabbay, D., Pelletier, F. J., and Woods, J. (Eds.), *Handbook of the History of Logic*, Vol. 11, *Logic: A History of its Central Concepts*, pp. 309–340. Amsterdam: North Holland.
21. Krakauer, B. 2013. What Are Impossible Worlds? *Philosophical Studies* 165: 989–1007.
22. Kripke, S. 1959. A Completeness Theorem in Modal Logic. *Journal of Symbolic Logic* 24: 1–14.
23. Kripke, S. 1963. Semantical Analysis of Modal Logic I: Normal Modal Propositional Calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 9: 67–96.
24. Kripke, S. 1965. Semantical Analysis of Modal Logic II: Non-Normal Modal Propositional Calculi. In Addision, J. W., Henkin, L., and Tarski, A. (Eds.), *The Theory of Models*, pp. 202–220. Amsterdam: North Holland.
25. McKinsey, J., and Tarski, A. 1944. The Algebra of Topology. *Annals of Mathematics* 45: 141–191.
26. McKinsey, J., and Tarski, A. 1946. On Closed Elements in Closure Algebras. *Annals of Mathematics* 47: 122–162.
27. Menzel, C. 2016. Possible Worlds. In E. Zalta (Ed.), S*tanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/possible-worlds/.
28. Nolan, D. 2013. Impossible Worlds. *Philosophy Compass* 8: 360–372.
29. Nolan, D. 2014. Hyperintensional Metaphysics. *Philosophical Studies* 171: 149–160.
30. Priest, G. 1992. What is a Non-Normal World? *Logique et Analyse*, 35: 291–302.
31. Priest, G. 1997a. Editor's Introduction. In Priest (1997b), pp. 482-487.

32. Priest, G. (ed.) 1997b. *Notre Dame Journal of Formal Logic*, 38(4).
33. Priest, G. 2005. *Towards Non-Being*. 2nd edn 2016. Oxford: Oxford University Press.
34. Priest, G. 2006. *Doubt Truth to be a Liar*. Oxford: Oxford University Press.
35. Priest, G. 2008. *Introduction to Non-Classical Logic*. Cambridge: Cambridge University Press.
36. Priest, G. 2021. Metaphysical Necessity: a Skeptical Perspective. *Synthese* 198: 1873–1885.
37. Ripley, D. 2013. Paradoxes and Failures of Cut. *Australasian Journal of Philosophy* 91: 139–164.
38. Routley, R., Plumwood, V., Meyer, R. K., and Brady, R. 1982. *Relevant Logics and their Rivals*, Vol. 1. Atascadero, CA: Ridgeview.
39. Scambler, C. 2020. Transfinite Meta-inferences. *Journal of Philosophical Logic* 49: 1079–1089.
40. Sillari, G. 2008. Quantified Logic of Awareness and Impossible Possible Worlds. *Review of Symbolic Logic* 4: 514–529.
41. Stuart, M., McLoone, B., and Grützner, C. 201+. Counterpossibles in Science: an Experimental Study. To appear.
42. Tanaka, K. 2018. Logically Impossible Worlds. *Australasian Journal of Logic* 15: article 3.12, https://ojs.victoria.ac.nz/ajl/article/view/4870.
43. Taylor, B., and Hazen, A. 1992. Flexibly Structured Predication. *Logique & Analyse* 139–140: 375–393.
44. Weber, Z, and Omori, H. 2019. Observations on the Trivial World. *Erkenntnis* 84: 975–994.
45. Yagisawa, T. 2010. *Worlds and Individuals, Possible and Otherwise*. Oxford: Oxford University Press.
46. Zalta, E. 1997. A Classically-Based Theory of Impossible Worlds. *Notre Dame Journal of Formal Logic* 38: 640–660.