

Mind: 131: 292-301 (2022).

Book Review

***Impossible Worlds*, by Francesco Berto & Mark Jago**

Oxford: Oxford University Press, 2019, pp. 324, ISBN: 978-0-19-881279-1

Koji Tanaka

Australian National University

Koji.Tanaka@anu.edu.au

1 Impossible!

Francesco Berto and Mark Jago are arguably two of the most prolific and dynamic philosophers alive. They have written numerous articles and books mainly on logic, metaphysics, epistemology, the philosophy of language and the philosophy of computation often invoking the concepts of *impossibility* and *impossible worlds*. In *Impossible Worlds*, they put their effort together and present a comprehensive discussion of impossibility and impossible worlds. The framework of *possible* worlds is an indispensable resource for modern philosophers. Berto and Jago have done the impossible task of demonstrating the equal importance of worlds other than the standard possible worlds. The book is a must read for any contemporary philosophers who make use of possible and impossible worlds.

After an introduction (Introduction and Chapter 1), the book starts with a discussion of the metaphysics of impossible worlds (Chapter 2): what are impossible worlds and how do they represent what they represent? After rejecting a range of suggestions to answer these questions (Chapter 2) and offering their own preferred view on the question (Chapter 3), Berto and Jago apply impossible worlds to various issues in logic, in particular, modal logics (Chapter 4), epistemic logics (Chapter 5), relevant logics (Chapter 6), and the logic of imagination (Chapter 7). In the final and most substantive part of the book, they consider a number of philosophical applications. They apply impossible worlds to an analysis of knowledge and belief as well as their content (Chapter 8), the theory of information including theories of identity, inference and vagueness (Chapter 9), the problems of logical omniscience and bounded rationality (Chapter 10), the issue of fiction and fictional entities (Chapter 11), and an analysis of counterpossibles (or counterfac-

tuals) (Chapter 12). They make use of impossible worlds to offer novel solutions to the outstanding problems on these issues.

The number of issues that Berto and Jago discuss is large. They try to sell impossible worlds (or the worlds other than the standard possible worlds) to a large number of people who have given possible world treatment of their favourite subjects. In this book review, I could take up any number of their analyses to discuss. Instead of examining what they say about any of the specific issues, however, I would like to discuss something more general in nature that applies to the whole book. That is the definition of impossible worlds: what worlds do Berto and Jago count as impossible? My question is not what counts as a world. I assume that we know what a world is as much as we know what it is when it is applied to possibilities. My question is: given a world, what makes it impossible? Without having any sense of what the answer might be, it is hard to know what exactly is talked about and achieved in the book. So that is a crucial question for understanding what the book is about. Berto and Jago provide various definitions of impossible worlds allowing them to invoke all sorts of impossible worlds. It is not clear, however, that the impossible worlds that do much of the work fit any of the definitions. At least, they have not explained how their impossible worlds are impossible according to those definitions. That is what I will show in this book review. I will do so not as a way to criticise the work of Berto and Jago but to contribute to the discussions of impossible worlds and help clarify the significance of the use of impossible worlds for contemporary philosophers.

2 Definitions of Impossible Worlds

Berto and Jago offer several definitions of impossible worlds. The definition that plays a significant role in early chapters (Chapters 2 and 3) is an ontological one: an impossible world is a world that represents what could not possibly exist. After those metaphysical chapters, they turn to logical characterisations of impossible worlds. In terms of logical characterisations, they explicitly list four definitions that they claim to be able to find in the literature:

IMPOSSIBLE WAYS: An impossible world is characterised as ways things could not have been.

LOGIC VIOLATORS: An impossible world is a world where the laws of logic fail.

CLASSICAL LOGIC VIOLATORS: An impossible world is a world where the laws of classical logic fail.

CONTRADICTION-REALISERS: An impossible world is a world where sentences of the form A and $\neg A$ hold. (pp. 31-32)

How do these definitions play out in the book? Before answering this question, I will first explain what kinds of worlds the above definitions give us.

I start with the last one: **CONTRADICTION-REALISERS**. An impossible world according to this definition is a dialetheic world. That is, it is a world where a contradiction (or contradictions) is (are) realised. If we were to think that such a world is non-trivial (i.e., not everything is true), then we cannot think that classical logic holds. Some classically valid inferences, in particular the *ex contradictione quodlibet* (ECQ) ($A, \neg A \models B$ for any A and B), have to be violated to make it a non-trivial world. An impossible world according to this definition is, thus, a special case of the third definition: **CLASSICAL LOGIC VIOLATORS**.

Berto and Jago describe an impossible world as a **CLASSICAL LOGIC VIOLATOR** to be a special case of an impossible world as a **LOGIC VIOLATOR**. According to **LOGIC VIOLATORS**, an impossible world is a world where the laws of logic fail. But this depends on which logic is under discussion. So, for some logic L , a world is impossible with respect to the logic L if it violates any of the laws of L . If L is classical, this definition delivers the same worlds as **CLASSICAL LOGIC VIOLATORS**. For instance, a world where the Law of Excluded Middle fails is a world where one of the classical laws is violated.

However, whether or not an impossible world as a **CLASSICAL LOGIC VIOLATOR** is a special case of an impossible world as a **LOGIC VIOLATOR** depends on what counts as a world. A world may or may not be closed under some logical consequence. If it is not closed under any logical consequence, then the world is not characterised in terms of the laws of logic. So the fact that the Law of Excluded Middle fails to obtain may not be indicative of the violation of any logical laws. However, if a world is closed under some logical consequence, then an impossible world as a **CLASSICAL LOGIC VIOLATOR** is a special case of an impossible world as an **IMPOSSIBLE WAY**. Consider a world at which the Law of Excluded Middle fails. If such a world is closed under some logical consequence, then it may be an intuitionistic world, i.e., a world that is closed under intuitionistic logical consequence. A (hard-core) classical logician would think that it is impossible for a world to be the way represented by intuitionistic logic. So it represents an impossible way things have been.

In order to see the significance of this point, let's further unpack the first two definitions of impossible worlds Berto and Jago offer: **IMPOSSIBLE WAYS** and **LOGIC VIOLATORS**. First, **IMPOSSIBLE WAYS**. There are various ways to conceive of the ways that could not have been and I cannot examine all of them. Impossible worlds do not have to be described in terms of closure of some logical consequence (for instance, Salmon (1984) and Yagisawa (1988)) or in logical terms at all (for instance, an ontological definition: an impossible world is a world that represents what could not possibly exist). However, if we take a world to be closed under some logical consequence, we need to be careful about the nature of the difference

between IMPOSSIBLE WAYS and LOGIC VIOLATORS.

Consider the analysis of impossible worlds as impossible ways presented by Restall (1997). Assume that we have all of the possible worlds. These possible worlds are standard. At a possible world x , the following holds for conjunction, disjunction and negation:

1. $x \Vdash A \wedge B$ iff $x \Vdash A$ and $x \Vdash B$
2. $x \Vdash A \vee B$ iff $x \Vdash A$ or $x \Vdash B$
3. $x \Vdash \neg A$ iff $x \not\Vdash A$

where \Vdash is an interpretation relating propositions to worlds. We assume that conditionals are defined in terms of disjunction and negation.

A *world* is then a nonempty set of possible worlds and truth and falsity at a world w (\Vdash^+ and \Vdash^-) are defined as follows:

1. $X \Vdash^+ p$ iff $x \Vdash p$ for some $x \in W$
2. $X \Vdash^- p$ iff $x \not\Vdash p$ for some $x \in W$
3. $X \Vdash^+ \neg A$ iff $X \Vdash^- A$
4. $X \Vdash^- \neg A$ iff $X \Vdash^+ A$
5. $X \Vdash^+ A \wedge B$ iff $X \Vdash^+ A$ and $X \Vdash^- B$
6. $X \Vdash^- A \wedge B$ iff $X \Vdash^- A$ or $X \Vdash^- B$
7. $X \Vdash^+ A \vee B$ iff $X \Vdash^+ A$ or $X \Vdash^+ B$
8. $X \Vdash^- A \vee B$ iff $X \Vdash^- A$ and $X \Vdash^- B$

A world, w , could be like x or y for two different possible worlds. However, it cannot be both like x and like y at the same time. It may be that that A is true at x , but it is false at y . There is nothing impossible about this. However, if a world consists of x and y , then it represents a way that A is both true and false. Assuming that a possible world is a world where no contradictions obtain, w represents a way things cannot be.

This analysis of impossible worlds as IMPOSSIBLE WAYS gives us impossible worlds where the laws of logic are *different*. Priest (2008) characterises the worlds where the laws of logic are different in analogy to the worlds where the laws of physics are different. If there is a world where an object travels faster than the speed of light, the laws of physics at that world must be different from those at our world. There may still be laws of physics that hold at such a world. However, they must be different from the laws of physics at our world. Similarly, there may be worlds where the laws of logic are different. If our world is a classical world, then an intuitionistic world, a world which is closed under intuitionistic consequence, is a world where the laws of logic are different. If we think of an intuitionistic world as representing an IMPOSSIBLE WAY along this line, then such a world can be described as impossible (see Cresswell (1973)).

Contrary to what Priest (2008) claims, which Berto and Jago follow, impossible worlds as LOGIC VIOLATORS are *not* the worlds where the laws of logic are *different*, however. To see this, consider the simplified semantics for basic relevant logic **B** (Priest and Sylvan (1992)). (I largely follow the presentation by Berto and Jago.) A *Routley-Meyer frame* for **B** is a structure $\langle W, N, R, * \rangle$ where W is the set of worlds, $N \subseteq W$ is the set of *normal* worlds ($W - N$ is the set of *non-normal* worlds), $R \subseteq W \times W \times W$ is a ternary relation on worlds satisfying the following condition called *normality condition*:

(NC) If $w \in N$, then Rww_1w_2 iff $w_1 = w_2$.

$*$ is the *Routley Star* which is an operation on W : $w^{**} = w$ for each $w \in W$.

To have a model, we add an evaluation function, ν , that assigns truth values to atomic sentences (or propositions) at a world. We then extend ν to the whole language as follows:

$$\nu_w(\neg A) = 1 \text{ if } \nu_{w^*} = 0, \text{ and } 0 \text{ otherwise.}$$

$$\nu_w(A \wedge B) = 1 \text{ if } \nu_w(A) = \nu_w(B) = 1, \text{ and } 0 \text{ otherwise.}$$

$$\nu_w(A \vee B) = 1 \text{ if } \nu_w(A) = 1 \text{ or } \nu_w(B) = 1, \text{ and } 0 \text{ otherwise.}$$

$$\nu_w(A \rightarrow B) = 1 \text{ if for all } w_1, w_2 \in W \text{ such that } Rww_1w_2 \text{ if } \nu_{w_1}(A) = 1 \text{ then } \nu_{w_2}(B) = 1, \text{ and } 0 \text{ otherwise.}$$

Validity is defined in terms of truth preservation at all *normal* worlds. The semantics is then sound and complete with respect to **B**.

The importance of (NC) can be felt once we see the way the semantics invalidate the irrelevant conditional: $A \rightarrow (B \rightarrow B)$. At a normal world, $w \in N$, because of (NC), the ternary relation R essentially collapses to a binary relation. And, for all $w_1 \in W$ such that Rww_1w_1 , if $\nu_{w_1}(B) = 1$ then $\nu_{w_1}(B) = 1$. So $\nu_w(A \rightarrow (B \rightarrow B)) = 1$. However, at a non-normal world, $w' \in W - N$, it may be that $\nu_{w_1}(q) = 1$ but $\nu_{w_2}(q) = 0$ such that $Rw'w_1w_2$ for some atomic q . So $q \rightarrow q$ may not hold at the non-normal world w' .

Given that validity (and logical truth) is (are) defined in terms of truth preservation at all *normal* worlds, $B \rightarrow B$ is a logical truth of **B**. However, at a *non-normal* world, a logical truth may come out false. If we think of logical truths and valid inferences as the laws of logic, then the laws of logic may fail at a non-normal world. Non-normal worlds used in the semantics for **B** are impossible in the sense that they violate the laws of logic that hold at those worlds. Non-normal worlds used in the semantics for non-normal modal logics (modal logics weaker than **K**) are similar in this respect (Tanaka (2013, 2018)).

Now, at a non-normal world, the laws of **B** are violated. However, such a world is not closed under any alternative logical consequence. This means that

the world that violates the laws of logic may *not* be the world where the laws of logic are different. If we describe them as impossible worlds, thus, we need to keep them separate. It is best to think of a world where the laws of logic are different as an IMPOSSIBLE WAY rather than as a LOGIC VIOLATOR. (See Sandgren and Tanaka (2019) for more details.) For the remainder of this review, I will focus on the violation of logical laws rather than logical difference as the main issue with Berto and Jago’s treatment of impossible worlds has to do with LOGIC VIOLATOR.

3 Impossible Worlds *a la* Berto and Jago

Having examined the definitions of impossible worlds, we can now go back to Berto and Jago. While they provide four definitions of impossible worlds, the definition that does most work in much of the book is different from the characterisations of impossible worlds that I provided in the previous section based on their definitions. To see this, consider their discussion of non-normal worlds used in the semantics for non-normal modal logics (modal logics weaker than **K**). After introducing non-normal worlds used in these semantics, Berto and Jago make the following claim:

A key idea in impossible worlds semantics of various kinds is that certain complex formulas are assigned arbitrary values at non-normal worlds. (p. 102)

The idea that plays a major role in the book is actually not this but a generalisation of that claim:

At a non-normal world, *all* sentences are assigned arbitrary values.

For instance, in Chapter 5, they introduce a *Rantala frame* to be used in epistemic logic. A *Rantala frame* is a structure $\langle W, N, R \rangle$ where W is the set of worlds, $N \subseteq W$ is the set of normal, possible worlds ($W - N$ is the set of non-normal, impossible worlds) and $R \subseteq W \times W$ is a binary relation on worlds. Once a Rantala frame is equipped with an evaluation function, we have a *Rantala model* $\langle W, N, R, v \rangle$. At normal worlds, atomic sentences are assigned 1 or 0 by the evaluation function and complex sentences are evaluated recursively in the standard way. However, at a non-normal world, *all* sentences, whether atomic or complex, are assigned a truth value by the evaluation function directly.

A non-normal world characterised in this way is, however, different from the non-normal world used in the semantics for non-normal modal logics (and relevant logics). To see this, consider a non-normal model which is a structure $\langle W, N, R, v \rangle$ where W is the set of worlds, $N \subseteq W$ is the set of normal worlds ($W - N$ is the set of non-normal worlds), $R \subseteq W \times W$ is a binary relation on W ,

and ν is an evaluation function that assigns 1 or 0 to atomic sentences. The evaluation ν is then extended to the whole language, in particular the sentences with modal operators as follows:

$$\nu_w(\Box A) = 1 \text{ for any } w \in N \text{ if for all } w' \in N \text{ such that } Rww', \nu_{w'}(A) = 1$$

and

$$\nu_w(\Box A) = 0 \text{ for any } w \in W - N.$$

We define $\Diamond A$ to be $\neg\Box\neg A$. Then $\nu_w(\Diamond A) = 1$ for any $w \in W - N$.

There are two main differences between these non-normal worlds and Berto and Jago's non-normal worlds. First, the evaluation of $\Box A$ at a non-normal world used in the semantics for non-normal modal logics assigns a truth value directly but *not* arbitrarily. A sentence of the form $\Box A$ always comes out false at a non-normal world. There is nothing arbitrary about this.

Second, and more importantly, the evaluation of $\Box A$ at a non-normal world is not only non-arbitrary but it has a 'principle' behind it. Consider how validity is defined in various non-normal modal logics. (In what follows, I will largely follow Tanaka (2013, 2018).) For Lewis systems **S2** and **S3**, validity is defined in terms of truth preservation at all *normal* worlds. For Lemmon systems **E2** and **E3**, however, validity is defined in terms of truth preservation at *all* worlds. Let's call the validity for Lewis systems *weak* validity and that for Lemmon systems *strong* validity. We denote them as \models_w and \models_s respectively. As a logical truth is a special case of validity, if $\models_w A$ then A is true at all *normal* worlds, and if $\models_s A$ then A is true at *all* worlds. Let's say that logic is *general* if a logical truth expresses a truth no matter what the situation is. So if the necessity of A is general, then $\Box A$ is a logical truth. If we use $\models A$ to represent the generality of logical truth A , then $\models_s A$ expresses that the necessity of A is general: $\models_s A \Leftrightarrow \models_w A$.

If we understand a law of logic to be expressed by a logical truth and if $\models \Box A$, then $\Box A$ expresses a law of logic. But if laws of logic are expressed by the sentences of the form $\Box A$, the failure of $\Box A$ means the failure of the laws of logic. Since $\Box A$ is false at a non-normal world, the laws of logic can be said to be violated at a non-normal world. In this way, a non-normal world used in the semantics for non-normal modal logics is an impossible world.

It is important to see that the sentences that are evaluated directly by an evaluation function at non-normal worlds are of the form $\Box A$ and, importantly, only of this form. It is because the sentences that fail at non-normal worlds have such a form that those worlds can be said to be the worlds where the laws of logic fail and, thus, that they can be said to be impossible by satisfying LOGIC VIOLATORS. It is the failure of specific sentences (or sentences of a specific form) that characterises the impossible nature of non-normal worlds.

Berto and Jago's impossible worlds are not like this. This comes out clearly in the counter-examples they provide for various inferences. For instance, consider the following closure principle for a knowledge operator K :

(C1) If KA and $A \models B$, then KB .

This is one of the closure principles that are identified in the context of logical omniscience. It says that we would know the consequence of knowledge we already have whether or not we know the consequence relation to hold. The knowledge operator that satisfies this principle, thus, applies only to those who are logically omniscient. In order to invalidate this principle based on various considerations, Berto and Jago provide the following counter-model within a Rantala frame that has been applied to epistemic context:



where $N = \{w\}$ (p. 113). (I am assuming that if a sentence does not appear in the graphic representation of the model, it is false.) Since p is true at w_1 , Kp is true at w (where K works just like \Box). Also, given that validity is defined in terms of truth preservation at all normal worlds, $p \models p \vee q$. But $p \vee q$ is false at w_1 . So $K(p \vee q)$ is false at w . Hence the closure principle in question does not hold.

The above model invalidates the closure principle by using a non-normal world where $p \vee q$ is false even though p is true. If we take Berto and Jago's generalisation onboard, there is no systematic relation between the evaluation of p and the evaluation of $p \vee q$ as *all* sentences are evaluated arbitrarily at non-normal worlds. They use a Rantala frame to achieve this result. In a Rantala frame (or model), various facts that obtain at a non-normal world are not recursively composable or decomposable by logical operations. But what happens at the factual level may affect nomological relations such as the principles that hold for a knowledge operator as the above counter-example demonstrates. It is this feature of their models that allows them to invalidate the problematic epistemic principle that gives rise to logical omniscience.

At a non-normal world used in the semantics for non-normal modal logics and relevant logics, sentences that can be thought to express nomological relations may come out false. That is not because of what happens at the factual level. In fact, the truth value of A is irrelevant to the truth value of $\Box A$ at these non-normal worlds. Berto and Jago's impossible worlds are, thus, different from the impossible worlds that are made use of in those logics. This is the case even though they claim to derive their characterisation of impossible worlds from the non-normal worlds of those logics.

4 What are Berto and Jago's Impossible Worlds?

Are Berto and Jago's impossible worlds LOGIC VIOLATORS? Perhaps. Even if they manage to show that their impossible worlds should count as LOGIC VIOLATORS, we must notice that the way they fit the definition is different from the way that non-normal worlds used in non-normal modal logics and relevant logics do. Non-normal worlds used in those logic are impossible because they violate the laws of logic that hold at those worlds. It is true that the laws of logic may be violated at Berto and Jago's impossible worlds. However, if the laws of logic are violated at their impossible worlds, that is because the facts that obtain at those worlds do not cooperate with logical operations. Those violations or differences are indicative of the mismatch between facts and logical operations rather than the violations of the laws of logic as such. Thus, even if their impossible worlds can be said to be LOGIC VIOLATORS, they satisfy the definition differently from the non-normal worlds used in non-normal modal logics and relevant logics. Hence, contrary to what they suggest, Berto and Jago's impossible worlds are not the same species as the non-normal worlds of non-normal modal logics and relevant logics.

References

- Cresswell, M. 1973, *Logics and Languages* (London: Methuen).
- Priest, G. 2008, *An Introduction to Non-Classical Logic* (Second Edition) (Cambridge: Cambridge University Press).
- and Sylvan, R. 1992, 'Simplified Semantics for Basic Relevant Logics', *Journal of Philosophical Logic* 21: 217-232.
- Restall, G. 1995, 'Four-Valued Semantics for Relevant Logics (and Some of Their Rivals)', *Journal of Philosophical Logic* 24: 139-160.
- Salmon, N. 1984, 'Impossible Worlds', *Analysis* 44: 114-117.
- Sandgren, A. and Tanaka, K. 2019, 'Two Kinds of Logical Impossibility', *Noûs*, Online First.
- Tanaka, K. 2013, 'Making Sense of Paraconsistent Logic: The Nature of Logic, Classical Logic and Paraconsistent Logic', *Paraconsistency: Logic and Applications*, K. Tanaka *et al.* (eds.), 15-25 (Dordrecht: Springer).
- 2018, 'Logically Impossible Worlds', *Australasian Journal of Logic* 15: 489-497.
- Yagisawa, T. 1988, 'Beyond Possible Worlds', *Philosophical Studies* 53: 175-204.