

Against Classical Paraconsistent Metatheory

Koji Tanaka

Australian National University

Patrick Girard

University of Auckland

Abstract

There was a time when ‘logic’ just meant classical logic. The climate is slowly changing and non-classical logic cannot be dismissed off-hand. However, a metatheory used to study the properties of non-classical logic is often classical. In this paper, we will argue that this practice of relying on classical metatheories is problematic. In particular, we will show that it is a bad practice because the metatheory that is used to study a non-classical logic often rules out the very logic it is designed to study.

1 Paraconsistent Metatheory

Metatheory is the study of the mathematical properties of logic(s). A logic systematically specifies valid inferences. It is a system that can tell us which inferences are valid or invalid. A metatheory is a theory about that system. But a metatheoretical proof of a formal system also requires valid inferences. Badia, Weber & Girard (2022) have shown that such a metalogic can be paraconsistent in the sense that a paraconsistent metalogic can ‘recapture’ famous classical metatheoretical theorems such as Completeness, Löwenheim-Skolem and Compactness in a substructural paraconsistent metatheory.

The work of Badia, Weber & Girard paves the path towards proving completeness (and various other metatheoretical results) of paraconsistent logics in a paraconsistent metatheory. Even though there has been some work done to achieve such results (Bacon 2013, Rosenblatt 2021, Weber, Badia, & Girard 2016), the standard practice is to study metatheoretical results of paraconsistent logics in a metatheory that behaves classically (Beall 2009, Priest 1987, Routley & Routley 1972): ‘the [paraconsistent] logic literature is suffused with classical logic at the top’ (Weber 2021: 93).¹ The result of Badia *et al.* raises hopes of fully developing

¹Weber is talking about nonclassical logic generally. However, given the focus on paraconsistent logic, he mainly has the paraconsistent logic literature in mind.

a paraconsistent metalogic to study paraconsistent logics. That is, we may now have the resources to see how paraconsistency can *ascend* to metatheory. With this *paraconsistent ascent*, '[t]he classical ladder [can] be kicked away' (Priest 2008: 585). Instead of a ladder, however, we have a spiral staircase where, from a horizontal angle, we ascend upwards but, from a vertical angle, we may stay in the same 'spot'. So the same paraconsistent logic can ascend the spiral staircase without losing paraconsistency. With such a spiral staircase, a paraconsistent logician can now have a resource to develop a paraconsistent metalogic and a paraconsistent metatheory (a theory that accords a paraconsistent logic) to 'talk about' a paraconsistent logic.

The excitement expressed above may not be shared by all paraconsistent logicians. Traditionally (at least since the time of Tarski), the metalogic to study the properties of a logical system is considered to be classical. Priest (2006), for instance, accepts this classical 'default': 'provided we stay within the domain of the consistent, which classical reasoning of course does (by and large), classical logic is perfectly acceptable. ... [W]e are justified in assuming consistency until and unless it is shown otherwise' (p. 222).

This classical default has been a point of contention by some critics of paraconsistent logic. For instance, Burgess (2005) writes:

How far can a logician who *professes* to hold that [paraconsistency] is the correct criterion of a valid argument, but who freely accepts and offers standard mathematical proofs, in particular for theorems about [paraconsistent] logic itself, be regarded as *sincere* or *serious* in objecting to classical logic? (p. 740)²

One may think that there is nothing embarrassing about the appeal to classical metatheory in theorising about paraconsistent logic (and nonclassical logic generally). For instance, Beall (2009, 2011, 2013a, 2013b, 2013c, 2018) takes the formal apparatus (or models) to have only an instrumental value and argues that a classically closed theory (or a world) is useful. An 'instrumentalist' of this kind may be able to sidestep Burgess's criticism since there is no profession to holding the view that paraconsistent logic is 'correct'.³

However, there are paraconsistent logicians who are committed to a stronger view of logic according to which logic *must* be paraconsistent given that there *are* true contradictions. For example, Priest claims that 'logic is a normative subject' (Priest 1979: 297) and holds that a paraconsistent logic is somehow 'canonical'

²This criticism is actually about relevant logic; however, the same criticism can be made about paraconsistent logic.

³Thanks go to Jc Beall for his comments on an earlier draft of this paper that led us to see this point.

or ‘correct’. For him, it is paraconsistent logic that holds at the actual world. This means that Priest is committed to there being normative constraints for the metatheory used to study a paraconsistent logic as the metatheory must respect the canonical status of paraconsistent logic. Thus, Priest or someone like him comes face to face with the criticism of paraconsistent logic based on the lack of paraconsistent metatheory.

In this paper, we will argue that proving metatheoretical results of paraconsistent logics in a classical metatheory is unsatisfactory if a paraconsistent logician holds that logic is normative. In so doing, we are not going to conduct a full-scale attempt to paraconsistently prove completeness and other metatheoretical theorems of a paraconsistent logic.⁴ Instead, we will argue that the practice of relying on a classical metalogic to talk about a paraconsistent logic, which is meant to be correct, is *bad*. We will articulate the technical sense in which we describe such a practice as ‘bad’ in what follows. But, to be a little bit more precise, we will show that the world (to be understood metaphorically) described by the classical metatheory for a paraconsistent logic is a bad world—a world whose logic is ruled out by the metalogic.

The structure of the rest of the paper is as follows. First, we will describe a paraconsistent world: that is, a world where a paraconsistent logic holds. Second, we will explain what an impossible world is like. There are three ways to characterise the impossible nature of impossible worlds and we will explain what they are. Third, we will use that discussion of impossible worlds to describe a world of a classical metatheory for a paraconsistent logic as a *bad world*. Fourth, we will use such a description to argue against the standard practice of setting up a classical metatheory and metalogic for a paraconsistent logic. The paper will thus show that, if paraconsistent logic is meant to be somehow correct, it needs to ascend all the way up.

2 Paraconsistent Worlds

We define a *world* (or a representation of a world)⁵ to consist in a set of (non-logical) facts and a set of logical laws. We let p, q, r, \dots represent facts that hold at the world and $A_1, A_2, \dots \models B_1, B_2, \dots$ represent a logical law that holds at the world where p, q, r are propositions and A s and B s are propositional variables. We are mainly concerned with the laws of logic that hold *at* or *in* a world and nothing of what we say in this paper has any bearing on the laws of logic that hold *over* or *of* a world (i.e., the consequence relations that hold by a closure of

⁴For a flying start, see Weber 2021.

⁵We take no stance on the metaphysical nature of worlds in this paper and the metaphysical status of worlds is independent of the discussion of the paper.

the world under some logic).⁶ Then, a *paraconsistent* world is a world where a paraconsistent logic holds. That is, it is a world where no logical laws of the form $A, \neg A \models B$ hold. Thus, a paraconsistent world may be non-trivial even if it is contradictory.

From a classical perspective, a paraconsistent world may be thought to be impossible since classical logic does not hold there. However, there does not seem to be anything extra-ordinary about paraconsistent worlds. If a paraconsistent world is consistent (the facts that obtain at the world are consistent), it may behave just like a classical world depending on which paraconsistent logic holds at that world.⁷ But, if it is inconsistent (and not already trivial), it does not explode into triviality. So it does not go all haywire even if it is inconsistent.

While acknowledging this, we will argue that the metalogic used to metatheorise about paraconsistent worlds, *in the way that it is often described*, makes a paraconsistent world not an *impossible* world but a *bad* world, a ‘deformed nether-world’ (Girard & Weber 2015: 94). We will show that a paraconsistent world may be an impossible world; however, it should not be a bad world. In so arguing, we will object to the standard practice of introducing a classical metatheory for a paraconsistent logic.

3 Impossible Worlds

An *impossible* world is a world where impossible things happen. While this much is clear, what exactly such a world is like is a matter of dispute. From a paraconsistent perspective, an inconsistent world is not necessarily impossible even though such a world may count as impossible from a classical perspective. Given that it is a paraconsistent world that is at issue, we cannot rely on classical logic to characterise impossible worlds.⁸ We need to provide a definition of impossible worlds that is logic-neutral (i.e., does not rely on any particular logic) to define impossible worlds. There are generally three such definitions in the literature on impossible worlds:

1. An impossible world is a world where the laws of logic are *violated*. (Sandgren & Tanaka 2020, Tanaka 2018, Tanaka & Sandgren 202+.)
2. An impossible world is a world where the laws of logic are *different*. (Berto & Jago 2019, Priest 1992, 2008, 202+)

⁶For the distinction between the laws *in* (or at) a world and the laws *of* (or over) a world, see Routley 2019: 7.

⁷This is, in fact, debatable. It is just that a consistent paraconsistent world looks like a classical world from a classical perspective. Thanks go to Zach Weber for pointing this out.

⁸See Nolan 1997 and Zalta 1997 for the definitions of impossible worlds from classical points of view.

3. An impossible world is a world which is *open* (i.e., not closed under *any* logical entailment). (Berto & Jago 2019, Priest 2005, 202+)

In this paper, we do not debate which characterisation best captures impossible worlds. We will, instead, use the notion of violation to characterise impossible worlds. This is the notion required to introduce *bad worlds*, the kind of worlds required to argue against the standard practice of setting up a classical metatheory and classically proving metatheoretical results of paraconsistent logics.

So, what is a world where the laws of logic are violated? In order to describe such a world, consider the semantics for what we now call non-normal systems of modal logic introduced by C.I. Lewis, especially *S2* and *S3*. To capture these semantics, Kripke (1965) introduced *non-normal* worlds. What is characteristic of a non-normal system is the failure of necessitation ($\models A$ does not entail $\models \Box A$). Given a set of worlds, normally, if A holds at every world, then $\Box A$ must hold everywhere too. So, in a semantics for a non-normal system, there must be a world where $\Box A$ fails to hold even if A holds everywhere. Thus, at a non-normal world, $\Box A$ fails to hold for any A . (By the interchangeability of $\Box \neg$ and $\neg \Diamond$, $\Diamond A$ holds for any A at a non-normal world.) Then, even if $B \vee \neg B$ is true everywhere (and, thus, $\models B \vee \neg B$), $\Box(B \vee \neg B)$ fails to hold at a non-normal world. Thus, if there is such a world, it is the case that $\not\models \Box(B \vee \neg B)$ even if $\models B \vee \neg B$. Hence, non-normal worlds characterise the failure of necessitation and the non-normal systems.

These non-normal worlds have, since then, been redescribed as impossible worlds (Berto & Jago 2019, Priest 1992, 202+, Tanaka 2018). In a non-normal system, validity, \models , is defined in terms of truth preservation at all *normal* worlds.⁹ So $\models \Box(B \vee \neg B)$ since $B \vee \neg B$ holds at every world and, thus, $\Box(B \vee \neg B)$ holds everywhere. If we take a logical truth to express a logical law, then a non-normal world is a world where the logical laws are violated. A non-normal world is then an impossible world in the sense that the logical laws are *violated*.

To be more precise, let's say that when a set of logical laws holds at a world, there is a list of *entailment statements* that specify valid inferences according to those logical laws. An entailment statement often takes the form: $A_1, A_2, \dots \models B_1, B_2, \dots$. In order to focus on paraconsistent logic rather than relevant logic, we assume that the language for the entailment statements does not contain a conditional (or we assume that a conditional is defined in terms of a negation and a disjunction ($A \rightarrow B =_{\text{def}} \neg A \vee B$)).¹⁰ *Counterexamples* of $A_1, A_2, \dots \models B_1, B_2, \dots$ at a world w are p_1, p_2, \dots and q_1, q_2, \dots such that p_i is *true* for all i and q_i is *un-true* for all i at w when p_i and q_i are in the form of A_1, A_2, \dots and B_1, B_2, \dots respectively.

⁹At least, this is the case in Lewis' systems.

¹⁰To generalise this to cover (full) relevant logics (whose languages allow nested relevant conditionals), one can think of a law to be expressed by some relevant conditional. See Priest 202+.

Truth is a ‘good’ value and un-truth is a ‘bad’ value. What exactly the ‘good’ and ‘bad’ values amount to depends on the logic in question. But, roughly speaking, a ‘good’ value is the one that is preserved in a valid inference and a ‘bad’ value is the one that serves as a counterexample to the validity of inferences. If there are counterexamples of $A_1, A_2, \dots \models B_1, B_2, \dots$ at w , the inference $A_1, A_2, \dots \models B_1, B_2, \dots$ takes you from a good value to a bad value at w . A law of logic is *violated* at a world w iff w contains a counterexample of that law. Then, a world is *impossible* iff it contains a violation of some set of laws.

It is important to note that, under the definition of impossible worlds as worlds where the logical laws are violated, laws and facts may come apart at an impossible world. Take a world w such that $A \models B$ holds, but where p and q are true and un-true instances of A and B respectively. Since $A \models B$ holds at w , it should not take you from a good value to a bad value. But p and q provide a counterexample, making w an impossible world. This shows that there may be a mismatch between the laws and the facts that hold at an impossible world where the logical laws are violated.¹¹

4 Bad Worlds

Bad worlds were introduced to analyse the worlds that are used in the semantics for some relevant logics (Girard & Weber 2015). These worlds include Routley star worlds.¹² In order to accommodate the idea that both A and $\neg A$ may take good values (the values that are preserved in a valid inference), Routley & Routley (1972) introduced what is now known as Routley star worlds:

$\neg A$ is true at w iff A is un-true at w^* .

Given that A and $\neg A$ are evaluated at different worlds, they may both be true. With such a metatheoretical tool, completeness of various relevant logics, many of which are paraconsistent, has been proven with respect to the semantics that is set up in that metatheory (Brady 2018, Routley *et al.* 1982).

Even though they were introduced to play an important role in the semantics for relevant logics, the metalogic of star worlds is classical. For instance, here’s what Routley & Routley (1972) say about star worlds (with notational adaptation):

When w differs from w^* tautologies will often fail or contradictions hold in one or other of w and w^* : to illustrate consider $A \wedge \neg A$ in case

¹¹See Tanaka & Sandgren 202+.

¹²Ternary relations that are introduced to account for the semantics for relevant conditionals are an important part of the development of the semantics for relevant logics. However, given that our focus is on paraconsistency, we do not need to consider them here.

$w \neq w^*$. Then $A \wedge \neg A$ is in w if A is in w and A is not in w^* , a consistent assignment. (p. 338.)

So, when the Routleys first introduced what is now known as the Routley star, they were working within a classical metatheory which requires consistent assignments in the metatheory. Thus, Routley star worlds are worlds whose metalogic and metatheory are classical.

We will now show that such a classical metatheory makes a paraconsistent world not an impossible world but a *bad world*. Before going on to show this, we should note that bad worlds have been defined as the worlds that ‘could not have been, worlds in which the laws of logic are different, arbitrary, go “haywire”, fail’ (Girard & Weber 2015: 94). Defined in this way, bad worlds appear to be impossible worlds. However, as we will see in this section, there is an important difference between impossible worlds and bad worlds, at least when we characterise impossible worlds as worlds where the laws of logic are violated.

Recall that, at an impossible world (a logically violated world), laws and facts may come apart. Let $A \models B$ be an entailment statement that holds at w . This means that, at w , if A is true, B must also be true. If there is a counterexample to this logical law, there are instances of A and B , p and q , such that p is true but q un-true at w . Such a world is impossible as some of the facts that obtain at the world (i.e., p and q) violate a logical law. At an impossible world, thus, some facts are decoupled from some logical laws even though those logical laws remain laws that hold at that world.

The situation is rather different at a Routley star world. As the Routleys make it clear, if $w \neq w^*$, contradictions may hold at w^* (or w). However, there is no reason to think that the facts that obtain at the world serve as counterexamples to any of the logical laws. What happens at w^* instead is that the metalogic that governs the assignment of semantic values is classical. It maintains a consistent assignment of values at w^* not necessarily in relation to itself but in relation to w as the Routleys make it clear in the passage quoted above. So, consistency is placed as a meta-criterion that the Routley star worlds have to satisfy. This means that the Routley star worlds force the truth of $A \wedge \neg A$ at w to be expressed as: the truth of A at w and the un-truth of A at w^* . In the metatheory, thus, no contradiction arises. However, a paraconsistent logic is a logic that allows a contradiction to arise. So, the Routley star world can be thought of as metatheoretically resolving the contradiction that arises at the object-level. Hence, the metalogic introduced to theorise about a paraconsistent logic in the form of the Routley star worlds is in tension with the logic that it is designed to study.

We are now in a position to provide a definition of bad worlds. At a bad world, there is a mismatch between the logic that hold at the world and the logic of the metatheory. In other words, at a bad world, the logic of the metatheory fails to

match the logic that is (meta)theorised.¹³ Thus, a bad world is not an impossible world. Rather, it is a ‘deformed netherworld, to be kept separated from the proper worlds’ (Girard & Weber 2015: 94).

But why is it bad that there is a mismatch between the logic of a metatheory and the logic that has been (meta)theorised? We will answer this question in the next section.

5 Against Classical Paraconsistent Metatheory

If the Routley star captures the ‘correct’ negation operation in the presence of true contradictions, then a story must be told to make sense of it. But, when the story is told of the Routley star worlds, a consistent assignment is required such that a contradiction is resolved by the metalogic. Instead of allowing A and $\neg A$ to be both true at w , the Routley star analyses such a contradiction as A being true at w and it being un-true at w^* . However, if a contradiction arises at the actual world as many paraconsistent logicians such as Priest and Routley claim, the metatheory that analyses the logic thought to hold at the actual world (i.e., the correct logic) undermines the very logic that it analyses. Hence, the metalogic that is used to (meta)theorise about the logic rules out that very logic to hold at the world. There may not be anything wrong with classically modelling a paraconsistent logic and seeing what the logic looks like if one takes models to have only instrumental values.¹⁴ However, a metalogic ruling out the very logic about which it (meta)theorises is more than an embarrassment for those paraconsistent logicians such as Priest and Routley/Sylvan who hold that paraconsistent logic is somehow ‘correct’.¹⁵ For them, (meta)theorising about paraconsistent logics is a *bad* practice. Thus, if a paraconsistent logician wants to claim that it is a paraconsistent logic that holds at the actual world (meaning that a paraconsistent logic is ‘true’ in some respect), then they cannot rely on classical metalogic to (meta)theorise about that paraconsistent logic. ‘[I]f we need a paraconsistent logic *somewhere*, then we need it *everywhere*’ (Weber 2021: 84). If paraconsistency is somehow correct, it must ascend all the way up.

¹³Thanks go to Zach Weber for suggesting to put the point in this way.

¹⁴See Beall 2009, 2011, 2013a, 2013b, 2013c, 2018.

¹⁵For their views about paraconsistent logic, see, for instance, Priest 1987, 2006 and Routley 2019.

6 Conclusion

Anyone who has worked on non-classical metatheory knows how arduous recovering even the introductory chapters of standard logic textbooks can be.¹⁶ However, anyone who has worked in the philosophy of non-classical logics also knows how arduous it is to give a good classical story of non-classical logics because the constraints of classical logic forces a narrative that tries to make sense of what that story disagrees with.¹⁷ If we want to ride the wave of cultivating non-classical logics away from classical logic,¹⁸ we have to find a way no matter how impossible it seems.

We end our discussion with a speculative generalisation of our result. We have argued that classical metatheories for paraconsistent logics are *bad* because they rule out the very logic they are designed to study. If paraconsistent logic is claimed to be ‘correct’, it is a bad practice to (meta)theorise about it with classical metalogic. Given that a paraconsistent logic may be a sublogic of classical logic,¹⁹ the classical metalogic is stronger than the paraconsistent object-logic. So, the generalisation of our result is that when the metalogic is stronger than the logic being (meta)theorised, *bad* results follow. Should a paraconsistent logician investigate classical logic as an object of study with a paraconsistent metalogic, we do not expect similar results. Thus, it seems to be a reasonable generalisation. We do not have enough space to prove it here. So we leave the paper on this conjecture: a normativist about logic must use a metalogic that is no stronger than their object logic, or else they will produce bad outcomes.²⁰

References

- Bacon, A. 2013. ‘Non-classical Metatheory for Non-classical Logics’ *Journal of Philosophical Logic* 42: 335-355.
- Badia, G., Weber, Z. & Girard, P. 2022. ‘Paraconsistent Metatheory: New Proofs with Old Tools’, *Journal of Philosophical Logic*, Online First.
- Beall, Jc. 2009. *Spandrels of Truth*, New York: Oxford University Press.

¹⁶For instance, Weber 2021.

¹⁷For instance, see Girard & Tanaka 2016, although the awkwardness is hidden from the narrative.

¹⁸Bourget & Chalmers 2020.

¹⁹See Priest 2008.

²⁰We would like to thank Jc Beall and Zach Weber for their comments on earlier drafts of this paper. Many thanks also go to the anonymous referees for *Analysis* which gave us an opportunity for improvement. Koji Tanaka is supported by an Australian Research Council Future Fellowship (FT160100360).

- 2011. ‘Multiple-Conclusion LP and Default Classicality’, *The Review of Symbolic Logic* 4 (2): 326-336.
- 2012a. ‘LP⁺, K3⁺, FDE⁺, and their ‘Classical Collapse’’, *The Review of Symbolic Logic* 4 (2): 742-754.
- 2013b. ‘A Simple Approach Towards Recapturing Consistent Theories in Paraconsistent Settings’, *The Review of Symbolic Logic* 6 (4): 755-764.
- 2013c. ‘Shrieking Against Gluts: the Solution to the ‘Just True’ Problem’, *Analysis* 73 (3): 438-445.
- 2018. ‘The Simple Argument for Subclassical Logic’, *Philosophical Issues* 28: 30-54.
- Berto, F. and Jago, M. 2019. *Impossible Worlds*, Oxford: Oxford University Press.
- Bourget, D. & Chalmers, D. 2020. ‘The 2020 PhilPapers Survey’, <https://survey2020.philpeople.org/>.
- Brady, R. (ed.) 2003. *Relevant Logics and Their Rivals* (Volume II), Aldershot: Ashgate.
- Burgess, J. 2005. ‘No Requirement of Relevance’, *The Oxford Handbook of Philosophy of Mathematics and Logic*, S. Shapiro (ed.), New York: Oxford University Press, 727-750.
- Girard, P. & Tanaka, K. 2016. ‘Paraconsistent Dynamics’, *Synthese* 193 (1): 1-14.
- Girard, P. & Weber, Z. 2015. ‘Bad Worlds’, *Thought* 4: 93–101.
- Kripke, S. 1965. ‘Semantical Analysis of Modal Logic II. Non-Normal Modal Propositional Calculi’, *The Theory of Models*, J.W. Addison, L. Henkin, and A. Tarski (eds.), Amsterdam: North-Holland, 206-220.
- Nolan, D. 1997. ‘Impossible Worlds: A Modest Approach’, *Notre Dame Journal of Formal Logic* 38: 535-572.
- Priest, G. 1979. ‘Two Dogmas of Quineanism’, *Philosophical Quarterly* 29: 289-301.
- 1987. *In Contradiction*, Dordrecht: Martinus Nijhoff.
- 1992. ‘What is a Non-Normal World?’, *Logique et Analyse* 35: 291-302.
- 2005. *Towards Non-Being*, Oxford: Oxford University Press.
- 2006. *In Contradiction* (Second Edition), Oxford: Oxford University Press.
- 2008. *An Introduction to Non-Classical Logic* (Second Edition), Cambridge: Cambridge University Press.
- 202+. ‘Mission Impossible’, *Saul Kripke on Modal Logic*, Y. Weiss and R. Padró (eds.), New York: Springer.
- Rosenblatt, L. 2021. ‘Towards a Non-classical Meta-theory for Substructural Approaches to Paradox’, *Journal of Philosophical Logic* 50: 1007-1055.
- Routley, R. 2019. *Ultralogic as Universal?*, Z. Weber (ed.), Cham, Switzerland: Springer Nature.
- & Routley, V. 1972. ‘The Semantics of First Degree Entailment’, *Noûs* 6 (4): 335-359.
- Routley, R. et al. 1982. *Relevant Logics and Their Rivals*, Atascadero: Ridgeview.
- Sandgren, A. & Tanaka, K. 2020. ‘Two Kinds of Logical Impossibility’, *Noûs* 54 (4):

795-806.

Tanaka, K. 2018. 'Logically Impossible Worlds', *Australasian Journal of Logic* 15 (2): 489-497.

——— & Sandgren, A. 200+. 'What's So Impossible About Impossible Worlds?' (ms.).

Weber, Z. 2021. *Paradoxes and Inconsistent Mathematics*, Cambridge: Cambridge University Press.

——— Badia, G. & Girard, P. 2016. 'What is an Inconsistent Truth Table?'. *Australasian Journal of Philosophy* 94 (3): 533-548.

Zalta, E. 1997. 'A Classically-Based Theory of Impossible Worlds', *Notre Dame Journal of Formal Logic* 38: 640–660.