The (In)Consistency of Consistency

Hitoshi Omori, Graham Priest, and Zach Weber

Abstract

This paper considers approaches to the notion of consistency. In the logics of formal inconsistency (LFIs), consistency can be given as an operator, $\circ$, which is itself required to behave consistently. We give reasons to think that, even on the interpretation of the LFIs favored by Carnielli, consistency itself may be inconsistent. This motivates considering other ways of implementing the notion of *inconsistent* consistency—most notably, using non-deterministic semantics.

Dedication: It is with great pleasure that we dedicate this essay to Walter Carnielli, for his many important contributions to the study of paraconsistency.

Contents

1 Introduction

On some approaches to paraconsistency, such as the *logics of formal inconsistency* or LFIs, there is a *consistency operator*. Such a device allows a 'recapture' of classically valid but paraconsistently invalid inferences. This is made possible in part by assuming that the notion of consistency itself is consistent, perhaps in conjunction with the thought that no contradictions are true. On other approaches to paraconsistency, such as a dialetheic one, some contradictions *are* true and in particular there are reasons to think that consistency is not itself consistent. A question then arises as to how closely a dialetheic consistency operator can still approach a classical consistency operator, and achieve (assuming it is an achievement) classical recapture. In this paper, we investigate.

There are two appendices. The first applies the ideas in the body of the paper to a critique of a dialetheic solution of the paradoxes of self-reference—notably concerning 'revenge paradoxes'. The second provides the proofs of various results stated in the paper.

2 Consistent Consistency

Our investigations here may be traced back to Newton da Costa's pioneering work on paraconsistent logic.[1] The aim of this was to produce a system of logic for reasoning from inconsistent information in a non-trivial fashion, but which, in other ways, remained as close as possible to classical logic.

Clearly, that A is inconsistent may be expressed by $A \wedge \neg A$. Da Costa's idea was to express the consistency of A, naturally enough, as $\neg(A \wedge \neg A)$. He wrote this as A°. The consistency of A can then be expressed in the language itself. Appropriate constraints on negation force A° itself to behave consistently in a certain sense, specifically: $A \wedge \neg A \wedge A^\circ \vdash B$.[2] And this allows the language the resources to *force* the validity of classically valid inferences in a certain way. Specifically, let Θ be the set of all formulas of the form A°. Then if the inference $\Sigma \vdash A$ is classically valid, the inference $\Theta \cup \Sigma \vdash A$ is valid in the systems.[3]

Da Costa's approach has been considerably extended and generalised by Carnielli, Marcos, Coniglio, and others to the LFIs. In these, $\circ$, now written in prefix, is taken as a primitive connective whose properties suffice to make it satisfy the conditions:[4]

$$p \wedge \circ p \not\models q$$

$$\neg p \wedge \circ p \not\models q$$

$$p \wedge \neg p \wedge \circ p \models q$$

The first two clauses are there to rule out inappropriate candidates for $\circ p$, such as $\perp$ (where $\perp \models A$ for arbitrary A). Again, given that the other

[1] See for example [**?**].

[2] This comes out particularly clearly in the non-truth-functional semantics for da Costa's systems. See [**?**].

[3] For further discussion, see [**?**, §4.4].

[4] See [**?**, p.20]. More generally, consistency can be enforced on a whole set of formulas, but we may ignore this generalisation here.

machinery in the language behaves classically in consistent situations, the presence of $\circ$ gives the language the resources to force the validity of classically valid inferences, as before.

There are many other nuances to the LFIs, but for our purposes the key idea is just this: in these logics, the consistency of a statement in a certain sense is expressible in the language, and this allows for a simple way to obtain 'classical recapture', that is, apparently ensuring that one can reason classically. Now, in non-classical logics the whole point, one might think, is *not* to reason entirely classically; but there is nevertheless a certain virtue in recovering classical logic in a controlled way. Classical reasoning may well be held to be correct in consistent contexts, and the machinations of $\circ$ would seem to provide the resources to account for how this is possible: enthymematically.

3 Problems with Consistent Consistency

In this section we turn a critical eye on the consistency of consistency.

3.1 Consistency in LFIs

The account we have just sketched is paraconsistent: it allows the possibility of inconsistency via an operator $\circ$ for separating out the well-behaved or 'consistent' sentences from sentences that may not so behave. The logic LFI1, introduced in [**?**], can be thought of as LP (or *logic of paradox* [**?**]) plus a primitive connective, $\circ$, with the truth table

A	$\circ A$
t	**t**
b	**f**
f	**t**

With $\neg$ as LP negation, one can then define a second operator, $\bullet$, which is a dual 'non-classicality' or *inconsistency* marker, $\bullet A := \neg\circ A$.[5]

Now, since $\circ$ and $\bullet$ only take classical values, then in LFI1 we have

[5]Indeed, adding $\circ$ to the language is very powerful: one may further define classical negation as $\sim A := \neg A \wedge \circ A$, a bottom constant $\perp := A \wedge \sim A$, and detachable conditional $A \to B := \sim A \vee B$.

$$\circ A, \bullet A \models B \qquad [Cons]$$

$$\models \circ A \vee \bullet A \qquad [Comp]$$

$$\circ A, A, \neg A \models B \qquad [RECQ]$$

Crucially, the $\circ$ and $\bullet$ operators are mutually exclusive by [Cons], and *cannot themselves be subject to non-classicality.* The basic rule [RECQ] (restricted ECQ), expresses that $\circ$ is 'classical'. Moreover, $\circ A, \neg\circ A \models B$ according to [Cons], and more simply, $\bullet\circ A$ explodes.[6]

In recent versions of this approach, the notion of consistency is generalized to that of 'classicality'. Rodrigues *et al* [**?**] use as a base logic FDE. In addition to the inferences above, we have the rule:

$$\circ A \models A \vee \neg A \qquad [LEM]$$

From all these on an FDE base, it follows that contradiction implies non-classicality:[7]

$$A, \neg A \models \bullet A$$

This shows how da Costa's original suggestion has been expanded: the operators $\circ$ and $\bullet$ are generalizations, since $A \vee \neg A \not\models \circ A$ and $\bullet A \not\models A \wedge \neg A$ [**?**, §3]. This approach presumes, in a slogan, the classicality of classicality, of which the consistency of consistency is a special case.

It is not unusual, of course, to make such commitments. Nevertheless, we will look at reasons one might *not* assume that consistency is consistent. First we look at some reasons for the inconsistency of consistency coming from Carnielli *et al*'s own philosophical position. Then we look at some further, independent reasons.

3.2 Inconsistent Information

The initial reason one might say that consistency *is* consistent is that one might presume *everything* is, at the end of the day, consistent. And indeed, in

[6] Further constraints, such as iterability $\circ A \models \circ\circ A$, or even more strongly, the rule $\models \circ\circ A$, may be plausible too: if one thinks A is well-behaved, then it would seem natural to think also that A's well-behavior is also well-behaved. We return to this in §4.4.

[7] *Proof*: Suppose $A, \neg A$. Either $\circ A$ or $\bullet A$, by [Comp]. But if $\circ A$, then on assumption, $\bullet A$, by [RECQ]. A 'law of excluded fourth' follows, $\models A \vee \neg A \vee \bullet A$, from [Comp] and [LEM].

several publications, Carnielli and collaborators have presented a stridently *anti-dialetheic* interpretation of the inconsistencies that appear in the LFIs.[8] An inconsistent A is such that $A \wedge \neg A$; dialetheism says that some of these sentences are true. Carnielli *et al*'s view is that dialetheism is committed to the existence of true contradictions in a metaphysical or ontological way. They urge instead to understand contradictions in an epistemic way: for $A, \neg A$ to hold under this interpretation is for there to be *evidence* for both.

> What property are we going to ascribe to a pair of accepted contradictory propositions such that it would be possible for a proposition to enjoy it without being true? Such a property has to be something weaker than truth. ... Our proposal is that the notion of evidence is well suited to be such an answer [**?**, p.3791].

This interpretation of 'inconsistency' is in the spirit of Belnap [**?**] and Dunn [**?**], taking these logics not to be about truth-preservation, but rather about the deflated notion of *information*.

For the record, dialetheism claims that there are true contradictions, but it does not say what 'truth' means. If one takes a naively realist correspondence view of truth, then there may be ontological or metaphysical contradictions that 'exist' in some sense; but one could also take an anti-realist view of various sorts, or indeed any other candidate view on truth. If 'truth' means the existence of some kind of mental construction, then dialetheism says there are inconsistent mental constructions. And so forth. The point is that all dialetheism requires is *truth*.[9]

Requiring truth, metaphysical or otherwise, is stringent. Truth sets a very high bar for taking a contradiction seriously. A dialetheia needs to be either proven as a theorem in a sound system, or in some other way established by rigorous truth-assuring means. On most understandings of truth, finding true contradictions is *difficult*. On the other hand, if we only read $A \wedge \neg A$ as something like 'there is (good) evidence for A, and (good) evidence for $\neg A$' then it is much *easier* to find such contradictory pairs. Indeed, Carnielli *et al* do not take 'real' contradictions to be of any concern. Their inconsistencies are in our minds, or language, or reasoning. As Carnielli and Rodrigues put it:

[8] See e.g. [**?**, **?**, **?**].

[9] See, further, [**?**, ch.2] and [**?**, §27.6]

> In fact, 'real contradictions' seem to be quite impossible.... If it were confirmed someday that real contradictions do exist ... [i]n such an improbable scenario, a considerable part of science, and also of philosophy, would collapse altogether [**?**, p.3812].

The statement shows that the target inconsistencies are significantly downgraded from a dialetheic notion and are, in a sense, very classical.[10]

Our point here is that Carnielli has adjusted down the notion of contradiction so that it is much *harder* to maintain that some A is *not* contradictory. In particular, if there is some *evidence* or reason to believe that consistency is not itself consistent, then by Carnielli's own lights that is sufficient to relax the assumption of consistency. Is there such evidence?

Minimally, we all know that Gödel proved that there will never be an absolute proof of consistency. That in itself is already a step down from certitude. It leaves open the space for incidents like in 2010, when Fields medalist Vladimir Voevodsky, professor of mathematics at Princeton University, presented a lecture to the Institute for Advanced Study called "What if Current Foundations of Mathematics Are Inconsistent?".[11] Or in 2011, Edward Nelson, another mathematics professor at Princeton, publicized claims to have proven the inconsistency of Peano Arithmetic. For some period, then, there was reason to think that the statement 'PA is consistent' is both consistent (everyone thinks so!) and also not (at least two qualified experts say maybe not).

To take a more mundane example, if we are thinking with Belnap of a computer database collecting statements, it is very easy to believe that at least one such statement *about* the $\circ$ operator will receive conflicting information. Consider a lightly imaginary example[12] of a tech company running a routine R to check their databases for inconsistencies. A unit in the company's cybersecurity team then issues a warning that R has been infiltrated

[10] The statement is also simply a piece of unsupported dogma. See [**?**, ch.9].

[11] See https://www.ias.edu/ideas/2012/voevodsky-foundations-of-mathematics.

[12] For a real example, the ride-share company Uber uses a platform called Uber Money to maintain data consistency between asynchronous platforms (https://www.uber.com/en-DE/blog/money-scale-strong-data/). On September 15, 2022, Uber discovered its network had been breached, when a message on it internal system told employees, "I announce I am a hacker and Uber has suffered a data breach" (https://www.nytimes.com/2022/09/15/technology/uber-hacking-breach.html). According to a security engineer, "They pretty much have full access to Uber. This is a total compromise, from what it looks like." The data consistency platform would then itself have been compromised.

by hackers. A second cybersecurity unit says that R is secure; but perhaps this unit itself has been compromised. What should the head of cybersecurity think about R when it issues statements like 'data fragment X is consistent'? There is reliable and conflicting—contradictory—information about consistency itself. Such examples are common. Information is cheap.

3.3 True Inconsistency (of Consistency)

Looking further afield at reasons to doubt the consistency of consistency, if we have a suitably expressive language, then it is easy to find a liar-like sentence, L, of the form:

$$L : \neg L \wedge \circ L$$

i.e. 'this sentence is false, and consistently so', or some appropriate modification depending on the LFI in question. Reasoning in a standard fashion, one can show that $\circ L$ and $\neg \circ L$. LFIs are not designed to handle such sentences. This means they join classical logic in being fundamentally incomplete, in the specific sense of not being able to fully account for themselves. The LFIs model reasoning about possibly inconsistent information, but they cannot model reasoning about the possibility of their *own* inconsistency.

Another way to see this is to consider how a set theory based on an LFI (as in Carnielli and Coniglio 2016 [**?**, ch.8]) has the same limitations as standard theories like ZF, where there famously cannot be a set of all sets.[13] Most crucially, while a predicate $C(x)$ on sets meaning 'x is consistent' is definable, the set of all such sets $\mathsf{cons} = \{x : C(x)\}$ cannot exist on pain of, well, inconsistency. So the classical universe the LFIs purport to pick out is either unspeakable within the LFIs, or else consistency is not (everywhere and always) consistent after all. We suggest that this dilemma, while all too familiar, is a significant *cost* for the consistent approach.

The existence of liar-like sentences and the inconsistency of a domain of universal quantification are simply handled from a dialetheic perspective: they are just dialetheias. Some contradictions are true. And when it comes to the consistency of consistency for the dialetheist, we get a rather different result from the LFI approach. A consistency operator would pick out all the non-dialetheias. But it is well known that there is an overlap in the categories of being and not being a dialetheia [**?**, p.294], which *a fortiori* means the set

[13] See [**?**] for a different approach, inspired by [**?**].

of consistent and inconsistent sentences overlap. So any operator picking out the consistent sentences will itself need to be inconsistent.

If there is reason to doubt consistency at all, there is, then, reason to doubt the consistency of consistency.[14]

4 Inconsistent Consistency

4.1 A Simple Approach

How then should one express consistency if it may itself be inconsistent?

There is, in fact, a simple way for the dialetheist to express consistency. This is with the expression 'A is not both true and false'. One can express this with a truth predicate as $\neg(T\langle A\rangle \wedge T\langle\neg A\rangle)$, or even more simply as $\neg(A \wedge \neg A)$.[15] Let us fix on the latter. This is, indeed, a theorem of LP: all contradictions are at least false. Now, it is known that for dialetheists, there are counterexamples to this theorem, too—but, insofar as $\neg$ expresses 'not' for the dialetheist, then this seems like the natural way to express 'A is not inconsistent'. It may be objected that this way of expressing consistency is not consistent; it may be replied that the dialetheic approach is, as advertised, not consistent, and that consistency cannot be enforced by the dialetheist or anyone else (see §5.4). There are further questions, e.g. about expressing non-triviality or unsatisfiability; but those are further questions, not about consistency *per se*. On this rather bracingly simple approach, in a slogan, 'just true' is just 'true'.[16]

A cost of this approach is that it does not deliver anything like a consistency operator as thought to be understood by non-dialetheists. Saying that A is consistent in the proposed way does not rule out that A is also inconsistent; indeed, every proposition is consistent in the proposed way, so it has the air of vacuity. At the very least, it is at odds with seemingly mainstream

[14] In the rest of this paper, we will take this as an opening to explore a dialetheic approach to consistency. It should be added though that a paraconsistent *non*-dialetheist such as Carnielli can accept our arguments to the effect that consistency may be inconsistent, but still do so within his own preferred epistemic interpretation, without accepting true contradictions. The debate about dialetheism is related to but can be separated from the question of the behavior of the consistency operator. Thanks to a referee here.

[15] Even with the T-scheme, these two definitions do not come to the same thing if truth does not commute with negation; cf. [**?**, p.70], [**?**, p.288].

[16] Due to Beall. See [**?**]. This slogan *does* assume that truth commutes with negation.

usage. While this might be a cost that is worth paying eventually (see [**?**, ch.10] for payment), in the rest of this section we want to consider a more widely-acceptable option.

The strategy we will consider is to take the system LFI1, where the consistency operator is a primitive connective, and systematically weaken it, to see to what extent we can obtain a dialetheically tenable notion of consistency. To do this, we will deploy the framework of non-deterministic semantics, which allows more flexibility in assignments of truth values.[17]

4.2 Non-Deterministic Semantics for Consistency

The general definition of non-deterministic semantics goes as follows.

Definition 1 A *non-deterministic matrix* (*Nmatrix* for short) for a language, $\mathcal{L}$, is a tuple $M = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$, where:

(a) $\mathcal{V}$ is a non-empty set of truth values.

(b) $\mathcal{D}$ is a non-empty proper subset of $\mathcal{V}$.

(c) For every n-ary connective $*$ of $\mathcal{L}$, $\mathcal{O}$ includes a corresponding n-ary function $\tilde{*}$ from $\mathcal{V}^n$ to $2^{\mathcal{V}} \setminus \{\emptyset\}$.

We say that M is (in)finite if $\mathcal{V}$ is so. A *legal valuation* in an Nmatrix M is a function $v : \mathsf{Form} \to \mathcal{V}$ that satisfies the following condition for every n-ary connective $*$ of $\mathcal{L}$ and $A_1, \ldots, A_n \in \mathsf{Form}$:

$$v(*(A_1, \ldots, A_n)) \in \tilde{*}(v(A_1), \ldots, v(A_n)).$$

Finally, the semantic consequence relation can be defined by requiring the preservation of designated values for all legal valuations.

Non-deterministic semantics generalize the standard many-valued semantics in the sense that if $\tilde{*}$ is always mapped to singletons, then we obtain the usual many-valued semantics.

In what follows, the only connective that will enjoy non-determinacy is the consistency operator. What is the minimal condition for the consistency operator? Following the LFIs, we take the following condition as the minimum condition:

[17] For non-deterministic semantics, see [**?**, **?**].

C1 $p \wedge \neg p \wedge \circ p$ is not satisfiable.

From here, let us explore which further 'consistency' conditions can be adopted coherently from a dialetheic perspective.

4.3 LP with Maximally Non-Deterministic Consistency Operator

If we implement this idea on top of LP with the help of non-deterministic semantics, we obtain the following non-deterministic semantics.

Definition 2 Let LPmnC (LP with maximally non-deterministic consistency operator) be the expansion of LP by $\circ$ with the following non-deterministic semantics.

x	$\tilde{\circ}x$
$\mathbf{t}$	$\{\mathbf{t}, \mathbf{b}, \mathbf{f}\}$
$\mathbf{b}$	$\{\mathbf{f}\}$
$\mathbf{f}$	$\{\mathbf{t}, \mathbf{b}, \mathbf{f}\}$

We refer to the semantic consequence relation as $\models_{LPmnC}$.

If we consider the extensions of LPmnC obtained by possibly reducing or eliminating non-determinacies, then a simple combinatorial argument reveals that there are 49 extensions of LPmnC, one being LFI1. Needless to say, there are again a number of different ways to group 49 consistency operators! Given that we also have dialetheism in our scope, we will divide by the definability of classical negation (which is a problematic notion for dialetheism when applied to a naive theory of truth or sets). By *classical negation*, which we write as $\neg_b$, we mean a connective that satisfies the following condition.

Classical negation $\tilde{\neg}_b x \subseteq \mathcal{D}$ iff $x \notin \mathcal{D}$

There are 16 extensions of LPmnC in which classical negation is *not* definable.

Proposition 1 Consider the 16 extensions of LPmnC, obtained by eliminating non-determinacy in the following manner:

- For the value of $\tilde{\circ}\mathbf{t}$, one of the following holds: $\tilde{\circ}\mathbf{t} = \{\mathbf{t}, \mathbf{b}, \mathbf{f}\}$, $\tilde{\circ}\mathbf{t} = \{\mathbf{t}, \mathbf{f}\}$, $\tilde{\circ}\mathbf{t} = \{\mathbf{b}, \mathbf{f}\}$, or $\tilde{\circ}\mathbf{t} = \{\mathbf{f}\}$,

- For the value of $\tilde{\circ}\mathbf{f}$, one of the following holds: $\tilde{\circ}\mathbf{f} = \{\mathbf{t}, \mathbf{b}, \mathbf{f}\}$, $\tilde{\circ}\mathbf{f} = \{\mathbf{t}, \mathbf{f}\}$, $\tilde{\circ}\mathbf{f} = \{\mathbf{b}, \mathbf{f}\}$, or $\tilde{\circ}\mathbf{f} = \{\mathbf{f}\}$.

Then classical negation is not definable in any of these extensions.

Proof. By considering the extension in which the consistency operator is interpreted as the bottom constant. □

On the other hand, for the other 33 extensions of LPmnC, classical negation is definable. Indeed, if $\tilde{\circ}\mathbf{t} = \{\mathbf{t}, \mathbf{b}\}$, then classical negation can be defined as $\neg A \wedge \circ\neg A$ which has the following non-deterministic matrix:

x	$\tilde{\neg}_b x$
$\mathbf{t}$	$\{\mathbf{f}\}$
$\mathbf{b}$	$\{\mathbf{f}\}$
$\mathbf{f}$	$\{\mathbf{t}, \mathbf{b}\}$

The same classical negation is definable when $\tilde{\circ}\mathbf{f} = \{\mathbf{t}, \mathbf{b}\}$. In this case, classical negation can be defined as $\neg A \wedge \circ A$.

Similarly, if $\tilde{\circ}\mathbf{t} = \{\mathbf{t}\}$ or $\tilde{\circ}\mathbf{f} = \{\mathbf{t}\}$, then classical negation can be defined as $\neg A \wedge \circ\neg A$ or $\neg A \wedge \circ A$, respectively, and classical negation has the following deterministic matrix:

x	$\tilde{\neg}_b x$
$\mathbf{t}$	$\{\mathbf{f}\}$
$\mathbf{b}$	$\{\mathbf{f}\}$
$\mathbf{f}$	$\{\mathbf{t}\}$

Finally, if $\tilde{\circ}\mathbf{t} = \{\mathbf{b}\}$ or $\tilde{\circ}\mathbf{f} = \{\mathbf{b}\}$, then classical negation can be defined as $\neg A \wedge \circ\neg A$ or $\neg A \wedge \circ A$, respectively, and the classical negation has the following deterministic matrix:

x	$\tilde{\neg}_b x$
$\mathbf{t}$	$\{\mathbf{f}\}$
$\mathbf{b}$	$\{\mathbf{f}\}$
$\mathbf{f}$	$\{\mathbf{b}\}$

Note that, from the two-valued relational semantic perspective, non-deterministic classical negation is obtained by specifying only the truth condition, without any falsity conditions. Moreover, the deterministic classical negations are obtained by requiring that $\neg_b A$ is false iff A is true, and $\neg_b A$ is always false, respectively.[18]

[18] See [?, ?] for some considerations of classical negation in the context of FDE family.

4.4 Consistency Conditions as Rules

For the purpose of discussing the properties of the consistency operators, it will be helpful to consider how these semantic conditions relate to rules or proof, via some correspondence results. Let us first recall the natural deduction presentation of LP.

Definition 3 A natural deduction system for LP can be presented as follows (cf. [**?**, p.309]).

$$\frac{A \quad B}{A\wedge B} \qquad \frac{A\wedge B}{A} \qquad \frac{A\wedge B}{B} \qquad \frac{A}{A\vee B} \qquad \frac{B}{A\vee B} \qquad \frac{A\vee B \quad \begin{array}{c}[A]\\ \vdots \\ C\end{array} \quad \begin{array}{c}[B]\\ \vdots \\ C\end{array}}{C}$$

$$\frac{\neg\neg A}{\overline{A}} \qquad \frac{\neg(A\vee B)}{\overline{\neg A\wedge\neg B}} \qquad \frac{\neg(A\wedge B)}{\overline{\neg A\vee\neg B}} \qquad \frac{}{A \vee \neg A}$$

We write $\Gamma \vdash B$ if and only if there is a derivation of B from some finite subset $\{A_0, ..., A_n\} \subseteq \Gamma$ using the above rules.

We can then introduce an expansion of LP by the consistency operator.

Definition 4 A natural deduction system for LPmnC can be obtained by adding the following rule to those for LP:

$$\frac{A \quad \neg A \quad \circ A}{B} \qquad \text{RECQ}$$

We also define derivability as above, and denote it as $\vdash_{LPmnC}$. As expected, we obtain the following soundness and completeness result.

Proposition 2 For all $\Gamma \cup \{A\} \subseteq \mathsf{Form}$,

$$\Gamma \vdash_{LPmnC} A \text{ iff } \Gamma \models_{LPmnC} A.$$

Proof. See Appendix. □

Now we can state the following correspondences, showing how various truth conditions play out.

Proposition 3 First, for the correspondence between the elimination of non-determinacy for $\tilde{\circ}\mathbf{t}$, we obtain the following result.

Elimination of non-determinacy	Additional rule
$\tilde{\circ}\mathbf{t} = \{\mathbf{t}, \mathbf{b}, \mathbf{f}\}$	None
$\tilde{\circ}\mathbf{t} = \{\mathbf{t}, \mathbf{f}\}$	$\dfrac{\circ A \quad \neg\circ A}{\neg A}$
$\tilde{\circ}\mathbf{t} = \{\mathbf{b}, \mathbf{f}\}$	$\dfrac{}{\neg A \vee \neg\circ A}$
$\tilde{\circ}\mathbf{t} = \{\mathbf{f}\}$	$\dfrac{\circ A}{\neg A}$

Second, for the correspondence between the elimination of non-determinacy for $\tilde{\circ}\mathbf{f}$, we obtain the following result.

Elimination of non-determinacy	Additional rule
$\tilde{\circ}\mathbf{f} = \{\mathbf{t}, \mathbf{b}, \mathbf{f}\}$	None
$\tilde{\circ}\mathbf{f} = \{\mathbf{t}, \mathbf{f}\}$	$\dfrac{\circ A \quad \neg\circ A}{A}$
$\tilde{\circ}\mathbf{f} = \{\mathbf{b}, \mathbf{f}\}$	$\dfrac{}{A \vee \neg\circ A}$
$\tilde{\circ}\mathbf{f} = \{\mathbf{f}\}$	$\dfrac{\circ A}{A}$

Proof. See Appendix. □

Let us now note facts about the inconsistency of consistency. To this end, we consider the following three different ways of expressing the consistency of consistency.

C2 $\models \circ\circ p$

C3 $\circ p \models \circ\circ p$

C4 $\circ p \wedge \neg\circ p$ is not satisfiable.

An *inconsistent* consistency operator is one for which at least one of **C2-C4** fails. We obtain the following results.

Proposition 4 All 16 of the consistency operators in Proposition 1 are inconsistent consistency operators, in the sense that $\not\models \circ\circ p$.

Proof. None of the 16 extensions satisfies **C2**, because we can assign $\mathbf{f}$ to $\circ\mathbf{f}$. $\square$

Propositions 5 If $\tilde{\circ}\mathbf{t} = \{\mathbf{f}\}$ and $\tilde{\circ}\mathbf{f} = \{\mathbf{f}\}$, or $\tilde{\circ}\mathbf{t} = \{\mathbf{t}, \mathbf{f}\}$ and $\tilde{\circ}\mathbf{f} = \{\mathbf{f}\}$, then **C3** holds. Otherwise, **C3** fails, $\circ p \not\models \circ\circ p$.

Proof. For the former, this is tedious but straightforward. For the latter, we assign the value $\mathbf{f}$ to p. If we can pick a valuation that assigns $\mathbf{b}$ to $\circ\mathbf{f}$, then we are done. Otherwise, we assign $\mathbf{t}$ to $\circ\mathbf{f}$, but we can then assign the value $\mathbf{f}$ to $\circ\mathbf{t}$. $\square$

Proposition 6 Consider the 4 extensions of LPmnC, obtained by reducing or eliminating non-determinacy in the following manner.

- For the value of $\tilde{\circ}\mathbf{t}$, one of the following holds: $\tilde{\circ}\mathbf{t} = \{\mathbf{t}, \mathbf{f}\}$ or $\tilde{\circ}\mathbf{t} = \{\mathbf{f}\}$,
- For the value of $\tilde{\circ}\mathbf{f}$, one of the following holds: $\tilde{\circ}\mathbf{f} = \{\mathbf{t}, \mathbf{f}\}$ or $\tilde{\circ}\mathbf{f} = \{\mathbf{f}\}$.

Then, in all of these extensions, **C4** holds. For the others, **C4** fails.

Proof. For the former, observe that all statements with consistency operator will receive one of the two classical values $\mathbf{t}$ or $\mathbf{f}$. For the latter, we can assign the value $\mathbf{b}$ as a value for statements with consistency operator. $\square$

Finally, one may consider the condition **C1** being too weak, and that the following two conditions should also be both satisfied.

C5 $p \land \circ p$ is satisfiable.

C6 $\neg p \land \circ p$ is satisfiable.

The effect of this requirement is that the option to interpret the consistency operator as the bottom constant gets ruled out, but the other 15 options of Proposition 1 will remain as options, even after demanding the above two conditions.

5 Classical Recapture

In the light of the above, let us return to the matter of classical recapture.

If one subscribes to a paraconsistent logic, there are various principles of classical logic that one must foreswear as invalid—notably, disjunctive syllogism (DS), $A, \neg A \vee B \vdash B$, since, modulo a few very natural assumptions, this delivers explosion. But there seem to be many cases where reasoning with DS seems to be perfectly fine. Susan knows that the capital of Nicaragua is either San José or Managua. Penelope tells her reliably that San José is the capital of Costa Rica, and so not Nicaragua; so Susan infers correctly that it is Managua. There must, it seems, be some explanation of why this seems legitimate.

Several well known answers have been proposed and discussed. Central to much of the discussion is thought that the counter-examples to classical principles—at least, those of concern here—all require an inconsistent situation. And if an inconsistent situation is not on the table, so the thought goes, then classical reasoning is fine.[19]

A satisfactory articulation of this thought is not straightforward, however. In particular, one cannot simply add a premise (schema) to the effect that things are consistent, as one can add all instances of Excluded Middle to intuitionist logic to get classical logic. For, as has frequently been observed, if a situation is inconsistent, one cannot make it consistent by adding a premise to the effect that it is consistent. That will simply make it *more* inconsistent.[20]

5.1 LFIs

A Carnielli-style LFI can be used to solve the problem, as we observed, in effect, in §2. We add a premise (or rather, a premise scheme) to the effect that every A is consistent, $\circ A$. This does not force the situation to be consistent. Rather, what it does, is to force the situation, if it is inconsistent, to be trivial.

[19] As Priest puts it: "It is true that many inferences that are classically acceptable must be acknowledged as invalid. This does not occasion an explanatory loss, however. For the only situations about which it makes sense to reason classically are consistent ones; and even paraconsistent logicians may employ classical logic in consistent situations (just as intuitionists may employ classical logic when reasoning about finite situations): it is a special case" [**?**, p. 128].

[20] As noted by Belnap and Dunn. See [**?**, 8.2] and [**?**, p.88].

But whether it is consistent or trivial, an instance of DS is truth-preserving.

The problem with this strategy, as noted in §3.3, is that even if it works for many applications of dialetheic paraconsistency, it completely trashes such a solution to the paradoxes of self-reference—which is, perhaps, the most popular application. For using a Carnielli-style consistency operator, we can define a Boolean negation which then delivers a strong liar—and triviality.[21]

The LFIs do not make claim to solving the paradoxes of self-reference, or indeed reckoning with true inconsistency at all. Nevertheless if one is already involved with paraconsistency, minimally, it would seem desirable to have an approach that can handle Liars and their like. How else may we proceed, which does not have this consequence?

5.2 $\to \perp$

One well known way of recapturing classical logic is to use the *falsum* constant, $\perp$, and a detachable conditional, $\to$. $\perp$ is a logical constant satisfying the condition that $\perp \vdash B$.[22] Let us write:

$$A^{\perp} := (A \wedge \neg A) \to \perp$$

Then $A, \neg A, A^{\perp} \vdash B$. So, given $A^{\perp}$, if we have A and $\neg A$, the situation collapses into triviality, and classical recapture can be obtained as with $\circ A$ in an LFI.[23]

We can use $A^{\perp}$ to define a negation-like operator, $-A$, as $\neg A \wedge A^{\perp}$. Presuming that $\to$ satisfies *modus ponens* then $A, -A \vdash B$; but unless $\to$ has some very unpalatable properties it will not be the case that $\vdash A \vee -A$. This being so, $-A$ is not Boolean negation; and as is well known, a theory of naive truth or sets using an appropriate relevant conditional can contain $\perp$ without triviality.[24]

So using $\perp$ gives us a way of recapturing classical reasoning enthymematically. However, it is implausible to take $A^{\perp}$ as expressing the consistency of A, since the mere truth of something false only should not lead to all hell

[21] Note that there is also a way to avoid triviality. See [**?**].

[22] See, e.g., [**?**, §7.6].

[23] As in §2, if Θ is the set of formulas of the form $B^{\perp}$, then if the inference from Σ to A is classically valid, then $\Theta \cup \Sigma \vdash A$. In *LP*, if the inference from Σ to A is classically valid, then for some B, $\Sigma \vdash A \vee (B \wedge \neg B)$. Given that $B^{\perp}$, A follows. See [**?**, 8.5].

[24] See [**?**, ch.4].

breaking lose. 'Donald Trump is a frog and not a frog', is false and consistently so. But from a paraconsistent perspective 'if Donald Trump is a frog and not a frog then he will become US president in 2024' has nothing to recommend it. Similarly, $A \to \perp$ does not express the thought that A is false only. 'Donald Trump is a frog' is just plain false. But the conditional 'if Donald Trump is a frog then the Continuum Hypothesis holds' is a fallacy of relevance.[25]

5.3 Default Reasoning

Another strategy for recapturing classical logic is to employ default reasoning.[26] The idea here is simply to apply classical reasoning until it becomes patent that it does not work. In particular, we use the disjunctive syllogism, $A, \neg A \vee B \vdash B$ until a proof of $A \wedge \neg A$ turns up. We then take back B. If we are working in a consistent theory, so that a contradiction never turns up, the result of all classical reasoning stands.[27]

The relevant question for present purposes concerns the rationale for this procedure. Like all default reasoning, the assumption is that it is licit to reason assuming that we are in a normal situation unless and until we are shown otherwise. If we add to this the thought that consistency is normal, and that in consistent situations classical inferences are truth-preserving, then default classical reasoning seems justified.[28] The claim we need to focus on here is that in consistent situations classical reasoning is truth preserving.

We may concentrate on the DS, since this delivers classical logic when added to LP. So consider the conditional:

- If A is consistent then: $A, \neg A \vee B \vdash B$.

Never mind exactly how we understand 'A is consistent', provided only that it may itself be inconsistent, we may have all of: A is consistent, A, and $\neg A$.

[25] A more technical reason is given by Murzi and Scambler [**?**, §3.6]. The Curry paradox gives us a sentence, K equivalent to $K \to \perp$. This cannot be true on pain of triviality, so it is false only. But then this cannot be expressed by $K \to \perp$.

[26] See, e.g., [**?**, §§4.9-4.12], and [**?**, ch.8].

[27] The articulation of this idea has produced much important technical work in adaptive logic, especially by Diderik Batens and his school. See, for a start, [**?**].

[28] The assumption of the normality of consistency is a substantial one, but let us pass over this for now. See [**?**, ch.8]; but also [**?**, ch.3]. For concerns about the classical default approach for dialetheism, see [**?**, ch.3].

This being the case, we can infer an arbitrary B. So much the worse for the conditional.

However, we can reason this way only if the conditional is detachable. To avoid this, one might take it to be a (non-detachable) material conditional.[29] In this case, inferring that $A, \neg A \vee B \vdash B$ from 'A is consistent' is an instance of the DS, and so is itself a default inference, acceptable till 'A is inconsistent', that is, A and $\neg A$, are established. The unfortunate consequence is then avoided.[30]

Now, proceeding in this way clearly attempts to justify a certain kind of default reasoning by using that very default reasoning. Is this problematic? It is a kind of bootstrapping; but bootstrapping is unavoidable in logic. In any soundness proof in metalogic, to show that an inference is sound, one needs to use that very inference. Perhaps the present case of bootstrapping is no worse than that. If so, we have a solution to the problem of classical recapture. It depends, however, on an ineliminable use of default reasoning.

Is there a way which avoids this?

5.4 Non-Determinacy

The material of §4 on a non-determinate consistency operator provides such a way. As explained, we can have a consistency operator which does not allow a definition of Boolean negation, with its 'strengthened liar', in a naive theory of sets or truth. Indeed, as shown, the most general semantics of the operator, as given in Definition 2, is sound and complete with respect to the rule:

$$\frac{\circ A \quad A \quad \neg A}{B}$$

and so classical reasoning may be captured in exactly the same way as with the consistent consistency operator of an LFI. Moreover, this rule may be

[29] For elaboration of the idea of default classicality and a non-detachable conditional, see [**?**].

[30] In Priest's implementation of the adaptive strategy, the crucial result is stated in [**?**, p.225], Fact 3, as follows:

- If Σ is classically consistent, $\Sigma^m = \Sigma^{CL}$

where Σ^m comprises the default consequences of Σ, and Σ^{CL} its classical consequences. This is proved in ZF; and in Priest's understanding of this, the *if* is indeed a material conditional—*ibid*, 18.4. 18.5 contains a discussion of a theory of validity based on default reasoning.

added to a naive theory of sets/truth without triviality. This is shown by the fact that the rule is valid if we simply interpret $\circ A$ as $\perp$, which, as we have observed, can be part of such a theory non-trivially. (We are not, of course, suggesting that this should be taken as a definition consistency. The point is simply to give a non-triviality proof.)

The costs of such a move are twofold. First, the consistency operator must itself behave inconsistently. Secondly, it must have a non-deterministic semantics.

Regarding the first of these, we have already observed (§3) that in the context of a naive theory, this is to be expected. Simply define a liar, L of the form: L is false and consistent. If it is true, it is false and consistent, so it is true, false, and consistent; that is, consistent and inconsistent. If it is inconsistent, it is true (and false); so it is consistent and inconsistent. The only other possibility is that it is false and consistent. In that case, it is true, and so consistent and inconsistent again.

It is often claimed that a dialetheist cannot express consistency without triviality. This is just false, as we noted in §4.1. What cannot be done is to express consistency consistently—a quite different matter. The objector may say that to function as desired, consistency must be expressed consistently. But not even a classical logician can express consistency consistently. However they express the thought that A is consistent, it may still transpire that both A and $\neg A$ hold in the situation. (Witness: Frege.) And if these so transpire, the situation collapses into triviality. But this is exactly a role that our $\circ A$ fulfills. So the classical logician and the paraconsistent logician are in exactly the same situation. Neither can force consistency, and both can force triviality. As with many more mundane desires, the simple desire that consistency be consistent does not make it so.[31]

Let us turn to the second point: non-determinacy. Note, for a start, that, arguably, the truth predicate is itself non-deterministic.[32] If A has truth-value $\mathbf{b}$, $T\langle A\rangle$ may be $\mathbf{b}$, or it may simply be $\mathbf{t}$. Thus, if A is a paradoxical sentence not involving truth (maybe concerning boundaries), it has the value $\mathbf{b}$, but to say that it is true may be simply true, $\mathbf{t}$. But if A itself concerns truth, this may not be the case. If A is $\neg T\langle A\rangle$ then it has the value $\mathbf{b}$, as, then, does $T\langle A\rangle$.

We are not dealing with truth here, but with consistency; but matters

[31] See, further, [**?**, 20.4], [**?**, 6.3], and [**?**].
[32] See [**?**, 5.4]. See also [**?**, §11].

may be similar. If A is a plain vanilla truth such as '2+2=4', it has the value $\mathbf{t}$. And it is natural enough to suppose that, 'A is consistent', $\circ A$, similarly, has the value $\mathbf{t}$. But if A itself concerns consistency, it may take some other value.

Consider a sentence, A, that says of itself that it is not consistent:[33]

- $A := T\langle\neg\circ A\rangle$

The value of A cannot be $\mathbf{b}$. For then $\circ A$ would have the value $\mathbf{f}$, so its negation and $T\langle\neg\circ A\rangle$ would be $\mathbf{t}$. If the value of A is $\mathbf{t}$, then the value of $\circ A$ cannot be $\mathbf{t}$, or its negation would be $\mathbf{f}$, as would $T\langle\neg\circ A\rangle$. So the value of $\circ A$ is either $\mathbf{f}$ or $\mathbf{b}$. It is possible that the value of A is $\mathbf{f}$ and the value of $\circ A$ is $\mathbf{t}$. For then the value of its negation is $\mathbf{f}$, as is the value of $T\langle\neg\circ A\rangle$. Nothing, then, forces a value on A. That may itself be a non-determinate matter. But one possibility is that A is $\mathbf{f}$ and $\circ A$ is either $\mathbf{f}$ or $\mathbf{b}$.

There are more complex cases which leave even less wiggle room for determinacy. Let:

- $A := T\langle\neg A \wedge \circ A\rangle$

A cannot have the value $\mathbf{b}$. For then $\circ A$ has the value $\mathbf{f}$, as do $\neg A \wedge \circ A$ and $T\langle\neg A \wedge \circ A\rangle$. A cannot take the value $\mathbf{t}$. If it did, $\neg A$ would take the value $\mathbf{f}$, as would $\neg A \wedge \circ A$ and $T\langle\neg A \wedge \circ A\rangle$. So A must take the value $\mathbf{f}$. But then $\neg A$ takes the value $\mathbf{t}$. So $\circ A$ cannot take the value $\mathbf{t}$, or $\neg A \wedge \circ A$ and $T\langle\neg A \wedge \circ A\rangle$ would do so. $\circ A$ must take the value $\mathbf{b}$ or $\mathbf{f}$. In the first case, $\neg A \wedge \circ A$ takes the value $\mathbf{b}$. So $T\langle\neg A \wedge \circ A\rangle$ takes the value $\mathbf{t}$ or $\mathbf{b}$, which is impossible. Hence, $\circ A$ takes the value $\mathbf{f}$, in which case $\neg A \wedge \circ A$ does so too, and $T\langle\neg A \wedge \circ A\rangle$ takes the value $\mathbf{f}$, as required. Hence, A has the value $\mathbf{t}$, and $\circ A$ takes the value $\mathbf{f}$. Of course, there are plain vanilla falsehoods, A, such as $2 + 2 \neq 4$, where $\circ A$ is $\mathbf{t}$. So we have non-determinacy.

We note also one important feature of a non-determinate consistency operator. Unlike consistency, inconsistency can be defined in a straightforward way. Read $\bullet A$ as 'it is inconsistent that A'. Then $\bullet A$ is naturally defined as $A \wedge \neg A$. Now, $\bullet A$ is not equivalent to $\neg\circ A$. For suppose it were, and that A is $\mathbf{t}$ and $\circ A$ is $\mathbf{f}$. Then $\bullet A$ would be $\mathbf{t}$. It would follow that $A \wedge \neg A$; so A is not $\mathbf{t}$.

[33] Recall that we are treating consistency as an operator, not a predicate, though we could have a consistency predicate, $\mathcal{C}$, satisfying the condition that $\mathcal{C}\langle A\rangle \leftrightarrow \circ A$.

Actually, the failure of inter-definability has nothing essentially to do with non-determinacy. If $\bullet A$ means $A \wedge \neg A$, then $\neg\bullet A$ means $\neg(A \wedge \neg A)$, i.e., $A \vee \neg A$. For the consistency operator of an LFI, suppose that A is $\mathbf{b}$. The $\circ A$ is $\mathbf{f}$, and $\neg\circ A$ is $\mathbf{t}$. But $A \wedge \neg A$ is $\mathbf{b}$.

Non-determinacy thus opens up avenues for expressing consistency from a dialetheic standpoint with attractive properties.

6 Conclusion

The notion of consistency is ubiquitous in logic. Carnielli has built a large body of work in a tradition that focuses squarely on internalizing this notion in a logic. That approach has the virtue of recovering classical reasoning in classical cases—assuming that consistency itself is a consistent notion. We've provided some reasons to think otherwise, and outlined several ways to express the notion of consistency inconsistently, while still retaining most of its desired power.

Carnielli takes an anti-dialetheic approach, and with it, accepts that, as quoted in §3.2 above, "if it were confirmed some day that real contradictions do exist ... a considerable part of science, and also of philosophy, would collapse altogether." We suggest that there is no need for such (ungrounded) pessimism. An inconsistent approach to consistency and paraconsistency means that even in the event of a contradiction, we can carry on and live to prove another day.

Appendix 1: The Failure of Revenge

The possibility of a dialetheic solution to paradoxes of self-reference has appeared centrally in the discussion of the paper, and one standard objection to this is that it is just as subject to 'revenge paradoxes' as consistent solutions. In this appendix, we bring the contents of this paper to bear on the matter.

A convenient way to do so is to take up the points raised recently by Murzi and Scambler (M&S).[34] They argue that an *LP*ish solution to the liar paradox can avoid revenge only if the notion of consistency—or truth only—cannot be expressed.[35] The basic idea is very familiar. If we can

[34]See [**?**]. Earlier criticisms are discussed in [**?**, 20.3].

[35]Note that 'A is false only' can be defined in terms of $\circ$. 'A is false only' is $\neg A \wedge \circ A$.

define consistency, $\circ$, we can define a Boolean negation $\$A := \neg A \wedge \circ A$, and we have triviality.

M&S discuss a number of supposedly unacceptable consequences of the revenge scenario (§§3.5, 3.6). The first is that in an LP context, if L is the liar sentence, $\neg T\langle L\rangle$, L is both true and not both true and false, $T\langle L\rangle \wedge \neg(T\langle L\rangle \wedge T\langle\neg L\rangle)$ (even if truth does not commute with negation). This is equally true of simple vanilla truths, such as '$0 = 0$'. So one cannot use this fact to distinguish between the two. This is indeed so, though, note, this is a feature of this very specific paradoxical sentence. It is not true of other dialetheias—those concerning, say, arithmetic or boundaries.[36]

But in any case, there are other ways in which one can distinguish between the two.[37] Perhaps most obviously, a dialetheist about L will assert both L and $\neg L$. However, they will assert $0 = 0$ and *deny* $0 \neq 0$. True, assertion and denial are speech acts, and so cannot be antecedents of complex sentences, such as conditionals. If it be felt that there should be something that can do this,[38] the consistency connective of 5.5 will do just what is required. The value of $\circ 0 = 0$ is t; whereas that of $\circ L$ is f.

A second problem concerns the fact that if we are to avoid triviality, one cannot express consistency consistently. In this case, we can have all of 'A is consistent', A, and $\neg A$. That being so, the claim:

- if a situation is consistent, one may reason classically about it

may appear to lead to triviality. The previous sections have shown us two reasons why this does not follow. The first is that we may take the *if* here to be material, and use default reasoning 'all the way up'. The second is that one may express the claim using a consistency operator of the kind discussed in 5.5. So, taking DS as an example, we have $\circ A, A, \neg A \vee B \vdash B$. Or if we have an appropriate detachable conditional, $(\circ A \wedge A \wedge (\neg A \vee B)) \to B$.

The third problem is that, the above notwithstanding, it remains the case that there is nothing a paraconsistent logician can say which will force them not to accept A. $\neg A$ will not do. By contrast, a classical logician can assert Boolean $\$A$. As we have seen, this contrast fails. Asserting $\$A$ does not force a classical logician to reject A. All that is forced is that if they accept A,

---

And 'A is true only' can be defined as '$\neg A$ is false only'. Conversely, 'A is consistent' can be defined as 'A is false only or true only'.

[36]See, further, [**?**, 20.4].

[37]See, further, [**?**, 20.4], and [**?**, 6.3]

[38]See [**?**, p.292].

they must accept everything. And a paraconsistent logician can do exactly the same with the consistency operator of 5.4. They can assert $\neg A \wedge \circ A$.[39]

Actually, M&S go on to concede the point that a paraconsistent logician can force an endorsement of A to occasion a collapse into triviality if the language contains a detachable conditional, by endorsing $A \to \bot$ (as we have already noted). They promise to take the point up in the next section, but when they do so, they change the subject, pointing out that $A \to \bot$ cannot express the thought that A is false only.[40] This is a point we have already noted in §5.2.

In other words, all of M&S's objections fail. Consistency can be expressed in the object language—albeit inconsistently—and the situation does not collapse into triviality. There is no revenge problem.

Appendix 2: Proofs of Theorems

In this appendix we provide the details of Proposition 2 and Proposition 3 in §4.3. We begin with Proposition 2.

The details of Proposition 2

Definition 5 Let Σ be a set of formulas. Then, Σ is a *theory* iff $\Sigma \vdash A$ implies $A \in \Sigma$, Σ is *prime* iff $A \vee B \in \Sigma$ implies $A \in \Sigma$ or $B \in \Sigma$, and Σ is a *non-trivial* iff for some $A \in \mathsf{Form}$, $A \notin \Sigma$.

Lemma 1[Lindenbaum] If $\Sigma \nvdash A$, then there is $\Sigma' \supseteq \Sigma$ such that $\Sigma' \nvdash A$ and Σ' is a non-trivial prime theory.

We now define the canonical valuation in the following manner.

Definition 6 For any $\Sigma \subseteq \mathsf{Form}$, let v_Σ from Form to $\{\mathbf{t}, \mathbf{b}, \mathbf{f}\}$ be defined as follows:

[39] And if K is the Curry sentence (fn. 25), its being false only can be expressed in the obvious way, by $\neg K \wedge \circ K$, where both conjuncts have the value t. To be fair to M&S, they are considering a language with only the connectives of LP, which does not contain $\circ$. But they appear to take it that any inconsistent way of expressing consistency will fall to the previous objection.

[40] M&S attribute the thought that it does so to Priest, but the references they give do not say this at all.

$$v_\Sigma(A) := \begin{cases} \mathbf{t} \text{ iff } \Sigma \vdash A \text{ and } \Sigma \nvdash \neg A; \\ \mathbf{b} \text{ iff } \Sigma \vdash A \text{ and } \Sigma \vdash \neg A; \\ \mathbf{f} \text{ iff } \Sigma \nvdash A \text{ and } \Sigma \vdash \neg A. \end{cases}$$

Note that we are defining the canonical valuation in a different manner compared to the usual way, e.g. for LP, reflecting the difference of how deterministic and non-deterministic semantics are introduced.

Lemma 2 If Σ is a non-trivial prime theory, then v_Σ is a well-defined three-valued LPmnC-valuation.

Proof. Note first that the well-definedness of v_Σ is obvious. Then the desired result is proved by induction on the number n of connectives. Base case: For atomic formulas, it is obvious by the definition. Induction step: We split the cases based on the connectives. Here we only deal with $\circ$. If $A = \circ B$, then we have the following three cases.

Cases	$v_\Sigma(B)$	condition for B	$v_\Sigma(A)$	condition for A i.e. $\circ B$
(i)	$\mathbf{t}$	$\Sigma \vdash B$ and $\Sigma \nvdash \neg B$	$\mathbf{t}, \mathbf{b}, \mathbf{f}$	—
(ii)	$\mathbf{b}$	$\Sigma \vdash B$ and $\Sigma \vdash \neg B$	$\mathbf{f}$	$\Sigma \nvdash \circ B$ and $\Sigma \vdash \neg\circ B$
(iii)	$\mathbf{f}$	$\Sigma \nvdash B$ and $\Sigma \vdash \neg B$	$\mathbf{t}, \mathbf{b}, \mathbf{f}$	—

By induction hypothesis, we have the conditions for B, for cases (i) and (iii), we have nothing to do. For (ii), we can use RECQ. Indeed, in view of LEM and that Σ is prime, it is sufficient to establish $\Sigma \nvdash \circ B$, given the assumption that $\Sigma \vdash B$ and $\Sigma \vdash \neg B$. Then, if we assume, for reductio, that $\Sigma \vdash \circ B$, then together with the assumptions that RECQ, we obtain that Σ is trivial, which is absurd. □

We are now ready to prove Proposition 2.

Proof. For the soundness direction, it can be shown by a straightforward verification that each rule preserves designated values. Here we only spell out the details for the validity of RECQ. Suppose, for reductio, that there is a three-valued LPmnC-valuation v_0 such that $v_0(A) \in \mathcal{D}$, $v_0(\neg A) \in \mathcal{D}$, and $v_0(\circ A) \in \mathcal{D}$. Then, there are two cases, either $v_0(A)$=$\mathbf{t}$ or $v_0(A)$=$\mathbf{b}$. If the former holds, then this is absurd in view of $v_0(\neg A) \in \mathcal{D}$ since $v_0(\neg A)$=$\mathbf{f}$. If the latter holds, then this is absurd in view of $v_0(\circ A) \in \mathcal{D}$ since $v_0(\circ A)$=$\mathbf{f}$.

For the completeness direction, assume $\Gamma \nvdash A$. Then, by Lemma 1, there is a $\Sigma \supseteq \Gamma$ such that Σ is a non-trivial prime theory and $A \notin \Sigma$, and by Lemma 2, a three-valued LPmnC valuation v_Σ can be defined with $v_\Sigma(B) \in \mathcal{D}$ for every $B \in \Gamma$ and $v_\Sigma(A) \notin \mathcal{D}$. Thus it follows that $\Gamma \not\models A$, as desired. □

The details of Proposition 3

For the details of Proposition 3, we need to check that (i) the rule preserves designated values, and (ii) Lemma 2 can be suitably modified.

Proof. We deal with all six cases.
Case $\tilde{\circ}\mathbf{t} = \{\mathbf{t}, \mathbf{f}\}$: suppose, for reductio, that there is a three-valued valuation v_0 such that $v_0(\circ A) \in \mathcal{D}$, $v_0(\neg\circ A) \in \mathcal{D}$, and $v_0(\neg A) \notin \mathcal{D}$. Then, in view of $v_0(\neg A) \notin \mathcal{D}$, we have $v_0(A){=}\mathbf{t}$. Then, in view of the Nmatrices, there are two cases, either $v_0(\circ A){=}\mathbf{t}$ or $v_0(\circ A){=}\mathbf{f}$. If the former holds, then this is absurd in view of $v_0(\neg\circ A) \in \mathcal{D}$. If the latter holds, then this is absurd in view of $v_0(\circ A) \in \mathcal{D}$.

For the modification of Lemma 2, we need to check the case (i) since that is the only difference from the basic case we covered above.

Cases	$v_\Sigma(B)$	condition for B	$v_\Sigma(A)$	condition for A i.e. $\circ B$
(i)	$\mathbf{t}$	$\Sigma \vdash B$ and $\Sigma \nvdash \neg B$	$\mathbf{t}, \mathbf{f}$	$\Sigma \nvdash \circ B$ or $\Sigma \nvdash \neg\circ B$
(ii)	$\mathbf{b}$	$\Sigma \vdash B$ and $\Sigma \vdash \neg B$	$\mathbf{f}$	$\Sigma \nvdash \circ B$ and $\Sigma \vdash \neg\circ B$
(iii)	$\mathbf{f}$	$\Sigma \nvdash B$ and $\Sigma \vdash \neg B$	$\mathbf{t}, \mathbf{b}, \mathbf{f}$	—

Assume, for reductio, that $\Sigma \vdash \circ B$ and $\Sigma \vdash \neg\circ B$. Then, by the additional rule, we obtain that $\Sigma \vdash \neg B$, but this is absurd in view of assumption.
Case $\tilde{\circ}\mathbf{t} = \{\mathbf{b}, \mathbf{f}\}$: suppose, for reductio, that there is a three-valued valuation v_0 such that $v_0(\neg A \vee \neg\circ A) \notin \mathcal{D}$. Then, in view of the Nmatrices, $v_0(\neg A) \notin \mathcal{D}$ and $v_0(\neg\circ A) \notin \mathcal{D}$. The former condition is equivalent to $v_0(A){=}\mathbf{t}$, and in view of the Nmatrices, $v_0(\circ A){=}\mathbf{b}$ or $v_0(\circ A){=}\mathbf{f}$. But this is absurd in view of $v_0(\neg\circ A) \notin \mathcal{D}$.

For the modification of Lemma 2, we need to check the case (i) since that is the only difference from the basic case we covered above.

Cases	$v_\Sigma(B)$	condition for B	$v_\Sigma(A)$	condition for A i.e. $\circ B$
(i)	$\mathbf{t}$	$\Sigma \vdash B$ and $\Sigma \nvdash \neg B$	$\mathbf{b}, \mathbf{f}$	$\Sigma \vdash \neg\circ B$
(ii)	$\mathbf{b}$	$\Sigma \vdash B$ and $\Sigma \vdash \neg B$	$\mathbf{f}$	$\Sigma \nvdash \circ B$ and $\Sigma \vdash \neg\circ B$
(iii)	$\mathbf{f}$	$\Sigma \nvdash B$ and $\Sigma \vdash \neg B$	$\mathbf{t}, \mathbf{b}, \mathbf{f}$	—

Then, by the additional rule, and the primeness of Σ, $\Sigma \nvdash \neg B$ implies $\Sigma \vdash \neg\circ B$, as desired.

Case $\tilde{\circ}\mathbf{t} = \{\mathbf{f}\}$: suppose, for reductio, that there is a three-valued valuation v_0 such that $v_0(\circ A) \in \mathcal{D}$ and $v_0(\neg A) \notin \mathcal{D}$. The latter condition is equivalent to $v_0(A)$=$\mathbf{t}$, and in view of the Nmatrices, $v_0(\circ A)$=$\mathbf{f}$. But this is absurd in view of $v_0(\circ A) \in \mathcal{D}$.

For the modification of Lemma2 , we need to check the case (i) since that is the only difference from the basic case we covered above.

Cases	$v_\Sigma(B)$	condition for B	$v_\Sigma(A)$	condition for A i.e. $\circ B$
(i)	$\mathbf{t}$	$\Sigma \vdash B$ and $\Sigma \nvdash \neg B$	$\mathbf{f}$	$\Sigma \nvdash \circ B$ and $\Sigma \vdash \neg\circ B$
(ii)	$\mathbf{b}$	$\Sigma \vdash B$ and $\Sigma \vdash \neg B$	$\mathbf{f}$	$\Sigma \nvdash \circ B$ and $\Sigma \vdash \neg\circ B$
(iii)	$\mathbf{f}$	$\Sigma \nvdash B$ and $\Sigma \vdash \neg B$	$\mathbf{t}, \mathbf{b}, \mathbf{f}$	—

Then, in view of LEM and that Σ is prime, it is sufficient to establish $\Sigma \nvdash \circ B$, given the assumption that $\Sigma \vdash B$ and $\Sigma \nvdash \neg B$. Then, if we assume, for reductio, that $\Sigma \vdash \circ B$, then together with the additional rule, we obtain that $\Sigma \vdash \neg B$, which is absurd in view of the second assumption.

Case $\tilde{\circ}\mathbf{f} = \{\mathbf{t}, \mathbf{f}\}$: suppose, for reductio, that there is a three-valued valuation v_0 such that $v_0(\circ A) \in \mathcal{D}$, $v_0(\neg\circ A) \in \mathcal{D}$, and $v_0(A) \notin \mathcal{D}$. Then, in view of $v_0(A) \notin \mathcal{D}$, we have $v_0(A)$=$\mathbf{f}$. Then, in view of the Nmatrices, there are two cases, either $v_0(\circ A)$=$\mathbf{t}$ or $v_0(\circ A)$=$\mathbf{f}$. If the former holds, then this is absurd in view of $v_0(\neg\circ A) \in \mathcal{D}$. If the latter holds, then this is absurd in view of $v_0(\circ A) \in \mathcal{D}$.

For the modification of Lemma 2, we need to check the case (iii) since that is the only difference from the basic case we covered above.

Cases	$v_\Sigma(B)$	condition for B	$v_\Sigma(A)$	condition for A i.e. $\circ B$
(i)	$\mathbf{t}$	$\Sigma \vdash B$ and $\Sigma \nvdash \neg B$	$\mathbf{t}, \mathbf{b}, \mathbf{f}$	—
(ii)	$\mathbf{b}$	$\Sigma \vdash B$ and $\Sigma \vdash \neg B$	$\mathbf{f}$	$\Sigma \nvdash \circ B$ and $\Sigma \vdash \neg\circ B$
(iii)	$\mathbf{f}$	$\Sigma \nvdash B$ and $\Sigma \vdash \neg B$	$\mathbf{t}, \mathbf{f}$	$\Sigma \nvdash \circ B$ or $\Sigma \nvdash \neg\circ B$

Assume, for reductio, that $\Sigma \vdash \circ B$ and $\Sigma \vdash \neg\circ B$. Then, by the additional rule, we obtain that $\Sigma \vdash B$, but this is absurd in view of assumption.

Case $\tilde{\circ}\mathbf{f} = \{\mathbf{b}, \mathbf{f}\}$: suppose, for reductio, that there is a three-valued valuation v_0 such that $v_0(A \vee \neg\circ A) \notin \mathcal{D}$. Then, in view of the Nmatrices, $v_0(A) \notin \mathcal{D}$ and $v_0(\neg\circ A) \notin \mathcal{D}$. The former condition is equivalent to $v_0(A)$=$\mathbf{f}$, and in view of the Nmatrices, $v_0(\circ A)$=$\mathbf{b}$ or $v_0(\circ A)$=$\mathbf{f}$. But this is absurd in view of $v_0(\neg\circ A) \notin \mathcal{D}$.

For the modification of Lemma 2, we need to check the case (iii) since that is the only difference from the basic case we covered above.

Cases	$v_\Sigma(B)$	condition for B	$v_\Sigma(A)$	condition for A i.e. $\circ B$
(i)	$\mathbf{t}$	$\Sigma \vdash B$ and $\Sigma \nvdash \neg B$	$\mathbf{t}, \mathbf{b}, \mathbf{f}$	—
(ii)	$\mathbf{b}$	$\Sigma \vdash B$ and $\Sigma \vdash \neg B$	$\mathbf{f}$	$\Sigma \nvdash \circ B$ and $\Sigma \vdash \neg\circ B$
(iii)	$\mathbf{f}$	$\Sigma \nvdash B$ and $\Sigma \vdash \neg B$	$\mathbf{b}, \mathbf{f}$	$\Sigma \vdash \neg\circ B$

Then, by the additional rule, and the primeness of Σ, $\Sigma \nvdash B$ implies $\Sigma \vdash \neg\circ B$, as desired.

Case $\tilde{\circ}\mathbf{f} = \{\mathbf{f}\}$: suppose, for reductio, that there is a three-valued valuation v_0 such that $v_0(\circ A) \in \mathcal{D}$ and $v_0(A) \notin \mathcal{D}$. The latter condition is equivalent to $v_0(A)=\mathbf{f}$, and in view of the Nmatrices, $v_0(\circ A)=\mathbf{f}$. But this is absurd in view of $v_0(\circ A) \in \mathcal{D}$.

For the modification of Lemma 2, we need to check the case (iii) since that is the only difference from the basic case we covered above.

Cases	$v_\Sigma(B)$	condition for B	$v_\Sigma(A)$	condition for A i.e. $\circ B$
(i)	$\mathbf{t}$	$\Sigma \vdash B$ and $\Sigma \nvdash \neg B$	$\mathbf{t}, \mathbf{b}, \mathbf{f}$	—
(ii)	$\mathbf{b}$	$\Sigma \vdash B$ and $\Sigma \vdash \neg B$	$\mathbf{f}$	$\Sigma \nvdash \circ B$ and $\Sigma \vdash \neg\circ B$
(iii)	$\mathbf{f}$	$\Sigma \nvdash B$ and $\Sigma \vdash \neg B$	$\mathbf{f}$	$\Sigma \nvdash \circ B$ and $\Sigma \vdash \neg\circ B$

Then, in view of LEM and that Σ is prime, it is sufficient to establish $\Sigma \nvdash \circ B$, given the assumption that $\Sigma \nvdash B$ and $\Sigma \vdash \neg B$. Then, if we assume, for reductio, that $\Sigma \vdash \circ B$, then together with the additional rule, we obtain that $\Sigma \vdash B$, which is absurd in view of the first assumption. □

Acknowledgments: This work was supported in part by the Alexander von Humboldt Foundation and the Royal Society of New Zealand.

References

[1] Avron, A., and Lev, I. (2005), 'Non-deterministic multiple-valued structures', *Journal of Logic and Computation* 15(3): 241–261.

[2] Avron, A., and Zamansky, A. (2011), 'Non-deterministic semantics for logical systems', pp. 227–304, Vol. 16, of D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, 2nd edn, Dordrecht: Kluwer Academic Publishers

[3] Batens, D. (1989), 'Dynamic dialectical logics', ch. 6 of G. Priest, R. Routley, and J. Norman (eds.), *Paraconsistent Logic: Essays on the Inconsistent*, Munich: Philosophia Verlag.

[4] Beall, J. (2013), 'Free of detachment: logic, rationality, and gluts', *Noûs* 49: 410–423.

[5] Belnap, N. (1977), 'A useful four-valued logic', in G. Epstein and J. M. Dunn (eds.), *Modern Uses of Multiple Valued Logics*, Dordrecht: Reidel.

[6] Carnielli, W., Coniglio, M., and Marcos, J. (2007), 'Logics of formal inconsistency', in D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, 2nd edn, Vol. 14, Dordrecht: Springer.

[7] Carnielli, W., and Coniglio, M. (2016), *Paraconsistent Logic: Consistency, Contradiction and Negation*, Cham: Springer.

[8] Carnielli, W., Coniglio, M., and Rodrigues, A. (2020), 'Recovery operators, paraconsistency and duality', *Logic Journal of the IGPL*, 28(5): 624–656.

[9] Carnielli, W., Marcos, J., and De Amo, S. (2000), 'Formal inconsistency and evolutionary databases'. *Logic and Logical Philosophy*, 8: 115–152.

[10] Carnielli, W., and Rodrigues, A. (2017), 'An epistemic approach to paraconsistency: a logic of evidence and truth,' *Synthese* 196: 3789–3813.

[11] Carnielli, W., and Rodrigues, A. (2021), 'On epistemic and ontological interpretations of intuitionistic and paraconsistent paradigms', *Logic Journal of the IGPL*, 29(4): 569–584.

[12] da Costa, N. (1974), 'On the theory of inconsistent formal systems', *Notre Dame Journal of Formal Logic* 15: 497–510.

[13] da Costa, N., and Alves, E. (1977), 'A semantical analysis of the calculi C_n', *Notre Dame Journal of Formal Logic* 18: 621–630.

[14] De, M., and Omori, H. (2015), 'Classical negation and expansions of Belnap–Dunn logic', *Studia Logica* 103(4): 825–851.

[15] Dunn, J. M. (1976), 'Intuitive semantics for first-degree entailments and "coupled trees"', *Philosophical Studies* 29:149–168.

[16] Murzi, J., and Scambler, C. (2022), 'Revenge', in L. Rossi (ed.), *The Liar Paradox*, Cambridge: Cambridge University Press.

[17] Omori, H. (2015), 'Remarks on naive set theory based on LP', *The Review of Symbolic Logic*, 8(2): 279–295.

[18] Omori, H., and Weber, Z. (2019), 'Just true? On the metatheory for paraconsistent truth', *Logique et Analyse* 248: 415–433.

[19] Priest, G. (1979), 'Logic of paradox', *Journal of Philosophical Logic* 8: 219–41.

[20] Priest, G. (1987), *In Contradiction*, Dordrecht: Martinus Nijhoff; 2nd edn, Oxford: Oxford University Press, 2006.

[21] Priest, G. (2002), 'Paraconsistent logic', pp. 287–393, Vol. 6, of D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, 2nd edn, Dordrecht: Kluwer Academic Publishers.

[22] Priest, G. (2006), *Doubt Truth to be a Liar*, Oxford: Oxford University Press.

[23] Priest, G. (2010), 'Hopes fade for saving truth', *Philosophy* 55: 109–140.

[24] Priest, G. (2017), 'What if? The exploration of an idea', *Australasian Journal of Logic* 14(1): art. 4, https://ojs.victoria.ac.nz/ajl/article/view/4028/3574.

[25] Priest, G. (2019), 'Some comments and replies', ch. 27 of C. Başkent and T. Ferguson (eds.), *Graham Priest on Dialetheism and Paraconsistency*, Cham: Springer.

[26] Restall, G. (1992), 'A note on naive set theory in LP', *Notre Dame Journal of Formal Logic*, 33(3): 422–432.

[27] Rodrigues, A., Bueno-Soler, J., and Carnielli, W. (2021), 'Measuring evidence: a probabilistic approach to an extension of Belnap-Dunn logic', *Synthese*, 198: 5451–5480.

[28] Szmuc, D., and Omori, H. (2022). 'Liberating classical negation from falsity conditions', pp. 131–136 of *2022 IEEE 52nd International Symposium on Multiple-Valued Logic (ISMVL)*, IEEE.

[29] Weber, Z. (2021), *Paradoxes and Inconsistent Mathematics*, Cambridge: Cambridge University Press.