Logique & Analyse **217** (2012), 3–21

PARACONSISTENCY ON THE ROCKS OF DIALETHEISM

CONRAD ASMUS

Abstract

Can one be a non-dialetheic paraconsistentist? I will show that on a standard model-theoretic approach to consequence the answer to this question depends on the philosophical motivation behind the models. If the models are interpretations of the formal language, the answer to the question is "No", but if the models are representations of how things are, the answer is less clear. That different approaches to semantic characterisations of consequence come apart in this way demonstrates that attention should be paid, especially by paraconsistentists, to the motivations behind the theories.

1. *Introduction*

Different philosophies of logic can combine with the same logic to entail different commitments. In particular, different views on semantic characterisations of paraconsistent consequence stand in different relations to dialetheism (the view that some contradictions are true). This demonstrates that careful attention should be paid to the motivations behind logics.

Semantic characterisations of consequence, particularly model theoretic approaches, claim that the conclusion of an argument is a consequence of the premises if the truth of the premises guarantees the truth of the conclusion. In model theoretic approaches this is equated with truth preservation across models:

> ϕ is a consequence of Γ if and only if ϕ is true in every model where each member of Γ is.

In this paper I will demonstrate the importance of elaborating on what models are meant to do. I will show that:

(1) on some semantic accounts of consequence, paraconsistent logics immediately give rise to dialetheism,

(2) in others, paraconsistent logics are in dire danger of crashing onto dialetheic rocks — but not all hope is lost, and
(3) other accounts do not, in any way, force paraconsistency towards dialetheism.

These conclusions will show that it is very important for paraconsistentists who give semantic characterisations of their logic to also elaborate on the importance of the models. This should, moreover, encourage *anyone* engaged in giving semantic theories of consequence to provide the philosophical grounds for their project. That the choices offered below have such significant import for paraconsistentists demonstrates that no one making use of semantic characterisations of consequence can *assume* that they need not elaborate on the motivation behind their theory.

2. *Consequence and Cases*

The conclusion of an argument, C, is a consequence of the set of premises Γ if and only if whenever each member of Γ is true, so is C. We can clarify this notion by explicitly quantifying over *cases* in which sentences are true.

> C is a consequence$_\alpha$ of Γ if and only if in every case$_\alpha$ in which each member of Γ is true, so is C.[1]

The inclusion of α highlights that there may be a plurality of acceptable classes of cases or, at least, that there may be reasonable dispute over which is the right class of cases. I will refer to a selection of cases as a consequence theory.

The minimal requirement on cases for a theory of consequence is that they deem some sentences true and others not. Anything which demarcates the sentences of a language into the true and the not true is a *case*. Models (standard models for first order classical logic, Kripke models for intuitionistic logic, Routley-Meyer models for relevant logics and so on) are typical cases. They provide a class of sentences which are true in the model. The class of cases and the class of models are different. The class of cases is broader than the class of models, but every model is a case. That said, most *cases* in which philosophers and logicians are interested can be *captured* by *models* of one sort or another (I will say more about this shortly). I will deal with the more general class in this paper for reasons which will become evident.

[1] This is an adaptation of the generalised Tarski thesis found in [2].

Other examples of cases are fictions, belief sets, possibilities, the games of game theoretic semantics, or even whiteboards with sentences written on them. In each example the case specifies a number of (hypothetically) true sentences.

Not every account of consequence uses cases. Proof theoretic accounts do not. Some versions of algebraic semantics do not. Algebraic semantics, or semantics based on algebras, often analyse consequence in terms of an ordering within an algebra. An argument is valid iff in every algebra of the right sort (perhaps Boolean algebras, or Heyting algebras), every valuation of the premises is *less than or equal* to the valuation of the conclusion. This approach makes no use of *truth* (or truth in a case) and so does not use cases. Often, however, this can be converted into a case-based characterisation of consequence. If we can show that a non-case-based account of consequence is equivalent (via soundness and completeness results) to a case-based one, then, even if the non-case-based approach is taken to be primary, what follows may still be relevant if the equivalent case-based approach is taken to be of some philosophical import.[2]

The more serious requirement that must be met for a collection of cases to characterise consequence is that the choice of cases has to be well motivated. In general we need to know why the choice of cases matters with respect to consequence. Using all the whiteboards in the world as the cases of a theory

[2] There are, unfortunately, two uses of the term "algebraic semantics" in the literature. The first is as used here: an algebraic semantics is constructed on an algebra rather than a relational structure. This should not be confused with the second usage introduced by Dummett. Dummett [8, 9] says that a "characterisation is algebraic rather than semantic when we lack any means of using the algebra to give the meanings of the logical constants" [9, pp. 40–41] and "Semantic notions are framed in terms of concepts which are taken to have a direct relation to the use which is made of sentences of a language ... Corresponding algebraic notions define a valuation as a purely mathematical object — an open set, or a natural number — which has no intrinsic connection with the uses of sentences" [8, pp. 293]. The merely algebraic characterisation offer no philosophical insight into consequence. A properly semantic theory connects in an appropriate way to our use of language (or attempts to, Dummett is happy to recognise classical semantics as a *semantics* even though he thinks they fail to provide the appropriate connection to our practice in certain contexts). This usage has become tangled in the literature. Copeland (for example [6, 7]) associates Dummett's algebraic semantics with Plantinga's "pure semantics" and contrasts them to "applied" or "depraved" semantics. Copeland argues that Routley-Meyer semantics for relevant logics are not applied, and thus, using the connection with Dummett, a merely algebraic semantics. The Routley-Meyer semantics is not algebraic in the sense I am using. It is a relational semantics using a notion of truth in a model and not based on an algebraic *less than or equals* ordering. As indicated previously, the Routley-Meyer semantics provide prime examples of cases for this paper. A model in the Routley-Meyer semantics provides a specification of true sentences relative to a base world, this is a case of the sort considered here. We can then consider truth preservation across these cases.

has no obvious connection to consequence but taking all the interpretations of a language may.

In this paper I will focus on interpretational and representational theories as expounded by Etchemendy in [10] and [11]. In interpretational theories the cases are all and only the semantically well behaved interpretations of a language (with the proviso that the interpretation of the logical vocabulary is not altered). The underlying motivation is that this declares an argument valid if and only if, whatever the non-logical vocabulary means, if the premises are true so is the conclusion. In representational theories each case represents a logically possible way that the world could be and each logical possibility is represented by some case. Again there is a straightforward (if seemingly circular) motivation for this type of theory. If an argument is representationally valid, then, in every logically possible way things could be, if the premises are true, so is the conclusion. If an argument isn't representationally valid, then there is a logically possible way the world could be which serves as a counterexample to this argument.[3]

The significant difference between interpretational and representational theories is that the former considers which sentences of a language would be true if the words meant something else, holding constant the way everything else is, while the latter considers which sentences would be true if the world were other than it is, holding constant the meaning of the expressions of the language. The following metaphor may be helpful. The cases of representational theories are like fiction books in that they represent the world as being a certain way. If the book contains the sentence "The Chinese conquered Africa in the early 20^{th} century", then the book represents the world as such that the mentioned sentence is true. Notice that the meaning of the sentence is used in determining the specifics of how the world is represented. The cases of interpretational theories, on the other hand are similar to dictionaries. The dictionary gives the meaning of the expressions of a sentence and on the basis of this, and how thing are, we determine whether the sentence is true. If the dictionary explained that the previously mentioned sentence meant that New Zealand has had female prime ministers then the sentence, given this interpretation, would be true.

There will be consequence relations (say consequence$_\alpha$ and consequence$_\beta$ defined by using the distinct classes of cases, cases$_\alpha$ and cases$_\beta$, respectively) which are equivalent with respect to the arguments they deem valid (call them extensionally equivalent) but different with respect to which cases they depend on (call them intensionally different). It is tempting to say that there is no need to choose between extensionally equivalent, but intensionally different theories of consequence. A way of understanding the

[3] Later in this paper I will investigate dropping the second half of representationalism.

current paper is that it shows this should not be a default position. In [10], Etchemendy argues that, even though interpretational and representational theories are extensionally equivalent for the standard first order language, interpretational theories are unacceptable as characterisations of consequence but (as clarified in [11]) representational theories are. I will provide another example of where potentially extensionally equivalent but intensionally different consequence relations can commit one to different positions.

2.1. *Models and Cases*

The *case* based approach to consequence is, of course, very similar to model-theoretic accounts of consequence. Models (whether first order classical models, second order models, Kripke models for modal logics, Routley-Meyer models for relevant logics, ...) are cases. They categorise the sentences of a language into the true and the false. A nice property of models is that they specify which sentences are true according to them in a systematic manner. This is done using recursive clauses which echo the recursive structure of the languages they are for. A model is usually taken to have greater significance than other types of cases — it is not *just* because they are easier to work with than some other cases that logicians study them so deeply. This is because they can, or have been thought to, play the role of an interpretation of a language. A model is taken as interpreting the language in accord with the semantic values it assigns to the non-logical vocabulary and the recursive clauses it uses to deal with logical vocabulary. A collection of models can be used as an interpretational theory of logical consequence, where the collection contains all and only the appropriate interpretations of the language.

Why do I only say "can be used", rather than "is"? While for some logicians and philosophers models and interpretations are interchangeable, I will not (and should not) take them to be. A model can be used as a representation instead of as an interpretation. If we start with an interpreted language, then assigning truth to some sentences can be taken as representing that the world is such that those sentences are true. In this way a collection of models can be used as a representational theory of consequence, where the collection contains all and only the appropriate representations.

Different models can either be taken as varying in interpretation or varying in representation. A model can serve as an interpretation of a language or as a representation of how the world can be. Models can stand in the place of many cases. As this paper will explore the differences between extensionally equivalent but intensionally different theories of consequence, it is important to consider what function a model is intended to perform.

Can models play the role of any sort of case? This depends on how far one wants to stretch the label "model". The collection of declarative sentences written in my diary is a case — it is a collection of sentences deemed to be true according to it. This seems not to be a model, as there is no structure to the collection. I take it to be integral to a model that it determines true sentences in a systematic, recursive manner, however I will take no stance on whether the collection of models is as broad as the collection of cases.

2.2. *Pure and Applied Theories*

I have said that a collection of cases requires some sort of motivation before it should be accepted as a philosophically interesting theory of consequence. I have also said that when one puts forward a collection of models as a theory of consequence one should include a description of the function these models are serving. The reader may legitimately wonder why I have not put either of these requirements in terms of the common distinction between pure and applied semantics.

In [15] Plantinga rightly points out that for Kripke model structures to provide a semantics for modal languages they must be an applied semantics. An applied semantics is one which gives meaning to the interpreted language. A pure semantics either does not, or has not been demonstrated to, provide the appropriate meaning. There are three important reasons I am steering clear of this terminology. The first is that I am primarily concerned with theories of consequence, not with semantics in general. In Carnap [5] and others, we find the position that a semantic theory, a theory of the meaning of the terms in a language, provides a theory of consequence. If we provide all the interpretations of a language, we can then look at truth preservation across these interpretations. If your theory of consequence is based on a semantic theory, then your theory of consequence should provide the meaning of, at least, the logical vocabulary and thus be (in Plantinga's terms) applied. It doesn't follow that the only way to provide an account of consequence is by providing a semantic theory. That a semantic theory provides a philosophically interesting theory of consequence (provided that the semantic theory is applied) is one thing, that a philosophically interesting theory of consequence must provide an applied semantic theory is something else. It may be that the latter condition must be met by theories of consequence, but this is not an issue on which I will take a stand. This is particularly warranted in this context as I will be exploring fundamentally different approaches to providing theories of consequence. Applied semantics may result from some theories of consequence, but may not need to result from others.

The second reason for avoiding the pure/applied terminology is the apparent divergence of use in the literature. The common place to cite is Plantinga's discussion of Kripke model structures. This is odd as Plantinga introduces the discussion with the phrase "Logicians commonly distinguish between ..." [15, pp. 127]. *One* earlier use of a similar distinction is by Carnap in [5]. Here pure and descriptive semantic pursuits are distinguished. A descriptive semantic pursuit is one where the goal is to investigate a particular historical language. This is, for Carnap the empiricist, an empirical matter. The contrast is to pure semantic investigations "where we ... set up a system of semantical rules, whether in close connection with a historically given language or freely invented". There is no distinction here between pure and applied/descriptive semantics — any semantic system (system of semantic rules) is part of a pure semantic investigation. Plantinga's distinction is applied to semantic *systems*. For Carnap a semantics gives an interpretation, it just may not be the *appropriate* interpretation.

In her textbook, *The Philosophy of Logics* [12], Haack gives four levels for looking at the sentential calculus: "(i) the axioms/rules of inference, (ii) the formal interpretation (matrices), (iii) the ordinary language reading of (i), (iv) the informal explanation of (ii)" [12, pp. 30]. She says that "levels (ii) and (iv) are dubbed by Plantinga ... 'pure' and 'depraved' semantics, respectively" [12, pp. 30]. I agree with Haack that all four levels are required in understanding the sentential calculus. That extensionally equivalent, intensionally different, theories of consequence commit one to different additional conclusions (this is the core of the position I will argue for here) implies that one should provide an informal explanation of the models used in a model theoretic characterisation of consequence. Perhaps this paper calls for a depraved semantics in Haack's sense, but this does not seem to fit with Plantinga's distinction. Plantinga is not merely requesting an informal explanation of the models of a semantics. He insists that the elements of the Kripke models *be* possible worlds, Kripke's informal explanation was not enough.

Copeland (for example [6, 7]) combines a number of distinctions. I have already commented on the unfortunate manner in which "algebraic" is used in multiple ways in the literature. Copeland may be incorrect in identifying Dummett's "merely algebraic" semantics with pure semantics [8, 9]. One of the hallmarks of Plantinga's pure semantics (for example, Kripke model structures) is that one can obtain an appropriate applied semantics by placing further restrictions on the models (the set of worlds really are possible worlds etc.). But Dummett says of "merely algebraic" characterisations of consequence that "No one would think of calling [them] a *semantic* theory" [9, pp. 81]. Dummett gives us two examples to work with: "The same contrast obtains between the interpretation of the modal system S4 in terms of sets of real numbers under the usual topology and its interpretation in terms of possible worlds, or of intuitionistic logic in terms of open subsets of the real line and in terms of Beth trees" [9, pp. 81]. In both cases the first of the pair is merely algebraic and the latter is not. This sounds similar to Plantinga's distinction. In the merely algebraic characterisation sets of real numbers are used as opposed to possible worlds. We shouldn't be too hasty; in this context "possible worlds" could mean the *real* possible worlds which Plantinga requires, or it could be an informal gloss of the Kripke style models. There is another label that Dummett uses which is less frequently raised. Dummett calls Kripke model structures and Beth Trees "skeletal" semantics [9, pp. 151–157]. They are not a semantic theory *but with the appropriate additions* may be transformed into one (in particular we need an independent specification of the possible worlds). What the appropriate additions are may diverge for Dummett and Plantinga, but skeletal semantics seem far closer to pure semantics than "merely algebraic" semantics do. I think that Copeland identified that wrong pair, but this simply is not very clear and requires further discussion if any identification is to be of more use than harm.

With these, on the face of it, different uses of the terms "pure" and "applied", and their misleading connection to "merely algebraic" semantics, it seems safest to avoid them.

The third reason is that, even on the most common interpretation of the pure/applied distinction, it does not strike at the heart of the issue under investigation here. Consider this passage taken from Plantinga:

> An important difference between a pure and an applied semantics is that the latter places more conditions upon the notion of modelhood ... in the applied semantics, therefore, a model structure will not be just *any* triple (G, K, R) where G is a member of K, and R is reflexive; K will be a set of possible worlds (not chessmen) — possible states of affairs of a certain kind — of which G is a member.

Even if the demands of Plantinga are met by a model theoretic semantics, this does not determine whether the result is an interpretational or representational theory of consequence. Suppose that someone answers Plantinga's challenge (perhaps implementing Plantinga's own proposed solution). In that context it will be clear that they have provided the intended interpretations of the language. They will be able to go on a provide and interpretational theory of consequence. But the same models may be taken as representations of logically possible ways the world could be. The interpretational and representational theories arising from these depraved models result in extensionally equivalent theories of consequence. As mentioned, one aim of this paper is to show that committing to extensionally equivalent, but intensionally different theories of consequence, can have significantly different results, so I am asking for something other than the standard notion of an applied semantics.

With this (rather extensive) groundwork out the way we can now get to the core of the paper.

3. *Paraconsistency and Dialetheism*

A logic is paraconsistent if it deems the argument form with the premises A and $\neg A$ to the conclusion B invalid. Any consequence theory in which this fails will be called a paraconsistent consequence theory. If the consequence fails as the result of a counterexample in the consequence theory then there is some case in which A and $\neg A$ are both true. I will show that this results in dialetheism for some choices of cases but not for others.

Some readers may be wondering why I will be arguing for the conditional claim that *if* there is an inconsistent (but not trivial) case *then* this results in dialetheism (depending on what the cases are). After all, it seems to follow from the failure of the consequence, and that consequence is defined by quantifying over cases, that there is such a case. Classically we can reason as follows:

- $\neg(\forall x)($ if A is true in x and $\neg A$ is true in x then B is true in $x)$
- So, $(\exists x)(A$ is true in x and $\neg A$ is true in x but B isn't true in $x)$
- Thus, $(\exists x)(A$ is true in x and $\neg A$ is true in $x)$

This does not hold true when reasoning with many non-classical logics. There are two, paraconsistently motivated, ways in which the above reasoning can be rejected. In paraconsistent logics there is room for two different conjunctions: the extensional, truth-functional connective $\wedge$ and the intensional, non-truth-functional connective $\circ$ (often called fusion). If the 'but' in the chain of inferences was extensional and the conditional is not the material conditional, then the first inference move is unacceptable. If the 'but' is intensional, then the first move is acceptable but the second is not. Either way, if the conditional is not the material conditional, the inference does not go through. This raises the interesting question of how premises should be combined in arguments. Mares [13] argues that the premises should be combined in an intensional manner. So, if he is right, it is the second of the inferences which must be the focussed on. It should also be noted that these claims involve restricted quantification — it is not clear what the right implementation of restricted quantification in many paraconsistent logics is (see [3]).

The assumption that a paraconsistent consequence theory contains an inconsistent case is, nonetheless, a good one. Most consequence theories for paraconsistent logics *do* contain an inconsistent case, and *rely* on these cases to block trivialising implications. Moreover, a lot of the meta-theory for non-classical is done with classical reasoning, in these cases the above reasoning goes through. One may legitimately avoid what is to follow by refusing to use classical reasoning in meta-theory and inconsistent cases (for example as Read [21] and Routley and Routley [22] do).[4] I applaud this approach to paraconsistency, but it should be noted that, if it is accepted, the significance of many meta-theoretic results will need to be revisited.

Dialetheism is the view (most forcefully defended by Priest [16] and Routley [23]) that there are true contradictions. That is, there is at least one sentence A such that both A and $\neg A$ are true.[5] I will call any such sentence a *truth glut* (similarly, any sentence A, such that neither it nor its negation are true will be called a *truth gap*). I will also refer to the weaker position, that it is logically possible for a contradiction to be true, as *weak dialetheism*.

What is the substance to a case in which a sentence and its negation are both true? It depends on what the cases of the theory are. In general, that a sentence and its negation are true in a case does not commit one to dialetheism. If A and $\neg A$ are both true according to a fiction, this does not imply that there is a sentence B such that it and its negation *are* true. That someone asserts that A and that $\neg A$ does not mean that there are true contradictions — that there are dialetheists does not mean that they are right. (In this example all the sentences that some agent asserts is taken to be a case; the true sentences of the case are all and only the sentences asserted by the agent.) I will show that paraconsistent interpretational theories lead directly to dialetheism and that dialetheism is a consequence of some representational paraconsistent theories.

4. *Cases as Interpretations*

In this section I will show that paraconsistent consequence theories with only interpretations as cases result in dialetheism. I will do this using two different notions of interpretation. The first is modelled on an agent interpreting another's language into their own. This, at the very least, maps expressions for the language being interpreted to the language of the interpreter (I will

[4] In relevant logic circles this is to take up the Scottish plan, rather than the Australian or American plans; see [20, Chapter 7].

[5] Given the discussion above a comment on *which* conjunction (extensional or intensional) is being used in this formalisation of dialetheism is in order. I will not choose one over the other. What is required is that the conjunction in premise combination and in the definition of dialetheism is the same. This can lead to two different types of dialetheism (intensional and extensional) but this difference is beyond the scope of the paper.

call them the interpreted language and interpreting language respectively). The second approach follows Etchemendy's account of Tarski and models as interpretations (see [10] and [11]). In this approach, interpretations are functions from expressions of a language to *semantics values* which, in general, are not expressions of a further language. The important aspect of both these approaches is that it is only the interpretation of the interpreted language which changes from case to case. Differences in truth values must be accounted for by the choice of interpretation; non-linguistic facts must remain constant from case to case.

When someone engages in the project of interpreting a foreign language they produce a mapping of expressions from that language to their own where the expressions of the two languages are intended to mean the same thing. Of course, there is far more to interpretation than this minimal translation, but this will be sufficient for our purposes. I will call this type of interpretation a *translation* (not because interpretations of this type give no more than a translation, but because this will be the crucial feature for our purposes and it gives a means for distinguishing this brand of interpretation from the other explored in this paper).

By our hypothesis: in a paraconsistent consequence theory there is some case in which A and $\neg A$ are both true. If this case is a translation then there are expressions in our interpreting language where the translations of both A and $\neg A$ are both true. This, by itself, is not sufficient to drive us to dialetheism — for example, translate them as the English sentences "Mars is a planet" and "Venus is a planet" respectively. This changes when some desirable restraints are placed on translations. We restrict ourselves to translations where the logical constants (including negation) are translated faithfully into the interpreting language and the translation of a sentence is determined by the translation of its parts. This further restriction determines that in any translation in which A and $\neg A$ are true there is some sentence A' in the interpreting language such that it and its negation (using the negation of the interpreting language) are both true. Paraconsistent translational consequence theories (which respect composition and the meaning of negation) result in dialetheism.

There are two reasons for moving to interpretations which assign semantic values directly to expressions. The first is that there may be insufficient expressions in the interpreting language. The second reason is that translations produce more information than is required. All we need to know is which sentences are true. We do not need to know the translation of each sentence; it is sufficient to know whether their translations are true or not.

This is good reason to move from translations to interpretations. Each expression is assigned a value which it *could* have. It is important to understand the modality involved in the '*could*' above. This does not allow us to consider counterfactual situations, possible worlds, impossible worlds or the like. What is important is that the value is the right type for the category of the expression. The expression ϕ can be assigned the value α (that is, there is some interpretation in which ϕ is assigned α) if:

(1) either there is some expression of the same semantic type of the language which *is* assigned this value
(2) or there is an expression of the same semantic type that can be added to the language with this value.

This gives us one way of understanding what Tarskian models are. A model assigns an acceptable value to each expression of our language. This assignment is an abstraction of an interpretation of the language. It does not give us the full interpretation of the language, but it does give us the parts of the interpretation relevant to consequence (again, see [10] for further details).

By our hypothesis, in a paraconsistent interpretational theory there is some assignment of semantic values where both A and $\neg A$ are assigned *true*. As with the translation theory, this does not give rise to a true contradiction until we place restrictions on the interpretation of $\neg$ and ensure that the interpretations are compositional. Unlike the translation approach we still do not have a true contradiction. What we have is that it is possible, with minimal change (all of which is linguistic), to have a true contradiction by introducing new vocabulary. On this view, the world is already such that it would be accurately described by a contradiction, but it may be that this particular contradiction is not in any existing language. This is not quite dialetheism but it is close enough to undercut attempts to resist it.

5. *Cases as representations*

The alternative account of consequence given by Etchemendy in [10] and [11] uses all and only representations of logically possible ways that world could be as cases. This is clearly a well motivated characterisation of consequence. C is a consequence of Γ if and only if it is logically impossible for the premises to be true and the conclusion not. Representational characterisations need not concern themselves with motivating their position but do need to demonstrate that they are *substantive* characterisations.

If A and $\neg A$ are true in a representation of the way things are, it by no means follows that there are true contradictions. There are many inconsistent representations of the world — there are inconsistent fictions, pictures, belief sets and so on.

Consider the following argument:

(1) There is some representation in which $A \wedge \neg A$ is true (by hypothesis)
(2) In every representation $\neg(A \wedge \neg A)$ is true.
(3) From (2), in no representation $A \wedge \neg A$ is true.
(4) But (1) and (3) are contradictory.

The paraconsistentist's options (in order to avoid dialetheism) are to reject (2) or the move from (2) to (3). In general (2) is false and the move from (2) to (3) fails.

Which representations should be included in a theory of consequence? This type of theory was introduced as containing all and only the representations which represent the world in a logically possible way (for shorthand call them *logically possible representations*) should be included in a theory of consequence. In this case (1) and (2) imply that:

(1′) It is logically possible that $(A \wedge \neg A)$ is true.
(2′) It is logically necessary that $\neg(A \wedge \neg A)$ is true.

But then, by the duality of possibility and necessity it follows that:

(3′) It is not the case that it is logically possible that $(A \wedge \neg A)$ is true.

and, again, we get a contradiction.

There are five possible responses:

(A) Be a dialetheist.
(B) Reject paraconsistency.
(C) Reject (2′).
(D) Reject the move from (2′) to (3′).
(E) Use different cases.

I will not say anything more about options (A) and (B), as these options are to give up the game at hand. The last of these options will be briefly explored in the next section.

Option (C) requires that there is some case of the theory where $\neg(A \wedge \neg A)$ fails to be true. This gives the paraconsistentist who wants to avoid dialetheism some clear guidelines as to which logics are available to them. Depending on the properties of $\neg$, $\wedge$ and $\vee$ this approach may be committed to rejecting excluded middle (that $A \vee \neg A$ is a logical truth) and the law of noncontradiction (that $\neg(A \wedge \neg A)$ is a logical truth). Notice that this is only an option for the representationalist and not the interpretationalist. The interpretationalist who uses a paraconsistent logic which rejects the logical truth of $A \vee \neg A$ is committed to truth gaps as well as truth gluts. Note also that the representationalist trying this escape route is committed to the logical possibility of both true gaps and truth gluts and is thus committed to weak dialetheism.

As with option (C), option (D) also leaves one committed to weak dialetheism.[6] Moreover, the duality between logical possibility and logical necessity is well entrenched. This does not mean it cannot be rejected, but doing so will be costly.

6. *Other Cases*

Paraconsistentists can avoid dialetheism by using different cases. As I have already observed, there are many options for cases which would not (obviously) commit paraconsistentists to dialetheism. The challenge is to show that these consequence theories are as well motivated as interpretational and representational theories.

Paraconsistentists often argue that their logics are better than classical logic as they are able to deal with inconsistent fictions, belief sets, different inconsistent theories and the like (for example see: [17], [4] and [19, Part Four]). It seems possible to convert these applications into cases for a consequence relation. For example: take all fictions (or more likely, mathematical representations of systematic fictions) to be cases of the theory. This is an approach well worth pursuing. The main areas of concern are showing that these cases are relevant to consequence and providing a reasonable bound on the cases which includes inconsistent cases but does not commit one to the possibility of contradictions.

In [2], relevant logic is motivated as using situations as cases. Situations come from situation theory as developed by Barwise and Perry (eg [1]). A rough characterisation is that a situation is a *part* of a possible world. Their distinguishing feature is that they can be incomplete. A consequence relation can be given for a language using situations which do not decide the truth of every sentence of the language. Beall and Restall do not only use incomplete situations; they also make use of inconsistent situations. There are two ways of understanding inconsistent situations. Situations may still be parts of possible worlds. This approach, as with the representational theories, commits one to weak dialetheism (strong dialetheism is avoided as there are also truth gaps in these paraconsistent theories). Beall, as a dialetheist, can accept this but Restall, a recovering dialetheist, cannot. The alternative,

[6] At this point Priest's argument in [18] for dialetheism becomes difficult to put off.

and the approach of [2], is that inconsistent situations are parts of *impossible worlds*, or ways that the world could *not* be. This then requires similar motivation as discussed earlier in this section: why should these cases be part of a theory of consequence? Beall and Restall do motivate the use of inconsistent situations. As they are engaged in defending logical pluralism they do this by showing that this approach to consequence has a number of uses. The logical monist who wants to make use of this approach needs to do much more.

6.1. *Representations and Impossible Worlds*

We can combine the representational approach and the impossible worlds of the last section. A theory of consequence which includes representations of how the world cannot logically be can escape dialetheism. In this section, I will make some brief comments on impossible worlds and impossible representations. Impossible worlds are very useful in non-classical logics and should not be rejected out of hand. Impossible worlds can play an important role in providing a semantic theory from entailment connectives with desirable properties.

Consider the formula $A \to (B \to B)$: various non-classical logicians, including relevant logicians and paraconsistentist more generally, reject it as a logical truth. Relational semantics can provide counterexamples to them with the help of impossible worlds. Let us consider an example semantics for a propositional language with the connective $\to$ with models $\langle W, N, R, @, v\rangle$ where W is non-empty set, N is a subset of W, R is a binary relation on W, $@$ is a member of W and v assigns subsets of W to each atomic proposition. Intuitively, the members of N are the possible worlds and W without N are the impossible worlds. $C \to D$ is true at w iff $w \in N$ and for every accessible $u \in W$ such that Rwu where C is true, D is true as well. This provides a counterexample to $p \to (q \to q)$. Consider the two element model $\langle \{a, b\}, \{a\}, \{\langle a, b\rangle\}, a, v\rangle$ where $b \in v(p)$. In this model p is true at b but $q \to q$ is not. As a result $p \to (q \to q)$ is not true at the base world a. We can take this model to represent the way things are. The model represents the world as such that the sentences true at a are true. The model represents the world as such that $p \to (q \to q)$ is not true. This is done by making important use of the impossible world b within the model.[7]

Does this model represent the world as a way that it could be? Or does it represent the world as a way that it couldn't? That depends. Is it logically possible for $p \to (q \to q)$ not to be true? If it is logically impossible,

[7] I am not endorsing this particular semantics. This semantics is only serving as *one* example.

then the model represents the world in an impossible way. If it is logically possible for the sentence not to be true, then (depending on all the other aspects of its representation) it represents the world in a logically possible way. Importantly, this example does not use a model to represent the world as such that $q \rightarrow q$ is false. We can similarly make use of impossible worlds *within* representations to construct counterexamples to $A \rightarrow (\neg A \rightarrow B)$ without representing the world as such that A is true and that $\neg A$ is true. Is this a representation of a way the world could be? That turns on whether the world could be such that $A \rightarrow (\neg A \rightarrow B)$ is false, not on whether A and $\neg A$ could both be true as in the case of the consequence from $A, \neg A$ to B.

The relevantist rejecting these candidate logical truths, and presumably paraconsistentists more generally, will believe that it is possible for them to be false (in fact, they will take some instances to actually be false) and thus the model represents the world a way it could be. This use of impossible worlds does not yet provide the paraconsistentist an escape route. For impossible worlds to be of use to the representationalist they have to include representations of the world as it could not be as cases. This is to rescind on part of the motivation behind representational theories. Representational theories set consequence up as truth preservation across all and only logical possibilities. The *all* is easy to motivate. If it is logically possible for the premises of an argument to be true and the conclusion false, then the conclusion does not follow from the premises as a matter of logic. But what about the *only*?

There are clearly some representations of the world which must be excluded from any informative theory of consequence. Should a representational theory of consequence include any representations of the world in impossible ways? Allow me to approach this by analogy. Suppose we are developing a theory of physical consequence. I claim that it is a physical consequence of A that B. You respond by constructing a representation of a physical impossibility where A is true but B fails to be true. Should this trouble me? Suppose we are constructing a theory of metaphysical consequence. I claim that B is a metaphysical consequence of A. You object by constructing a representation of a metaphysical impossibility where A is true but B fails to be true. Should this trouble me? I say "No, it should not!" In each instance the counterexample is beyond the range of the modality involved. Similarly it seems that the way the world could not be should not be a counterexample to the validity of an argument. It follows from this that the representations should *only* be of logical possibilities.

This is, of course, far from a watertight argument. It does not show the impossibility of providing reasons for including impossible representations in a theory of consequence. All it aims at doing is highlighting the work the paraconsistentist is required to do. One reason that can be provided by the paraconsistentist is that impossible worlds have been demonstrated to do significant work in semantic theories for conditionals, as well as elsewhere. This pragmatic response is far from ideal. It would be much better to have some principled reason and preferably a principled reason stemming from the nature of logical consequence itself. Nothing in this paper rules out the use of impossible worlds as providing a semantics for conditionals like that above. Nothing in this paper rules out "in the story" logics; we are entitled to include impossible representations if we are interested in what follows when we systematically alter what counts as logically possible. We leave the paraconsistentist with the task of providing reasons for why the logically impossible plays a role in logical consequence.

If this challenge were met, my main conclusion would be strengthened. In this case we will have two different types of representational theories of consequence. On one approach theories contain all and only representations of logical possibilities and on the other theories contain all (but not necessarily only) representations of logical possibilities. One can have two extensionally equivalent, paraconsistent, theories of each sort. The first leads to dialetheism[8] but the second, if successfully developed, does not. We have extensionally identical, intensionally different theories leading to different ramifications. This is another example of my main thesis.

7. *Truth Gaps*

Similar results can be shown for other nonclassical theories of consequence. Consider logics where $A \vee \neg A$ is not a logical truth. In that case it fails to be true in some case. If the cases are interpretations, then, as before, there must *be* a truth gap — a sentence such that neither it nor its negation is true.

If cases are logically possible representations then it is far harder (we require more assumptions) to conclude either that there is a truth glut or gap. This highlights a potential asymmetry between truth gaps and gluts as well as between paraconsistent and paracomplete logics (see [14] for a description of the apparent symmetry).

8. *The Challenge*

The previous arguments lead to the following challenge for the paraconsistentist:

[8] Given the provisos from the earlier section on representational theories.

EITHER Accept dialetheism (or give up excluded middle, or the duality between logical possibility and necessity) OR provide another account of the cases in your theory of consequence.

This challenge should also encourage others to investigate the ramifications of their theories of consequence. In [10] Etchemendy puts forward a strong case, using similar arguments to those in this article, for the unacceptability of classical interpretational consequence theories — even if they are extensionally equivalent to acceptable representational theories.

I have shown that there are significant differences in what commitments are entailed by extensionally equivalent, but intensionally different, theories. Interpretational theories of paraconsistent logics commit one to dialetheism but the same is not true of representational theories. In representational theories additional commitments can lead the paraconsistentist to dialetheism, but there are theories for which this is not guaranteed. The representationalist is, nonetheless, committed to weak dialetheism.

School of Historical and Philosophical Studies
University of Melbourne
Australia
E-mail: conrad.asmus@gmail.com

REFERENCES

1. Jon Barwise and John Perry, *Situations and attitudes*, CSLI Publications, 1999.
2. J.C. Beall and Greg Restall, *Logical pluralism*, Oxford University Press, 2006.
3. J.C. Beall, Ross T. Brady, A.P. Hazen, Graham Priest, and Greg Restall, Relevant restricted quantification, *Journal of Philosophical Logic* (2006).
4. Nuel D. Belnap, A useful four-valued logic, In *Modern Uses of Multiple-Valued Logics* (J. Michael Dunn and George Epstein, eds.), Reidel, 1977, pp. 8–37.
5. Rudolf Carnap, *Introduction to semantics and formalazation of logic*, Hardvard University Press, 1943.
6. B.J. Copeland, On when a semantics is not a semantics: some reasons for disliking the Routley-Meyer semantics for relevance logic, *Journal of Philosophical Logic* 8 (1979), pp. 399–413.
7. B.J. Copeland, Pure semantics and applied semantics, *Topoi* 2 (1983), pp. 197–204.

8. Michael Dummett, The justification of deduction, In *Truth and other enigmas*, Harvard University Press, 1978, pp. 290–318.
9. Michael Dummett, *The logical basis of metaphysics*, Harvard University Press, 1991.
10. John Etchemendy, *The concept of logical consequence*, Harvard University Press, 1990.
11. John Etchemendy, Reflections on consequence, In *New Essays on Tarski and Philosophy* (Douglas Patterson, ed.), Oxford University Press, 2008.
12. Susan Haack, *Philosophy of logics*, Cambridge University Press, 1978.
13. Edwin Mares, *Relevant logic: A philosophical interpretation*, Cambridge University Press, 2004.
14. Chris Mortensen, *Inconsistent mathematics*, Kluwer Academic Publishers, 1995.
15. Alvin Plantinga, *The nature of necessity*, Clarendon Press, 1975.
16. Graham Priest, *In contradiction: A study of the transconsistent*, Martinus Nijhoff, 1987.
17. Graham Priest, Sylvan's box, *Notre Dame Journal of Formal Logic* 38 (1997), pp. 573–82.
18. Graham Priest, Motivations for paraconsistency: The slippery slope from classical logic to dialetheism, In *Frontiers of Paraconsistency* (Batens, Mortensen, Priest, and Van Bendegem, eds.), Kluwer Academic Publishers, 2000, pp. 223–232.
19. Graham Priest, Richard Sylvan, and Jean Norman (eds.), *Paraconsistent logic: Essays on the inconsistent*, Philosophia Verlag, 1989.
20. Stephen Read, *Relevant logic: A philosophical examination of inference*, Basil Blackwell, 1988.
21. Stephen Read, Monism: The one true logic, In *A Logical Approach to Philosophy*, Springer, 2006, pp. 193–209.
22. R. Routley and V. Routley, The semantics of first degree entailment, *Noûs* 6 (1972), pp. 335–359.
23. Richard Routley, *Exploring Meinong's jungle and beyond*, Philosophy Department, RSSS, Australian National University, 1980.