Paradoxes of Logical Equivalence and Identity

Andrew Bacon*

September 4, 2013

Assuming the unrestricted application of classical logic, the paradoxes of truth, sets and properties make trouble for naïve intersubstitutivity principles, such as the principle that allows one to substitute, in non-intentional contexts, the claim that ϕ for the claim that 'ϕ' is true, or the claim that x is F for the claim that x has the property of being F. One way to respond to these paradoxes is to reject the logical assumptions the paradoxes rest on, allowing one to instead accept naïve intersubstitutivity principles governing sets, truth and properties.

In this paper I show that even in these weakened logics intersubstitutivity principles can wreak havoc. In these discussions intersubstitutivity principles are normally formulated using relatively weak metalinguistic rules; the following paradoxes arise when a stronger intersubstitutivity axiom, formulated entirely in the object language, is assumed.

In section 1 I outline a few different applications of these paradoxes: truth theorists, for example, might want to endorse a principle stating that logically equivalent sentences are substitutable *salve veritate*. Property and set theorists might want to endorse a version of Leibniz's law. In section 2 I present two paradoxes that show that in either case they cannot endorse the principle in question, given background assumptions. The second of these paradoxes uses very little in the way of logical machinery, and thus applies to most logics developed to deal with the semantic and set theoretic paradoxes (see for example Bacon [1], Beall [3], Brady [5], Priest [14].) In the second half of section 2 I note that both paradoxes, when interpreted in terms of the notion of logical equivalence, are similar in spirit to recent versions of Curry's paradox that employ the notion of a valid argument (see, for example, [19] and [4].) I then show that the present paradoxes can be formulated so as not to depend on any distinctive structural rules and, as a result, they are problematic for recent approaches to the validity Curry paradox that relax the rule of structural contraction (Zardini [20], Priest [15], Murzi & Shapiro [12].)

1 Applications of the paradoxes

1.1 Substitutability of logical equivalents

In [7] Hartry Field presents us with a theory of truth that not only allows us to keep the T-schema in full generality, but also allows us to substitute ϕ for the claim that 'ϕ' is true. This further property follows from the fact that in Field's logic you can quite generally substitute logical equivalents for one another and that, given the T-schema, ϕ and the claim that 'ϕ' is true are logically equivalent. This fact about the logic, according to Field, represents a significant improvement over rival theories such as Priest's in [14], in which logically equivalent sentences are not intersubstitutable.

Two points require some clarification: we firstly need to know what it means for two sentences to be logically equivalent and we also need to know what it means for two sentences equivalent in this sense to be intersubstitutable.

*I'd like to thank Hartry Field for some extremely helpful comments on an earlier draft of this paper, and three referees for this journal for their many helpful suggestions and corrections.

I shall take, as a matter of terminology, two sentences A and B to be logically equivalent iff the biconditional $A \leftrightarrow B$ is a logical truth.[1] Other than connect the word 'logical equivalence' to 'logical truth' this tells us very little. Nonetheless, it is clear that in order to engage in the debate about whether logical equivalents are intersubstitutable you have to have some notion of 'logical equivalence' (and 'logical truth') in mind, and the following paradoxes provide limitative results on what that notion might be. That said, there are a few things one could mean by 'logical equivalence' that I expect will not serve the purposes of those interested in the notion. One might, for example, reserve the term 'logical truth' for a sentence provable from a purely quantificational logic – from principles governing the logical connectives and quantifiers alone. This is indeed compatible with the way many people following Quine use the word 'logic', but it would clearly not suffice for Field's purposes since the T-schema is not a logical truth in this sense, and without that one cannot assume the equivalence, and therefore intersubstitutivity, of S and 'S is true'.

More generally, the relation that holds between two sentences when they can be proved to be equivalent relative to some formal system (which possibly includes truth theoretic principles) is not general enough either. It is extremely natural to think that if two arithmetical sentences have the same truth value in the standard model of arithmetic, they should be intersubstitutable. However no definition of equivalence in terms of provability will ensure this due to the incompleteness theorems (one can always construct a Gödel sentence which ought to be intersubstitutable with $0 = 0$.)[2]

A better way to at least characterise the extension of the relation of logical equivalence would be to select a suitable class of models and define logical truth as truth in all models in that class.[3] Of course this helps us little unless we know which models are suitable, and this depends in turn on the prior conception of logical truth we are trying to characterise. Field himself prefers to call a sentence logically true only if we should (in a very objective sense) fully believe it, and expresses sympathy towards the converse of this claim (see §2(d).)[4] Provided one can make sense of this objective use of 'should', this interpretation appears to do a lot better in many regards. Since Field is our primary target at this juncture it will serve as a useful starting place. My strategy in the paper, however, will be simply to take the notion as a primitive; any results we can derive may be seen as constraints on what can be said about each of these ways of understanding equivalence.

The second point of clarification concerns what it means for two equivalent sentences to be intersubstitutable. It is indisputable that Field's logic has a rule that allows one to substitute logical equivalents within logical truths in a way that preserves logical truth.[5] However this rule is far too weak, allowing us only to substitute within sentences we can prove – it says nothing about substitutability within (say) contingent truths.

Field's logic also contains the rule $A, B \leftrightarrow C \vdash A[B/C]$ (which follows from the law

[1] It is important to distinguish this from a weaker notion of logical equivalence that obtains when A and B entail each other. Given modus ponens and conjunction elimination two sentences are equivalent in my sense only if they are equivalent in this sense. In logics without conditional proof, however, the converse does not hold.

[2] Moreover, given this incompleteness, the dependence on the choice of formal system can begin to seem arbitrary.

[3] These might be a class of many valued models described in a classical set theory, as Field does in [7]. Alternatively one could run the model theory within a non-classical theory of sets as suggested in [2]. In the former case one can expect 'S is logically true' to behave classically (i.e. it will conform to the laws of classical logic) whereas in the latter one cannot. The paradoxes I consider in what follows do not assume that these notions behave classically.

[4] See Field [8]. It should be noted that I am adopting Field's initial rough way of describing the view. The precise view is a bit more intricate, especially as it is designed to given a general account of logical consequence with multiple premises and multiple conclusions. The precise account of logical truth (the 0 premise 1 conclusion instance of this theory) requires that A is a logical truth (if and) only if ones credence in A must be 1 conditional on fully accepting C and fully rejecting D for any C and D. It's also worth noting that Field does not take claims about what we should believe to constitute an analysis of validity, although for my purposes material equivalence is sufficient.

[5] And more generally, it allows us to substitute logically equivalent sentences in valid arguments in such a way as to preserve validity.

$B \leftrightarrow C \vdash A \leftrightarrow A[B/C]$) which allows one to substitute B for C when one already knows that B and C are materially equivalent. This allows substitutability of logical equivalents within contingent and other non-logical truths to a limited extent, but I think it is still not enough. If logical equivalents truly are substitutable we ought to be able to say, even when one does not know whether B and C are logically equivalent, that *if* they are, and A is true, then $A[B/C]$ is.

The crucial shortcoming of these principles is that they take the equivalence of two sentences to be the premise of *rule* rather than the antecedent of a conditional. The first rule, for example, guarantees that anyone who is in a position to know the premises, A and $B \leftrightarrow C$, is in a position to know $A[B/C]$ but says nothing of someone who is not in a position to know the premises. Yet surely if logical equivalents really are substitutable, someone who knows that A is true ought to be able to infer that every substitution of A with a logical equivalent is true, no matter what they know about what's logically equivalent to what. That is to say that someone who knows that A is true ought to be able to know that $A[B/C]$ is true if B and C are equivalent, and they ought to be able to say this even if they're unsure about whether B and C are equivalent, suggesting the rule:

> (SSV*) From the fact that A is true infer that if B is logically equivalent to C then $A[B/C]$ is true.

$A[B/C]$ is just the sentence you get by substituting C everywhere for B in A. Special issues arise when C appears embedded under attitude verbs within A so I shall restrict myself to sentences, A, in which C does not appear embedded in this way. This is, of course, a weakened version of the principle that logical equivalents are substitutable *salve veritate*:

> (SSV) If A is true and B is logically equivalent to C then $A[B/C]$ is true

This principle says that substituting something for a logical equivalent does not change the truth value of the sentence it occurs in. In Field's logic SSV is in fact stronger than SSV* – I shall only ever need the weaker principle.

A natural reaction to the following paradoxes might be to deny that we ever needed the intersubstitutivity of logical equivalents in the first place. Perhaps all one needs is something a bit Field's rule: from A is true and the claim that B and C are logically equivalent infer that $A[B/C]$ is true. I would disagree. One reason we need the stronger principle SSV* is that Field's logic is highly non-recursive. Even if you are excellent at deductions, and have infinite patience, there will be many cases of logically equivalent sentences whose equivalence will be impossible to determine. If one, as a theorist, wants to assert that any substitution of some accepted principle with a logical equivalent is true they will not be in a position to do this on the basis of the weaker rule alone, unless they have determined, for every pair of sentences, whether they are equivalent. In short, one is simply not entitled to this metatheoretic assertion when you are ignorant about what is equivelent to what.

Moreover, if there are genuine paradoxes involving logical equivalence statements there could be cases where we could never know whether B and C are logically equivalent because it is indeterminate whether they are equivalent. In these cases the weaker rule will again be of little use since we are not in a position to know the relevant premises – yet if you know A is true it shouldn't be left open whether $A[B/C]$ is true if B and C are equivalent, even when it's impossible to determine whether B and C are equivalent.

Let Tr represent the truth predicate and let E represent a binary relation stating that two sentences are logically equivalent. A natural logic, then, might take the weakened substitution rule SSV* and combine it with two further natural principles governing logical equivalence: that every thing is equivalent to itself and the rule that from a proof of $A \leftrightarrow B$ allows you to infer that A and B are equivalent.

LL(E) $Tr(\ulcorner A \urcorner) \vdash E(\ulcorner B \urcorner, \ulcorner C \urcorner) \rightarrow Tr(\ulcorner A[B/C] \urcorner)$

I(E) $E(\ulcorner A \urcorner, \ulcorner A \urcorner)$

RE(E) If $\vdash A \leftrightarrow B$ then $\vdash E(\ulcorner A \urcorner, \ulcorner B \urcorner)$.

Here RE(E) allows one to prove that A and B are logically equivalent given a proof of $A \leftrightarrow B$ from logical principles that may include LL(E), I(E) and even RE(E) itself. The following paradoxes pose problems for this combination of principles given various background logical and truth theoretic assumptions. Note that RE(E) should be understood so as to be applicable to proofs from these background assumptions (it will always be clear from context which principles are being considered to be under the scope of RE(E)'s application.)

Field's own remarks on similar paradoxes involving notions such as validity and logical truth provide us with one possible avenue for evading these paradoxes whilst keeping the intersubstitutivity of logical equivalents (see the discussion in section 20.5 in [7].) The idea is to accept principles like LL(E) and I(E) but treat them as *non-logical* truths. According to this view the blame falls on principles like RE(E). Whilst you may indeed be able to infer that two claims are logically equivalent given a proof of their equivalence from only logical assumptions, you cannot in general infer this if the proof depends on non-logical assumptions. Since LL(E) and I(E) are non-logical on this view one cannot apply RE(E) when LL(E) and I(E) have been used earlier in a proof. On this account, then, there is some way of interpreting 'logical equivalence' that allows us to assert that logical equivalents are intersubstitutable – more precisely, some reading of E which permits true readings of LL(E) and I(E) – but according to this reading, neither LL(E) nor I(E) are logical truths in the same sense.

This response is good as far as it goes, but it does leave one wondering about the status of truths like LL(E) and I(E) on this reading. It does not seem as though these very general principles concerning any of the candidate notions of logical equivalence are empirical claims that can be discovered by investigation, scientific or otherwise. If they are true at all, they are presumably discoverable *a priori*. Similarly they do not seem to be contingently true either; for example, what would the world have to be like for a sentence not to be equivalent to itself? These observations are also suggestive for Field's own account of validity. Given that they are both truths, non-empirical truths at that, it is natural to think that one ought, in the relevant objective sense, to fully believe LL(E) and I(E) whatever our evidence is.[6]

At any rate, when $A \leftrightarrow B$ is provable from necessary *a priori* assumptions, it follows that $A \leftrightarrow B$ is necessary and *a priori*. Thus, whatever status $A \leftrightarrow B$ must have in order for A and B to be counted logically equivalent and intersubstitutable for one another, it must be more demanding than necessary *a priori* truth; if it were as demanding or less demanding RE(E) would be acceptable. On the other hand it cannot be too demanding. Being a theorem of a pure quantificational logic is more demanding than being necessary, *a priori*, but as we have noted already this is too demanding for Field's purposes since it doesn't count S and 'S is true' as logically equivalent.

One possible model for this more demanding kind of equivalence is a highly hyperintensional one: a biconditional has the status that suffices for the logical equivalence of its arguments (in the sense that permits true readings of LL(E) and I(E)) only if the sentences flanking both sides of the biconditional are literally identical. This is surely too demanding too: it does not permit one to substitute A for $A \wedge A$. Thus, presumably, if LL(E) and I(E) are true on any plausible candidate interpretation for E it will have to be a notion that is more demanding that necessary *a priori* equivalence and less demanding than strict identity. So while Field's response can certainly be modified to evade these paradoxes, it raises questions of its own: in what sense must we read 'logical equivalence' for the substitutivity of logical equivalents to come out as a true principle?

[6] The relevant sense of 'ought' presumably is not so objective as to not depend on our evidence, but this is not at issue here since the two principles in question seem to be *a priori*.

1.2 Naïve property and set theory

Informally I shall call a theory of properties a 'naïve' property theory if it permits the substitution of sentences of the form 't has the property of being an x such that ϕ' with sentences of the form $\phi[t/x]$. A naïve property theory can be strengthened to a set theory by including some form of the principle of extensionality. The ensuing paradoxes at no point assume extensionality so everything I say about property theory also applies to set theories.

Formally we can represent the property of being such that ϕ with the term forming subnective $\langle x : \phi\rangle$, where ϕ may or may not contain x free (more complicated theories which allow for relations can be considered but are not needed for the following paradoxes.) In order to state when x instantiates y and when x is identical to y we introduce the relations $x\epsilon y$ and $x\doteq y$. The logical principles that drive this version of the paradox are:

LL($\doteq$) $A \vdash t\doteq s \to A[t/s]$

I($\doteq$) $t\doteq t$

RE($\doteq$) If $\vdash A \leftrightarrow B$ then $\vdash \langle x : A\rangle\doteq\langle x : B\rangle$

The principle of self-identity, I($\doteq$), should be self explanatory.[7] The rule RE($\doteq$) is subject to the same caveats we discussed in section 1.1, however it far harder to deny I($\doteq$) and LL($\doteq$) the status of logical truth in this case.

The first of these principles is a weakening of Leibniz's law. The standard version of Leibniz's law is formulated as an axiom rather than a rule: $t\doteq s \to (A \to A[t/s])$ or $A \to (t\doteq s \to A[t/s])$. Without making assumptions about the conditional we cannot assume that these two axioms are equivalent.[8]

The principle we are considering is a rule and not an axiom. It is fairly trivially weaker than the second formulation of the axiom (assuming only modus ponens), and given reasonable (although not indisputable) assumptions is also a weakening of the first formulation.[9] One could in principle block the arguments by accepting the former formulation of Leibniz's law and not the latter (provided one also rejects the logic that allows one to show they are equivalent.) However it is quite hard to philosophically justify one without the other, and conversely, hard to provide a principled philosophical reason to reject one of these formulations that doesn't extend to the other.

It is not hard to produce algebraic models in which LL($\doteq$) holds when A is atomic, but fails when A is a complex formula, even in a relatively strong logic like Łukasiewicz logic.[10] One might take the existence of such models as evidence against the unrestricted version of LL($\doteq$), and the axiom versions that it appears to follow from.

One thing these models do is establish that it is literally possible to adopt a combination of attitudes in which you (i) accept all the Łukaisewicz-consequences of things you accept, (ii) accept that a is F whilst (iii) rejecting (and therefore not accepting) the claim that b is F if a and b are identical. One simply accepts the sentences that get value 1 in these models and rejects everything else.

While this formal point is surely unassailable, it doesn't address the crucial question: is it *coherent* to accept that a is F whilst rejecting the claim that b is F if a and b are identical. To illustrate, compare it to the question of whether it is coherent to accept that

[7] Although not necessarily uncontentious. Some reject I($\doteq$) when the arguments are non-denoting terms.

[8] Indeed there are many different non-equivalent ways of stating Leibniz's law in non-classical logics; see for example the discussion in Priest [13], sections 24.6 and 24.7.

[9] The assumption in question is the rule $A \to (B \to C), B \vdash (A \to C)$.

[10] I am grateful to Hartry Field and an anonymous referee for drawing my attention to this. In Łukasiewicz's three valued logic we can ensure that instances of LL($\doteq$) where A is atomic come out true if we stipulate that for every atomic predicate F, the value of Fa and Fb differs by no more than 1 minus the value of $a = b$. However, if the value of $a = b$ is a half, Fa one and Fb a half then $\neg(Fa \to \neg Fa)$ will have value one but $a = b \to \neg(Fb \to \neg Fb)$ will have value a half. See also the discussion of Leibniz's law in relevant logics in 24.6 and 24.7 of Priest [13].

Fred is an unmarried man whilst denying he's a bachelor. Pointing out that there's a model in which 'bachelor' and 'unmarried man' have different extensions doesn't really address the complaint that the combination of attitudes ascribed seem incoherent when taken at face value.

On the other hand, it's not clear that we can afford to take these model more seriously than a consistency proof. The domain of any classical set theoretic model, for example, will consist only of objects that are determinately distinct from one another, whereas the putative counterexamples to Leibniz's law will necessarily involve indeterminate identities. While models described within classical set theory can often provide insight into non-classical ways of thinking, it would be unwise to take them as more than a helpful heuristic. At the end of the day we must evaluate a claim by what it actually states, and in this regard LL($\doteq$) is difficult to deny. It is just too hard to see how one could coherently endorse the claim that a is F whilst rejecting the claim that b is F if a and b are identical (and the appearance of incoherence in no way depends on whether F denotes an atomic predicate or not.[11])

1.3 Propositional identity

The paradoxes we consider can also be generated if we wish to introduce a propositional identity connective, $=$, into the language (see, for example, Cresswell [6].) Formally analogous principles can be formulated for this connective:

LL $A \vdash B = C \rightarrow A[B/C]$

I $A = A$

RE If $\vdash A \leftrightarrow B$ then $\vdash A = B$

Visually the proofs are more pleasing if we adopt these axioms and take the connective $A = B$ as a primitive. In a setting in which the vocabulary of section 1.1 or 1.2 is taken as primitive one can make analogous arguments by making the following substitutions.

> In the first case $A = B$ can be replaced by $E(\ulcorner A \urcorner, \ulcorner B \urcorner)$. If we have a validity predicate, V, for stating when a sentence is valid in the language we can also replace this with: $V(\ulcorner A \leftrightarrow B \urcorner)$. In some logics substitutivity of A and B holds only if both they and their negations are logical equivalents (see [1]); in these cases one can formulate similar paradoxes by adopting a different definition of equivalence: $V(\ulcorner (A \leftrightarrow B) \wedge (\neg A \leftrightarrow \neg B) \urcorner)$.)[12]
>
> In the second case $A = B$ can be replaced by $\langle x : A \rangle \doteq \langle x : B \rangle$ where $\doteq$ is the ordinary binary relation of identity and x does not appear free in A or B.

In effect, then, we have three different paradoxes depending on how we interpret the $=$ sign. The first kind of interpretation is only available in contexts in which we have names for each sentence of the language (when we are discussing truth theories, for example, but not set theories.) In these cases the paradoxes formulated using connectives should be reformulable using relations or predicates applying to sentences, although the details become a bit more fiddly. It is worth noting, however, that in general the distinction between operators and predicates has little logical impact in naïve truth theories of the

[11] If we define a determinacy operator as $A \wedge \neg(A \rightarrow \neg A)$ and take Field's three-valued model to guide us in what to assert we get even more bizarre commitments. For example, one would have to assert that a is F, b is not F and that it's not determinate that a is distinct from b.

[12] In some of the theories we will discuss one can a putatively stronger notion of logical equivalence using a so called fusion connective, $\circ$: $V(\ulcorner ((A \rightarrow B) \circ (B \rightarrow A)) \urcorner$. In the following arguments substituting this notion weakens the premises even further.

form I am considering (when ϕ and $Tr(\ulcorner\phi\urcorner)$ are intersubstitutable then, given a connective $C(A_0, \ldots, A_n)$, one can always define an equivalent relation between sentences by the formula $C(Tr(\ulcorner A_0\urcorner), \ldots, Tr(\ulcorner A_n\urcorner))$.)[13]

With the first definition in place one can prove LL, I and RE from LL(E), I(E) and RE(E) and the intersubstitutivity of ϕ with $Tr(\ulcorner\phi\urcorner)$. With the second definition in place one can prove LL, I and RE from LL($\doteq$), I($\doteq$), RE($\doteq$) and the intersubstitutivity of ϕ with $x\epsilon\langle x : \phi\rangle$.

2 Two paradoxes

The theories I will discuss can be formulated in the propositional language $\mathcal{L}$ whose logical connectives are given by the set $\{\rightarrow, \wedge, \bot\}$, a propositional identity connective, $=$, and which contains, for each formula of the language, ϕ, a propositional constant A governed by the following pair of axioms

FP $A \rightarrow \phi[A/B], \phi[A/B] \rightarrow A$

Here B can be any propositional letter (possibly occurring in ϕ), $\phi[A/B]$ the result of substituting B everywhere for A. Principles with the same logical form arise in the context both of naïve property/set theory and naïve truth theories.[14] Formulating things this way, however, allows us to abstract away from the details of the specific device of self-reference and allows us to formulate the paradoxes in a setting of pure propositional logic.

The second paradox we shall consider will rely on the following three principles, mentioned above:

LL $A \vdash B = C \rightarrow A[B/C]$

I $A = A$

RE If $\vdash A \leftrightarrow B$ then $\vdash A = B$

In the framework I have outlined one can also introduce a notion of logical necessity, which we may formally define as follows.[15]

$$\Box A := A = (A = A)$$

With this definition in place one can state a rule of necessitation that is weaker than RE (the rule of equivalence) which plays an important role in the first paradox.

RN If $\vdash A$ then $\vdash \Box A$

Given the rule of equivalence and the identity axiom one can prove RN with some natural background logic.[16] RN, however, is strictly weaker than RE.

2.1 A Warm Up Paradox

The paradoxes in this section are formulated within a Hilbert style formalism in which all of the structural rules are being assumed. In everything that follows I shall assume the rule of modus ponens. The first paradox appeals to LL, I and RN. In addition to these we shall

[13] I should say, however, that I only do this to simplify the discussion. Notions applying to sentences, such as logical equivalence, are clearly quite different from the notion of propositional identity expressed using a connective, for example.

[14] For example, if we set A to be the formula $Tr(S)$ where S is a name for the sentence $\phi[Tr(S)/B]$ then the T-schema gives us that $Tr(S) \leftrightarrow \phi[Tr(S)/B]$ as required.)

[15] Alternatively one could take the notion of logical necessity as primitive and define a notion of logical equivalence as $\Box((A \rightarrow B) \wedge (B \rightarrow A))$.

[16] Suppose you can prove A, so $\vdash A$, and by I, you also have $\vdash A = A$. All one needs then is enough conditional logic to infer that $A \leftrightarrow (A = A)$, from which one can infer $A = (A = A)$ by RE.

need two logical principles. The first is a principle of transitivity for the conditional. The other is the rule of assertion, RA, which is slightly more controversial.

While this rule does not appear to be responsible for any of the standard semantic and set theoretic paradoxes (and indeed there are consistent naïve truth and set theories that contain the principle, see Grisin [10]) it is not validated in some recent theories (see e.g. Bacon [1], Beall [3], Brady [5], Field [7], Priest [14].) However the fact that these theories reject the principle is a fairly significant limitation of this result – we will show how to drop RA in the next section where we present a much more general paradox.

TR $A \to B, B \to C \vdash A \to C$

RA $A \vdash (A \to B) \to B$

LL $A \vdash (B = C) \to A[B/C]$

I $A = A$

RN If $\vdash A$ then $\vdash \Box A$

Observation: applying RN to I shows that $\Box(A = A)$ is a theorem. The proof of triviality proceeds as follows.

1. $(\Box C \to \bot) \to C$ instance of FP.
2. $C \to (\Box C \to \bot)$ instance of FP.
3. $\Box C \to (C = C \to (\Box(C = C) \to \bot)$ from 2 by LL and definition of $\Box$.
4. $(C = C \to (\Box(C = C) \to \bot)) \to (\Box(C = C) \to \bot)$ by RA and I.
5. $\Box C \to (\Box(C = C) \to \bot)$ 3, 4 and transitivity
6. $(\Box(C = C) \to \bot) \to \bot$ RA and observation
7. $\Box C \to \bot$ 5, 6, transitivity
8. C by 1
9. $\Box C$ by necessitation.
10. $\bot$.

2.2 The Main Paradox

The most significant weakness in the previous argument was the use of the rule of assertion. The next argument dispenses with RA, but uses the rule of equivalence, which in some logics (e.g. [7] and [1]) is strictly stronger than the rule of necessitation. In this argument we also have to assume a standard axiom governing the falsum constant.[17] The logical assumptions, TR and F, are accepted by pretty much everyone engaging in this debate (e.g. Bacon [1], Beall [3], Brady [5], Field [7], Priest [14].)

TR $A \to B, B \to C \vdash A \to C$

LL $A \vdash (B = C) \to A[B/C]$

I $A = A$

F $\bot \to A$

[17] This axiom assumes that there *is* a such constant. One can define such a thing in the truth and property theories considered provided you have an axiom of universal instantion (for example, in a naïve truth theory you can achieve this by identifying $\bot$ with $\forall x Tr(x)$.)

RE If $\vdash A \leftrightarrow B$ then $\vdash A = B$.

As before we assume modus ponens and FP. In this case we will also appeal to conjunction introduction (although the step at which it is used can be eliminated by appealing to the rule (a variant of RE): 'if $\vdash A \to B$ and $\vdash B \to A$ then $\vdash A = B$'.)

1. $C \leftrightarrow (C = \bot)$ from FP and conjunction introduction.
2. $C = (C = \bot)$ by RE.
3. $(C = \bot) \to ((\bot = \bot) = \bot)$ by LL and 2.
4. $((\bot = \bot) = \bot) \to \bot$ by LL and I
5. $(C = \bot) \to \bot$ by TR on 3 and 4
6. $C \to \bot$ by 1 5 and TR.
7. $\bot \to C$ axiom
8. $C = \bot$ form 5 and 6 by RE.
9. C by 1
10. $\bot$ by 6, 9 and modus ponens.

2.3 The paradoxes of validity

The above paradoxes arise when certain notions, such as logical equivalence and propositional identity, are expressible in the object language. Recently a number of paradoxes have been discussed that involve the related notion of logical entailment (see Whittle [19], Beall & Murzi [4].)

The preceding paradoxes differ from the paradoxes of validity in a couple of respects. The first difference is that they require less expressive resources. Given a predicate expressing the validity of an argument from A to B, $V(\ulcorner A \urcorner, \ulcorner B \urcorner)$, one can express the logical equivalence of A and B with the formula $V(\ulcorner \top \urcorner, \ulcorner A \leftrightarrow B \urcorner)$ – i.e. by saying that the argument from a tautology to $A \leftrightarrow B$ is valid. On the other hand, however, it is not possible, given a predicate expressing logical equivalence $E(\cdot, \cdot)$, to express the the fact that A entails B. We can certainly define the notion of a sentence being logically valid (i.e. being the conclusion of a logically valid argument with no premises) by defining a predicate, $L(\ulcorner A \urcorner)$, with the formula $E(\ulcorner A \urcorner, \ulcorner \top \urcorner)$ (contrast this with our earlier definition of $\Box$ from $=$.) But in logics in which conditional proof is not a permissible form of inference this is not sufficient for us to recover the notion of logical entailment. Saying that $A \to B$ is valid is not the same as saying that A entails B; there can be cases where A entails B but $A \to B$ is not valid.

The other sense in which these paradoxes differ from the validity paradoxes is, of course, that they make use of different assumptions. Our paradoxes make essential use of LL, which on this interpretation represents the substitution of logical equivalents *salve veritate*. The paradoxes of validity are, in effect, just versions of Curry's paradox. One must therefore assume the analogue of principles that suffice for deriving Curry's paradox. So, for example, the pair of principles below would suffice. Here I used $A \Rightarrow B$ to mean that A entails B[18]

CP If $A \vdash B$ then $\vdash A \Rightarrow B$

MP $A, A \Rightarrow B \vdash B$

[18] As before I present these arguments with a connective, $A \Rightarrow B$, for expressing the fact that A entails B, rather than a predicate $V(\cdot, \cdot)$ for ease of reading. As mentioned before, the differences are insubstantial when there is a predicate, Tr, such that A and $Tr(\ulcorner A \urcorner)$ are intersubstitutable.

as would MP plus the following four principles.

PMP $(A \wedge (A \Rightarrow B)) \Rightarrow B$.

CI $A \Rightarrow B, A \Rightarrow C \vdash A \Rightarrow (B \wedge C)$.

I $A \Rightarrow A$.

TR $A \Rightarrow B, B \Rightarrow C \vdash A \Rightarrow C$.

For example in the latter case one can begin with the sentence $C \leftrightarrow (C \rightarrow \bot)$ (by FP) and $C \rightarrow C$ (by I) to infer $C \rightarrow (C \wedge (C \rightarrow \bot)$. By PMP we have $(C \wedge (C \rightarrow \bot)) \rightarrow \bot$. So by TR we have $C \rightarrow \bot$, from which we could infer C and finally $\bot$.

MP informally states the self-evident fact that we can validly move from A and the fact that A entails B to B. PMP, on the other hand, is just a formalisation of the preceding sentence. I says that the inference from A to A is valid, TR encodes the idea that the consequence relation is transitive and CI is just a formalisation of a version of the principle of conjunction introduction (provided A entails B and A entails C, A entails $B \wedge C$.)

This latter argument, a variant of Whittle's, is of particular interest as it can easily be formulated so as not to use any distinctive structural rules. Despite this, the version of the paradox that uses MP, CP in addition to the structural rule of contraction, has received the most attention recently, and has prompted many to consider relinquishing structural contraction (an issue we will treat in more detail in the next section.) Existing proposals along these lines, however, have given up conjunction introduction (Zardini [20]) or modus ponens (Priest this volume [15]).[19]) The above argument demonstrates that this dilemma is inevitable, given the other background assumptions.

Interestingly these responses typically retain the rule of conditional proof, and its zero premise version (which we've already encountered and dubbed RN, provided one defines $\Box A$ as $\top \Rightarrow A$.)

CP If $A \vdash B$ then $\vdash A \rightarrow B$

RN If $\vdash A$ then $\vdash \Box A$.

The former (and thus presumably the latter) is retained in many of these recent proposals.

I think it is far from obvious whether or not we should accept CP and RN. An important issue, one I have not addressed adequately yet, concerns how to think of rules such as RN, CP and RE. Let us focus on the simpler zero premise version, RN. My remarks should extend to CP and RE.

Although, strictly speaking RN does not commit us to this alone, one might think that RN preserves validity. To say that the rule RN preserves validity is just to say that if A is valid then the claim that $\Box A$ is valid. This principle can in fact be formulated in the object language, since we are assuming that $\Box$ provides us with the means to express validity. It is therefore just the principle that if A is valid then the claim that A is valid is valid, $\Box A \rightarrow \Box\Box A$, that is characteristic of the modal system $\mathsf{S4}$. Thus while RN initially looks like it might be guaranteed to be true in virtue of the fact that $\Box$ expresses validity, it is actually somewhat controversial. The $\mathsf{S4}$ principle, for example, plays a crucial role in the following paradox employing the notion of necessity.

Premises:

TR $A \rightarrow B, B \rightarrow C \vdash A \rightarrow C$

EX$\Box$ $\Box A, \neg\Box A \vdash B$

[19]In Priest's set-up things are complicated by the fact that he has two conjunction symbols. According to one of these PMP is valid and according to the other CI is, however neither makes both principles true at once.

CO $A \to B \vdash \neg B \to \neg A$

T $\Box A \to A$

$\mathsf{4}$ $\Box A \to \Box\Box A$

R$\Box$ If $\vdash A$ then $\vdash \Box A$

R$\neg\Box$ If $\vdash A \to \Box A$ and $\vdash A \to \neg\Box A$ then $\vdash \neg\Box A$.

These principles are mostly unremarkable. The two most notable principles are $\mathsf{4}$, and the rule R$\neg\Box$, which states that when you can prove that $A \to \Box A$ and $A \to \neg\Box A$ then one can infer that A is not logically valid. The intuition behind the latter principle is that if a sentence implies contradictory claims then one ought to be able to infer that it is not logically true. Indeed the principle is weaker even than that: one need only establish that the sentence implies contradictory claims *about whether it is itself a logical truth* to infer that it is, in fact, not a logical truth.

One might think that paradoxes involving this principle, and the restricted principle of explosion EX$\Box$, should be of no concern to a paraconsistent logician who would reject certain generalisations of these principles for reasons to do with the liar paradox. However both principles trade on the rejection of contradictions concerning what is logically true – the idea that one can accept contradictions concerning what is logically true certainly takes us beyond the standard paraconsistent line on the semantic paradoxes.

At any rate, for Field and others both are extremely natural rules to endorse. The following therefore demonstrates a limitation of the approach of Zardini [20], who is committed to the $\mathsf{S4}$ axiom modulo definitions, and the other principles apart from R$\neg\Box$. It is a strike against this proposal, I suggest, that it cannot accomodate R$\neg\Box$. The proof then proceeds:

1. $A \to \neg\Box A$, $\neg\Box A \to A$ (fixed point of $\neg\Box$.)
2. $\Box A \to A$ by T
3. $A \to \neg\Box A$ by 1
4. $\neg\Box A \to \neg\Box\Box A$ T and contraposition
5. $\Box A \to \neg\Box\Box A$ by 2,3 and 4
6. $\Box A \to \Box\Box A$ by $\mathsf{4}$
7. $\neg\Box\Box A$ by R$\neg\Box$
8. $\neg\Box\Box A \to \neg\Box A$ by contraposition on $\mathsf{4}$
9. $\neg\Box A$ by 7 and 8.
10. A by 1
11. $\Box A$ R$\Box$
12. $\perp$ by EX$\Box$.

This paradox arises from treating RN as a rule that preserves validity. Before we move on it should be noted that the rule of necessitation as it is stated and used does not strictly commit us to the $\mathsf{S4}$ principle – formally or informally. Technically speaking RN is a rule of proof – it says, informally, that if we can prove that A then we can prove that A is valid. Due to the incompleteness of various formal logics with respect to validity we cannot move from the fact that a rule preserves provability to the fact that it preserves validity.[20] The rule of necessitation, as we have used it in proofs, is thus far weaker; it merely states that provable sentences should be provably valid.[21] If there can be cases of indeterminately valid sentences, due to the validity paradoxes perhaps, they must surely not be among the valid sentences which are provable.[22] Whether something is provable or not is always a clear cut matter, so if the system is half decent it will not clearly prove any sentence if it is unclear (and thus not valid) that it is valid – a good system should only prove sentences which are valid, validly valid, and so on.

2.4 Going substructural

When talking about arguments in the above setting I have assumed a standard formalism in which the premises of an argument are given by a set of sentences. This implicitly commits us to certain structural rules, most notably the rule of structural contraction:

SC If $A, A \vdash B$ then $A \vdash B$

which is guaranteed by the fact that $\{A\} = \{A, A\}$. SC bears a striking resemblance to the following rule of contraction for the conditional:

RC $A \to (A \to B) \vdash A \to B$

Logics that contain FP, modus ponens and the above rule of contraction are well known to be trivial (they prove every sentence) – a fact that is shown by Curry's paradox: 'if this sentence is true then A', where A can be any sentence. As mentioned already, many authors have exploited similarities between the consequence relation and the conditional, like the similarity above, to argue that analogous paradoxes arise for the validity predicate.

The standard response to Curry's paradox is to relinquish the rule of contraction. Several authors, particularly those mentioned in section 2.2, have argued that the correct way to respond to the validity paradoxes is, by analogy, to give up the rule of structural contraction (see for example Zardini [20], Priest (this volume), Murzi and Shapiro [12].)

It is of particular interest, then, to ascertain whether the present paradoxes involving logical equivalence and related notions can be generated once various structural rules have been relaxed. Here I answer the question in the negative: in the following argument the only characteristically structural rule is a simplified version of cut, SCut, which plays roughly the same role as TR does in the original paradoxes.[23]

In order to reason substructurally we cannot treat valid sequents as relating sets of sentences to conclusions, for otherwise structural contraction, and other rules, would be validated automatically. A sequent $\Gamma \vdash A$, therefore, consists of a sequence (not a set) of premises, Γ, and a conclusion formula A. I shall use commas to separate the arguments of theses sequences and a blank space to represent the empty sequence. A sequent calculus is

[20] See, for example, the incompleteness of axiomatic systems of second order logic with respect to the semantic notion of validity for those languages. More to the point: the concept of validity which Field endorses in [7] is highly non-recursive (see [18]) and so has no complete axiomatisation. Note that to infer that a rule preserves validity from the fact that it preserves provability requires both soundess and completeness (whereas to infer that a rule or principle is valid from the fact that it is provable one only needs soundeness.)

[21] By 'provable' I mean 'provable from the axioms and rules of the background logic *and* the rule RN.' The principle RN is thus impredicative and allows us to prefix arbitrary strings of $\Box$'s to theorems. In fact the proofs we present only apply RN once or twice in a give proof, so this aspect of the strength RN is not really the issue.

[22] The 'paradoxes of provability' can be represented by completely determinate (albeit unprovable) facts about the natural numbers, as Gödel has showed, but it is not obvious that the paradoxes of validity need to be completely determinate (see for example Schiffer [17].) Indeterminate validities arise even in the context of model theoretic accounts of validity if the model theory is carried out in a non-classical metatheory (see my [2].)

[23] Thus, of course, this argument does not apply directly to those who respond to the semantic paradoxes by denying the transitivity of entailment (such as Ripley [16] and Weir (this volume)).

a list of rules for deriving sequents from other sequents. I shall run a version of the second paradox in a relatively weak sequent calculus consisting of only the following rules:

$$\frac{A \vdash B, B \vdash A}{\vdash A = B} \ (RE') \qquad \frac{\vdash A}{B = C \vdash A[B/C]} \ (Sub)$$

$$\frac{}{\vdash A = A} \ (Id) \qquad \frac{A \vdash B, B \vdash C}{A \vdash C} \ (SCut) \qquad \frac{}{\perp \vdash A} \ (Bot) \tag{1}$$

In place of FP I shall just help myself to the sequents $C \vdash C = \perp$ and $C = \perp \vdash C$, where C denotes a fixed point for the formula $X = \perp$. The following argument is completely analogous to the paradox presented in §2.1 (proof on separate page.)[24]

[24] A referee has pointed out to me that these proofs rely on the fact that we can apply Sub to make substitutions of the same formula within the scope of = at different depths. If you restricted Sub to permit only substitutions of the same depth, then one would need to use it twice and apply contraction.

$$
\dfrac{
\dfrac{
C \vdash C = \bot
\qquad
\dfrac{
\dfrac{
\dfrac{
\dfrac{C \vdash C = \bot \qquad C = \bot \vdash C}{\vdash (C = \bot) = C}
}{C = \bot \vdash (\bot = \bot) = \bot}
\qquad
\dfrac{\vdash \bot = \bot}{(\bot = \bot) = \bot \vdash \bot}
}{C = \bot \vdash \bot}
}{}
}{C \vdash \bot}
\qquad
\bot \vdash C
}{\vdash C = \bot}
\qquad
\dfrac{
\dfrac{
\dfrac{C \vdash C = \bot \qquad C = \bot \vdash C}{\vdash (C = \bot) = C}
}{C = \bot \vdash (\bot = \bot) = \bot}
\qquad
\dfrac{\vdash \bot = \bot}{(\bot = \bot) = \bot \vdash \bot}
}{C = \bot \vdash \bot}
}{\vdash \bot}
$$

3 Concluding remarks

In this paper we have presented some difficulties for the principle that logical equivalents are substitutable *salve veritate* (and formally analogous principles.) We have also shown that recent responses to similar paradoxes that involve weakening the structural rules of the logic do not seem to provide much relief in this context.

It is worth remarking that classical theories of truth – theories that are not committed to theorems conforming to the fixed point schema FP – do not have to give up the principle that logical equivalents be substitutable *salve veritate* (see in particular the theories FS and FS_n described in [11].[25]) This reversal of fortunes is worthy of note; while the non-classical logician must make certain concessions regarding the logical connectives (notably the rule of contraction, conditional proof, and so on) the upside is a simple and intuitive theory of truth. The substitutivity of logical equivalents *salve veritate*, however, is surely a part of the naïve conception of truth. Yet it is an example of a principle explicitly concerning truth (one that does not principally govern the logical connectives), that the classical logician can retain but which the non-classical logician apparently cannot.

References

[1] A. Bacon. A new conditional for naive truth theory. *Notre Dame Journal of Formal Logic*, 54(1):87–104, 2013.

[2] A. Bacon. Non-classical Meta-theory for Non-classical Logic. *The Journal of Philosophical Logic*, 42(2):335–355, 2013.

[3] JC Beall. *Spandrels of truth.* Oxford University Press, USA, 2009.

[4] JC Beall and J. Murzi. Two flavors of curry paradox. *Journal of Philosophy*, Forthcoming.

[5] R. Brady. *Universal logic.* CSLI Publications, 2006.

[6] M.J. Cresswell. Propositional identity. *Logique et Analyse*, 40:283–291, 1967.

[7] H. Field. *Saving truth from paradox.* Oxford University Press, USA, 2008.

[8] H. Field. What is logical validity? In Colin Caret and Ole Hjortland, editors, *Foundations of Logical Consequence.* Oxford University Press, forthcoming.

[9] H. Friedman and M. Sheard. An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33:1–21, 1987.

[10] V.N. Grišin. Predicate and set-theoretical calculi based on logic without the contraction rule, mathematical ussr izvestiya 18 (1982), 41–59. *English transl., Izvéstia Akademii Nauk SSSR*, 45:47–68, 1981.

[11] V. Halbach. A system of complete and consistent truth. *Notre Dame Journal of Formal Logic*, 35(3):311–327, 1994.

[12] J. Murzi and L. Shapiro. Validity and truth-preservation. In *Unifying the Philosophy of Truth.* Springer, forthcoming.

[13] G. Priest. *An introduction to non-classical logic.* Cambridge University Press, 2001.

[25] For example each instance of the schema $Pr_{FS}(\ulcorner A \leftrightarrow B\urcorner) \to (Tr(\ulcorner C\urcorner) \to Tr(\ulcorner C[A/B]\urcorner)$, where Pr_{FS} is an arithmetical formula expressing provability in FS, is true in the revision sequence described in Friedman and Sheard [9] (Pr_{FS} can in fact be strengthened to 'provability from FS and true arithmetic'.) Differences between the motivations of this kind of theory unfortunately prevent a direct comparison to the non-classical approach.

[14] G. Priest. *In contradiction: a study of the transconsistent.* Oxford University Press, 2006.

[15] G. Priest. Fusion and confusion. *Topoi*, Forthcoming.

[16] D. Ripley. Paradoxes and failures of cut. *Australasian Journal of Philosophy*, 91(1):139–164, 2013.

[17] S. Schiffer. *The things we mean.* Clarendon Oxford, 2003.

[18] P. Welch. Ultimate Truth vis-a-vis Stable Truth. *The Review of Symbolic Logic*, 1(01):126–142, 2008.

[19] B. Whittle. Dialetheism, logical consequence and hierarchy. *Analysis*, 64(284):318–326, 2004.

[20] E. Zardini. Naive modus ponens. *Journal of Philosophical Logic*, 42(4):1–19, 2013.