Paradoxes of logical equivalence and identity

Author

April 11, 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 the claim that ϕ for the claim that 'ϕ' is true, or the claim that x is F for the claim that x belongs to the set of F's in ordinary contexts. 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 shall argue that this strategy is more limited than some philosophers have realised; when one tries to formulate reasoning involving intersubstitutivity principles not as a metalinguistic rule but as a principle in the object language, new paradoxes arise.

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 minimal 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 [2], Brady [4], Priest [10].) 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, [13] and [3].) 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 [14], Priest [11], Murzi & Shapiro [9].)

1 Applications of the paradoxes

1.1 Substitutability of logical equivalents

In [5] 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 (A and B are logical equivalents if the biconditional $A \leftrightarrow B$ is a logical truth.)[1] This fact about the logic, according to Field, represents a significant improvement over rival theories such as Priest's [10], in which logically equivalent sentences are not intersubstitutable.

What does it mean for two sentences to be intersubstitutable? It is indisputable that Field's logic has a rule that allows one to substitute logical equivalents within logical truths

[1] I don't want to take sides on what it might mean to call something a logical truth, but I assume that in order to engage in the debate about whether logical equivalents are intersubstitutable you have to have some notion in mind. In what follows I will not assume that it can be captured by provability in some system, or indeed, that statements of logical truth or equivalence satisfy classical laws such as the law of excluded middle.

in a way that preserves logical truth.[2] However this rule is far too weak, allowing us only to substitute within sentences we can prove – it says nothing about substitutability within ordinary contingent truths. Field's logic also contains the rule $A \leftrightarrow B \vdash C \leftrightarrow C[A/B]$, which allows one to substitute A for B when one already knows that A and B are materially equivalent (and also, presumably, if one already knows that A and B are logically equivalent.) This allows substitutability of logical equivalents within contingent truths to a limited extent, but it is not enough. If logical equivalents truly are substitutable we ought to be able to say, even when one does not know whether A and B are logically equivalent, that *if* they are, then C is true iff $C[A/B]$ is.

To properly address the problems that Field raises for Priest, for example, one needs to formulate and assess the principle that logical truths are substitutable *salve veritate*, which just means:

> (SSV) If A is true and 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 principle says that substituting something for a logical equivalent does not change the truth value of the sentence it occurs in. It is, of course, stronger than the principle that substituting logical equivalents in logical truths leaves you with logical truths; but that is only because the latter is exceptionally weak and does not properly encode the intersubstitutivity thought. The principle that logical truths are substitutable *salve veritate* can be reformulated in different ways. Here are a couple

> If A is true, then if B is logically equivalent to C then $A[B/C]$ is true.
>
> From the fact that A is true infer that if B is logically equivalent to C then $A[B/C]$ is true.

Assuming classical logic these really are just reformulations of the original principle; however in Field's logic they (successively) weaken the principle that logical truths are substitutable *salve veritate.*

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 weakest of these substitution principles and combine it with two further natural principles governing logical equivalence.

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

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

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

The following paradoxes pose problems for this combination of principles given various background logical assumptions.

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 interubstitutivity of logical equivalents (see section 20.5 in [5].) 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): you cannot in general infer that two claims are logically equivalent from the fact that one can prove they are equivalent from non-logical assumptions. On this account, then, there is some way of interpreting 'logical equivalence' that allows us to say that logical equivalents are intersubstitutable – more precisely, some

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

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?

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 that necessary *a priori* truth; if it were less demanding RE(E) would be acceptable.

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: 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?

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. The rule RE($\doteq$) is subject to the same caveats we discussed in section [REF], 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.[3]

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

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.[4] 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.[5] One might take this as evidence against the unrestricted version of LL($\doteq$), and the axiom versions that it appears to follow from.

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.[6])

1.3 Propositional identity

The paradoxes we consider can also be generated if we wish to introduce a propositional identity connective, $=$, into the language. Formally analogous principles can be formulated for this connective:

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

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 [REF] or [REF] is taken as primitive one can make analogous arguments, with a bit of fiddling here and there, 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 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: $V(\ulcorner (A \leftrightarrow B) \wedge (\neg A \leftrightarrow \neg B) \urcorner)$.)[7]
>
> 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.

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

[5] I am grateful to Hartry Field and an anonymous referee for this observation. 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 [REF].

[6] If we define a determinacy operator as $A \wedge \neg(A \to \neg A)$ 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.

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

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 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 $C(Tr(\ulcorner A_0\urcorner), \ldots, Tr(\ulcorner A_n\urcorner))$.)[8]

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 $\{\rightarrow, \wedge, \perp\}$, a propositional identity connective, $=$, and which contains, for each formula of the language, ϕ, a propositional constant A governed by the following axiom

FP $A \leftrightarrow \phi[A/B]$

Here B can be any propositional letter (possibly occurring in ϕ) and $A \leftrightarrow B$ is short hand for $A \rightarrow B \wedge B \rightarrow A$. Axioms of this form arise in the context both of naïve property/set theory and naïve truth theories.[9] 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.[10]

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

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.[11] RN, however, is strictly weaker than RE.

[8] 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.

[9] 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.)

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

[11] 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.

2.1 The first paradox

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 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 it does not appear to be responsible for any of the standard paradoxes (and indeed there are consistent naïve truth and set theories that contain the principle Grisin [7]) it is not validated in some recent theories (see e.g. Bacon [1], Beall [2], Brady [4], Field [5], Priest [10].) We shall consider paradoxes that drop RA in the next section

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. $C \leftrightarrow (\Box C \to \bot)$ instance of FP.
2. $C \to (\Box C \to \bot)$ by 1.
3. $\Box C \to (C = C \to (\Box(C = C) \to \bot)$ from 2 by LL
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 second 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 is strictly stronger than the rule of necessitation. In this argument we also have to assume a standard axiom governing the falsum constant.

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

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

I $A = A$

F $\bot \to A$

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

As before we assume modus ponens and FP.

1. $C \leftrightarrow (C = \bot)$ an instance of FP.
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 [13], Beall & Murzi [3].)

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[12]

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

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

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 $A \Rightarrow A$.

[12]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.

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

For example in the latter case one can begin with the sentence $C \leftrightarrow (C \to \bot)$ (by FP) and $C \to C$ (by I) to infer $C \to (C \wedge (C \to \bot)$. By PMP we have $(C \wedge (C \to \bot)) \to \bot$. So by TR we have $C \to \bot$, 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 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 [14]) or modus ponens (Priest [11]).[13]) 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 $\vdash A \to B$

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 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 the formalism does not commit us to this, it is very natural to think of RN as a rule that 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, $\Box A \to \Box\Box A$, 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 $A \to B, B \to C \vdash A \to C$

EX $A, \neg A \vdash B$

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

T $\Box A \to A$

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

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

[13]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.

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

These principles are mostly unremarkable. The two most notable principles are $\mathsf{4}$, and the rule R¬N, which states that when you can prove that $A \to C$ and $A \to \neg C$ then one can infer that A is not logically valid. Neither this nor the principle of explosion, EX, will be acceptable to a paraconsistent logician, however for Field and others this is a very natural rule to endorse. (The following is thus problematic for Zardini [14], who is committed to the $\mathsf{S4}$ axiom modulo definitions although, apparently, not R¬N.) 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¬N
8. $\neg\Box\Box A \to \neg\Box A$ by contraposition on $\mathsf{4}$
9. $\neg\Box A$ by $\mathsf{7}$ and 8.
10. A by 1
11. $\Box A$ RN
12. $\perp$ by EX.

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[14] we cannot move from the fact that a rule preserves provability to the fact that it preserves validity. 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.[15] 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.[16] 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.

[14] 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 [5] is highly non-recursive (see [12]) and so has no complete axiomatisation.

[15] 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.

[16] the 'paradoxes of provability' can be represented by completely determinate (albeit unprovable) facts about the natural numbers, as Gödel has showed.

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 [14], Priest (this volume), Murzi and Shapiro [9].)

It is of particular interest, then, to ascertain whether the present paradoxes involving logical equivalence and related notions can be generated once various structural rule 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.

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 $\Gamma \rhd A$, therefore, consists of a sequence (not a set) of premises, Γ, and a conclusion formula A. A sequent calculus is 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:

$$\frac{A \rhd B, B \rhd A}{\rhd A = B}\ (RE') \qquad \frac{\rhd A}{B = C \rhd A[B/C]}\ (Sub)$$
$$\frac{}{\rhd A = A}\ (Id) \qquad \frac{}{\bot \rhd A}\ (Bot) \qquad (1)$$
$$\frac{A \rhd B, B \rhd C}{A \rhd C}\ (SCut)$$

In place of FP I shall just help myself to the sequents $C \rhd C = \bot$ and $C = \bot \rhd C$, where C denotes a fixed point for the formula $X = \bot$. The following argument is completely analogous to the paradox presented in §2.1

1. $C \rhd C = \bot$
2. $C = \bot \rhd C$
3. $\rhd (C = \bot) = C$ by RE$'$
4. $C = \bot \rhd (\bot = \bot) = \bot$ from 3 by Sub
5. $\rhd \bot = \bot$ Id
6. $(\bot = \bot) = \bot \rhd \bot$ Sub from 5
7. $C = \bot \rhd \bot$ SCut

8. $C \rhd \bot$ from 1, 7 and SCut
9. $\bot \rhd C$ by Bot
10. $\rhd C = \bot$ RE$'$
11. $\rhd C$ by 2, 10 and SCut
12. $\rhd \bot$ 8 and 11 by SCut.

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 [8].[17]) 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] Andrew Bacon. A new conditional for naive truth theory. *Notre Dame Journal of Formal Logic*, 54(1):87–104, 2013.

[2] J. Beall. *Spandrels of truth.* Oxford University Press, USA, 2009.

[3] Jc Beall and Julien Murzi. Two flavors of curry paradox. *Journal of Philosophy*, Forthcoming.

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

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

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

[7] VN 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.

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

[17] For example each instance of the schema $Pr_{FS}(\ulcorner A \leftrightarrow B \urcorner) \rightarrow (Tr(\ulcorner C \urcorner) \rightarrow Tr(\ulcorner C[A/B] \urcorner)$, where Pr_{FS} is an arithmetical formula expressing provability in FS, is true in the revision sequence described in Friedman and Sheard [6]. Differences between the motivations of this kind of theory unfortunately prevent a direct comparison to the non-classical approach. For example some of the axioms of FS are most naturally thought of as contingent principles about the actual notion of truth (FS states that a disjunction, 'A or B', is true iff one of the disjuncts is. This principle would have been false had 'or' meant and.) On the other hand non-classical theorists typically treat the theorems of their theories to be logically necessary.

[9] Julien Murzi and Lionel Shapiro. Validity and truth-preservation. In *Unifying the Philosophy of Truth Springer.* Springer, forthcoming.

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

[11] Graham Priest. Fusion and confusion. *Topos*, Forthcoming.

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

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

[14] Elia Zardini. Naive modus ponens. *Journal of Philosophical Logic*, pages 1–19, 2012.