Curry's Revenge: the costs of non-classical solutions to the paradoxes of self-reference

Greg Restall
Department of Philosophy
The University of Melbourne
restall@unimelb.edu.au

July 12, 2006

. . . .

The paradoxes of self-reference are *genuinely* paradoxical. The liar paradox, Russell's paradox and their cousins pose enormous difficulties to anyone who seeks to give a comprehensive theory of semantics, or of sets, or of any other domain which allows a modicum of self-reference and a modest number of logical principles.

One approach to the paradoxes of self-reference takes these paradoxes as motivating a *non-classical* theory of logical consequence. Similar logical principles are used in each of the paradoxical inferences. If one or other of these problematic inferences are rejected, we may arrive at a consistent (or at least, a coherent) theory.

In this paper I will show that such approaches come at a serious cost. The general approach of using the paradoxes to restrict the class of allowable inferences places severe constraints on the domain of possible propositional logics, *and* on the kind of metatheory that is appropriate in the study of logic itself. Proof-theoretic and model-theoretic analyses of logical consequence make provide different ways for non-classical responses to the paradoxes to be defeated by *revenge* problems: the redefinition of logical connectives thought to be ruled out on logical grounds. Non-classical solutions are not the "easy way out" of the paradoxes.[1]

1 Non-Classical Solutions

In this section I will sketch the structure of non-classical approaches to the paradoxes. They have straightforward general features: Firstly, we keep whatever semantic, or set-theoretic principles are at issue. For example, if it is the liar paradox in question, we can keep the naïve truth scheme, to the effect that

$$T\langle A\rangle \leftrightarrow A$$

where $\langle — \rangle$ is some name-forming functor, taking sentences to names, and where $\leftrightarrow$ is some form of biconditional. This scheme says, in effect, that $T\langle A\rangle$ is true under the same circumstances as A. To assert that A is true is saying no more and no less than asserting A.

[1]It does not follow that non-classical accounts of the paradoxes are misguided or wrong-headed. On the contrary, I think that the general approach is quite sane, and have argued as much in print [18].

Secondly, we allow our language to contain a modicum of self-reference. We wish to express sentences such as the liar: "This very sentence is not true." If the language in question is a natural language, then indexicals will do the trick. If the language is a formal language without indexicals, some other technique will be needed to construct sentences analogous to the liar. A Gödel numbering and a means of diagonalisation will do nicely to give the required results.[2]

With such machinery at hand, we can reason as follows: Use a means of diagonalisation to construct a statement λ such that λ is equivalent to $\sim T\langle\lambda\rangle$. Then reason as follows: $\lambda \leftrightarrow \sim T\langle\lambda\rangle$, but by the T-scheme, $\lambda \leftrightarrow T\langle\lambda\rangle$. Therefore $T\langle\lambda\rangle \leftrightarrow \sim T\langle\lambda\rangle$, and equivalently, $\lambda \leftrightarrow \sim\lambda$. We can then deduce $\lambda \wedge \sim\lambda$ (from an inference such as *reductio*: $p \rightarrow \sim p \vdash \sim p$) and we have a contradiction.

If your favourite paradox is Russell's, instead of the liar, the non-classical approach will keep the naïve class abstraction scheme

$$x \in \{y : \phi(y)\} \leftrightarrow \phi(x)$$

and you reason similarly, from the definition of the Russell class r as $\{x : x \notin x\}$. If $r \in r$ then $r \notin r$, and if $r \notin r$ then $r \in r$. The same holds for Berry's paradox, the Burali-Forti paradox, and many others.[3]

The non-classical response to these paradoxes is to find fault with the logical principles involved in the deduction. Most approaches to the paradoxes take them to be important lessons in the behaviour of *negation*. There are two different lessons we might learn. One is that the inference from $A \leftrightarrow \sim A$ to $A \wedge \sim A$ fails, since A might be (speaking crudely) neither true nor false. Another possible lesson is that the inference from $A \wedge \sim A$ to an arbitrary B fails, since A can be (speaking less crudely this time) both true and false. However negation works, it cannot be *Boolean*. Boolean negation allows *both* inferences, and inferring *every* statement from a the existence of liar sentence or a Russell set is just too much. Boolean negation is rather too strong, so an alternative logic of negation must be found [5, 6, 16, 17].

If you wish to define negation non-classically, there are many options available. You can define negation *inferentially*, taking $\sim A$ to mean that *if* A, *then* something absurd follows,[4] or it can be defined by way of the equivalence between the *truth* of $\sim A$ and the *falsity* of A, and allowing truth and falsity to have rather more independence from one another than is usually taken to be the case: say, allowing statements to be *neither* true nor false, or *both* true and false.[5] The former account takes truth as primary, and defines negation in terms of a rejected proposition and implication. In the context of the semantics of relevant logics, this approach is sometimes called the *Australian Plan*. The account which takes truth and falsity as on a par is sometimes called the *American Plan*. In either case, there are many options for the theorist seeking an alternative account of negation.

[2] See Boolos and Jeffrey [4] for a review of the standard approach, and see Smullyan [21] for more on what a language must contain to feature self-reference.

[3] A compendium of such paradoxes is given by Graham Priest [17].

[4] See, for example, Meyer and Martin's account of negation as implying falsehood, and its idiosyncrasies when combined with a *relevant* notion of implication [11]

[5] Three examples are four valued semantics of relevant logics, used by Dunn [8, 9] and Belnap [2, 3], the semantics of Priest's *In Contradiction* [17] and the semantics of Nelson's constructible falsity [15] and its extensions by Heinrich Wansing [23].

I sketch this general typology of negation merely to indicate that I need not take a stand on it. In what follows we will see that the paradoxes have more to teach us than this. If we wish to be non-classical, we need to work with much more than the logic of negation.

2 Curry's Paradox

The paradox I have in mind can be found in a logic independently of its stand on negation. The deduction appeals to no particular principles of negation, as it is negation-free. Any deduction must use some inferential principles. Here are the principles needed to derive the paradox.

A TRANSITIVE RELATION OF CONSEQUENCE: We write this by '$\vdash$'. I take $\vdash$ to be a relation between statements, and I require that it be transitive: if $A \vdash B$ and $B \vdash C$ then $A \vdash C$.

CONJUNCTION AND IMPLICATION: I require that the conjunction operator $\wedge$ be a greatest lower bound with respect to $\vdash$. That is, $A \vdash B$ and $A \vdash C$ if and only if $A \vdash B \wedge C$. Furthermore, I require that there be a *residual* for conjunction: a connective $\supset$ such that

$$A \wedge B \vdash C \text{ if and only if } A \vdash B \supset C$$

This is our connective of implication. (You may wonder how we might come across such a connective. There are many ways to construct it. In the next section we will examine some.)

A PARADOX GENERATOR: We need only a very weak paradox generator. We take the T scheme in the following *enthymematic* form:

$$T\langle A\rangle \wedge C \vdash A \qquad A \wedge C \vdash T\langle A\rangle$$

for some true statement C. The idea is simple: $T\langle A\rangle$ need not *entail* A.[6] Take C to be the conjunction of all required background constraints. It is true that the *sentence* "snow is white" could be true without snow being white. However, if "snow is white" is true and some background semantic theory is holds, then it follows that snow is white. Conversely, if snow is white and the background semantic theory holds, then "snow is white" is true. Let C be that background semantic theory. It must simply give A under some background constraints (such as some facts about language) we can infer A. (If you find it difficult to construct the required background semantic theory. Do not worry. Take C to be the conjunction of *all* truths: a maximally specific statement. Then we need simply that there is no instance of A for which in which $T\langle A\rangle$ is true and A fails to be true, or vice versa.)

DIAGONALISATION To generate the paradox we use a technique of diagonalisation to construct a statement λ such that λ is equivalent to $T\langle\lambda\rangle \supset A$, where A is

[6] So, we need not take the range of the T-scheme to be *propositions* [10], as we do not need to commit ourselves to the *equivalence* of $T\langle A\rangle$ and A.

any statement you please. Then, with this A chosen, we reason as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{C \wedge T\langle\lambda\rangle \vdash \lambda \quad \lambda \vdash T\langle\lambda\rangle \supset A}{C \wedge T\langle\lambda\rangle \vdash T\langle\lambda\rangle \supset A}}{C \wedge T\langle\lambda\rangle \wedge T\langle\lambda\rangle \vdash A}}{C \wedge T\langle\lambda\rangle \vdash A}\,(*)}{C \vdash T\langle\lambda\rangle \supset A} \quad T\langle\lambda\rangle \supset A \vdash \lambda}{C \vdash \lambda} \quad C \wedge \lambda \vdash T\langle\lambda\rangle}{C \vdash T\langle\lambda\rangle} \quad \dfrac{\dfrac{\text{from } (*)}{C \wedge T\langle\lambda\rangle \vdash A}}{T\langle\lambda\rangle \vdash C \supset A}}{C \vdash C \supset A}}{C \wedge C \vdash A}}{C \vdash A}$$

This is a problem. Our true C entails an arbitrary A.

This inference arises independently of any treatment of negation. The form of the inference is reasonably well known. It is *Curry's paradox*, and it causes a great deal of trouble to any non-classical approach to the paradoxes [12, 13, 14, 20]. In the next section I show how the tools for Curry's paradox are closer to hand than you might think. Avoiding this paradox severely constrains the non-classical theorist.

3 The Revenge Problem

There are many different ways to get the logical tools necessary for our problematic deduction. In particular, there are many ways to get a connective $\supset$ which residuates conjunction. We will examine them one at a time.

BOOLEAN NEGATION: If Boolean negation is present (write it "$\sim$") then we can define $A \supset B$ to be $\sim A \vee B$. However, the non-classical theorist has explicitly rejected Boolean negation, so we need not tarry here. This is not a problem by itself.

INTUITIONISTIC LOGIC: The rule for the residual is satisfied by the conditional of intuitionistic logic. Any semantic account which motivates intuitionism motivates the residual of conjunction. Now no non-classical theorist of the paradoxes is going to *explicitly* use the intuitionistic conditional, for it is well known to suffer from Curry-style paradoxes. Our point in the rest of the paper is to show that the *implicit* acceptance of this conditional is deeply embedded in our practices of logic.

INFINITARY DISJUNCTION: A Curry-paradoxical conditional can arise as a revenge problem for the non-classical theorist without *explicitly* motivating intuitionistic implication. If we have *infinitary* disjunction at hand, such that a (finite) conjunction distributes over infinitary disjunction, we can define $B \supset C$ to be

$$\bigvee \{A : A \wedge B \vdash C\}$$

This will satisfy the definition of $\supset$. If $A' \wedge B \vdash C$ then $A' \vdash \bigvee\{A : A \wedge B \vdash C\}$, since $A' \in \{A : A \wedge B \vdash C\}$. Conversely, if $A' \vdash B \supset C$, we have $A' \vdash \bigvee\{A : A \wedge B \vdash C\}$. Then $A' \wedge B \vdash B \wedge \bigvee\{A : A \wedge B \vdash C\}$ and by the distribution of conjunction over disjunction, $A' \wedge B \vdash \bigvee\{A \wedge B : A \wedge B \vdash C\}$ and clearly $\bigvee\{A \wedge B : A \wedge B \vdash C\} \vdash C$, so $A' \wedge B \vdash C$ by the transitivity of entailment. Therefore, *any* semantic theory which motivates infinitary disjunction and distributive lattice logic motivates the residual $\supset$ of conjunction, and our problematic inference. This seriously constrains non-classical solutions to the paradoxes, for infinitary disjunction can be motivated in many different ways.

Proof Theory: If your favoured way to introduce connectives is by way of natural deduction (introduction and elimination rules) then infinitary disjunction is no less motivated than ordinary disjunction. To infer $\bigvee X$ from a statement A, it is sufficient to infer a member of X.

$$\frac{A \vdash B_i}{A \vdash \bigvee\{B_i : i \in I\}}$$

If you can infer A from each element of X, then you can infer A From $\bigvee X$ too.

$$\frac{A_i \vdash B \text{ (each } i \in I)}{\bigvee\{A_i : i \in I\} \vdash B}$$

This rule is the left-hand Gentzen rule. For a traditional elimination rule for a natural deduction system, you use

$$\frac{C \vdash \bigvee\{A_i : i \in I\} \quad A_i \vdash B \text{ (each } i \in I)}{C \vdash B}$$

which is equivalent, given the transitivity of entailment. These rules seem to motivate the connective straightforwardly. However, a non-classical theorist of the paradoxes must do one of two things. One response is to allow the connective but to deny the distribution of conjunction over disjunction: that is, we do not have

$$A \wedge \bigvee\{B_i : i \in I\} \vdash \bigvee\{A \wedge B_i : i \in I\}$$

Such an approach has its own difficulties: however, it may be attempted. The second response is to reject the definition of $\bigvee$ in some way. It must be argued that this does not define a connective. This will require giving a precise account of what proof-theoretical principles are *permissible* in the account of a logical connective, and which principles are illicit. To avoid doing this is to leave the theory open to Curry's revenge.

The problem does not end here, however. The non-classical theorist must also have something to say in areas other than proof theory, for we can define disjunction in many different ways.

The Algebra of Propositions: Some logicians treat the class of propositions as an *algebra*. This algebra is closed under various operations, which have different algebraic properties. The algebra of propositions is *complete* if it is closed under arbitrary conjunctions and disjunctions. The non-classical theorist (who accepts the distribution of conjunction over disjunction) must hold that the "intended" algebra of propositions is incomplete. This is not a particularly great burden in and of itself. However, it becomes a burden when we consider the constructions available which naturally *complete* incomplete lattices [7, 19].

Here is one result of this nature. If the lattice of propositions is incomplete, then define a *new* lattice of propositions like this. The new propositions are *ideals* of the old lattice. A set I of propositions is an ideal if and only if it is closed under converse entailment (if a entails b and $b \in I$ then $a \in I$ too) and disjunction (if $a \in I$ and $b \in I$ then $a \vee b \in I$ too). You can think of I as a set of propositions such that you would like *one* of them to be true. Our conditions ensure that we add into the set any other proposition such that making it true will be enough to make one of our original choices true. (The smallest ideal containing $\{a, b, c\}$ is the set of all propositions entailing $a \vee b \vee c$. If any of these propositions are true, then $a \vee b \vee c$ is true, which ensures that either a or b or c is true.)

Now the ideals behave *just like propositions.* The conjunction of a class of ideals is the intersection of that class. The disjunction of a class of ideals is the smallest ideal containing that class. The entailment relation among ideals is just the relation of inclusion. The collection of ideals forms a *complete* lattice. Every set of ideals has a disjunction and a conjunction. The logic of the set of ideals is very similar to the logic of propositions out of which it was constructed. However, it is complete.[7]

This is not merely a mathematical construction with no possibility for interpretation. Given an algebra of propositions, any ideal in the structure can be treated as a proposition, with simple truth conditions: I is *true* just when one member of I is true. Given a class I of propositions, it makes sense to commit yourself to the claim that one member of I is true, and this claim ought to be true just when one member of I is true.[8]

To avoid revenge, the non-classical theorist explain the point at which this reasoning breaks down. There is some ideal in the structure such that the truth of a member of I is not expressible in the domain of propositions. This is a strange result indeed, and it is a cost to the non-classical theory. It seems that the class of propositions of the language must forever remain incomplete.[9]

Once this version of the revenge problem is avoided, there yet is *another* way in which Curry's paradox might take its revenge.

State Models: Perhaps the simplest way to construct infinitary disjunction is by way of what we might call "state models" of our logics. In a state model, each proposition is modelled by the set of states in which that proposition is true. Possible worlds models for modal logics are one form of state model.

Given a state model, it seems that infinitary disjunction is close at hand. Take a class of propositions. Their disjunction is true at the union of the class of sets of states at which each proposition in that class is true. The disjunction is true at a state just when one member of the disjunction is true

[7] The proof is not difficult, but I will not rehearse it here [7, 19].

[8] It would be very odd for a non-classical logician to reject *this* step, for she is the one defending the equivalence of $T\langle A\rangle$ and A.

[9] It will not do, either, to say that there are too many ideals to be expressible in a finitary language. For we do not *need* infinitary expressions to justify the existence of the residual $\supset$: The only infinitary disjunctions we need are those of the form $\bigvee\{A : A \wedge B \vdash C\}$ and these can be expressed in a finitary fashion.

at that state. This will define a proposition, which is the infinitary disjunction in the language. This construction relies on the notion that this class of states gives rise to a proposition. The non-classical theorist is free to reject this. However, to do so would require an explanation of which classes of states *do* give rise to propositions and which do not, and to explain why it rules out the kind of infinitary disjunction sufficient for generating the conditional for Curry's revenge.

4 Choices

Here, then, are the choices for any theory which seeks to give an account of the paradoxes of self-reference.

REJECT LARGE DISJUNCTIONS: This requires formulating responses to each of the arguments of the previous section. This has not been done, as yet, and it is unclear what a non-classical theory which takes those arguments seriously might look like. In particular, it would aid the cause of the non-classical theorist to be able to point to a particular class of propositions and to explain why *that* class has no disjunction. It is unclear what such an explanation could look like.

REJECT DISTRIBUTION: A crucial step in each argument has been the distribution of conjunction over disjunction. This inference has been under question for a number of reasons; primarily in quantum logic and in substructural logics. It is unclear how to motivate the failure of distribution in *this* context. It would be very nice to be able to point to a particular case of distribution and to have an explanation of why the premise is true but the conclusion fails. Such explanations are forthcoming in quantum logic (even if they are not always convincing). We need one to motivate the failure of distribution for non-quantum reasons. Uwe Petersen has the most fully developed non-classical theory of the paradoxes which lives without distribution [16]. However, no-one has given an explanation of *why* distribution fails, other than as an artefact of the proof theory. J. L. Bell has developed a semantics for quantum logic which motivates the failure of distribution [1]. It would be a great advance too if such a semantics could help *explain* a failure of distribution in the case of the paradoxes.

REJECT THE TRANSITIVITY OF ENTAILMENT: This may be seen to be cutting off one's nose to spite one's face, but this approach has its proponents. Neil Tennant gives one theory of consequence which abandons the transitivity of entailment [22]. Tennant does not do this for reasons of the paradoxes, and Tennant's non-transitive logical systems do not seem to help in this case. For Tennant, if we have a proof from A to B and a proof from B to C is valid, then we do not necessarily have a proof from A to C, but we *do* have either a refutation of A (a proof from A to the empty conclusion) or a proof of C (a proof of C from the empty premises) or a proof from A to C. So, for our purposes we may talk of arguments being *weakly* valid if and only if there is a Tennant-proof of some *superset* of the premises to some *superset* of the conclusions, and this notion of consequence is transitive, and it does all that we need to generate the paradoxes. Avoiding the transitivity of consequence in Tennant's style does not suffice for avoiding Curry's revenge.

REJECT THE STRONG LAWS: To live without the T-scheme or the naïve class comprehension scheme is to give up the goal of giving a non-classical account of the paradoxes. If the fault *isn't* with the logic but is with the semantics or the mathematics or whatever else we used, then the paradoxes do not motivate a non-classical logical theory. A classical one will do.

Each approach has its cost. None are straightforward. There is much work left to do, if we wish to give *any* account of the account of the paradoxes, including those which involve revising logical theory.[10]

References

[1] J. L. Bell. A new approach to quantum logic. *British Journal for the Philosophy of Science*, 37:83–99, 1986.

[2] Nuel D. Belnap. How a computer should think. In G. Ryle, editor, *Contemporary Aspects of Philosophy*. Oriel Press, 1977.

[3] Nuel D. Belnap. A useful four-valued logic. In J. Michael Dunn and George Epstein, editors, *Modern Uses of Multiple-Valued Logics*, pages 8–37. Reidel, Dordrecht, 1977.

[4] George Boolos and Richard Jeffrey. *Computability and Logic*. Oxford University Press, third edition, 1989.

[5] Ross T. Brady. The simple consistency of a set theory based on the logic csq. *Notre Dame Journal of Formal Logic*, 24:431–449, 1983.

[6] Ross T. Brady and Richard Routley. The non-triviality of extensional dialectical set theory. In Graham Priest, Richard Routley, and Jean Norman, editors, *Paraconsistent Logic: Essays on the Inconsistent*, pages 415–436. Philosophia Verlag, 1989.

[7] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, Cambridge, 1990.

[8] J. Michael Dunn. An intuitive semantics for first degree relevant implications (abstract). *Journal of Symbolic Logic*, 36:363–363, 1971.

[9] J. Michael Dunn. A Kripke-style semantics for R-mingle using a binary accessibility relation. *Studia Logica*, 35:163–172, 1976.

[10] Paul Horwich. *Truth*. Basil Blackwell, Oxford, 1990.

[11] Robert K. Meyer and Errol P. Martin. Logic on the australian plan. *Journal of Philosophical Logic*, 15:305–332, 1986.

[12] Robert K. Meyer, Richard Routley, and J. Michael Dunn. Curry's paradox. *Analysis*, 39:124–128, 1979.

[13] J. Myhill. Levels of implication. In A. R. Anderson, R. C. Barcan-Marcus, and R. M. Martin, editors, *The Logical Enterprise*, pages 179–185. Yale University Press, New Haven, 1975.

[14] J. Myhill. Paradoxes. *Synthese*, 60:129–143, 1984.

[15] D. Nelson. Constructible falsity. *Journal of Symbolic Logic*, 14:16–26, 1949.

[16] Uwe Petersen. Diagonal method and dialectical logic. (Unpublished Book), 1992.

[17] Graham Priest. *In Contradiction: A Study of the Transconsistent*. Martinus Nijhoff, The Hague, 1987.

[10]Thanks to audiences at the 1998 Australasian Association of Philosophy Conference and at Stanford University (especially John Etchemendy, Grigori Mints, Chris Mortensen and Graham Priest) for helpful comments on earlier versions of this paper.

[18] Greg Restall. Deviant logic and the paradoxes of self reference. *Philosophical Studies*, 70:279–303, 1993.

[19] Greg Restall. *An Introduction to Substructural Logics.* Routledge, 2000.

[20] Moh Shaw-Kwei. Logical paradoxes for many-valued systems. *Journal of Symbolic Logic*, 19:37–40, 1954.

[21] Raymond M. Smullyan. *Diagonalization and Self-Reference*, volume 27 of *Oxford Logic Guides.* Clarendon Press, 1994.

[22] Neil Tennant. The transmission of truth and the transitivity of deduction. In Dov Gabbay, editor, *What is a Logical System?*, volume 4 of *Studies in Logic and Computation*, pages 161–177. Oxford University Press, Oxford, 1994.

[23] Heinrich Wansing. *The Logic of Information Structures.* Number 681 in Lecture Notes in Artificial Intelligence. Springer-Verlag, 1993.