Ciro De Florio and Alessandro Giordani
From Arithmetic to Metaphysics
De Gruyter Studium

Massimilano Carrara

Naïve Proof and Curry's Paradox

1 Introduction

In classical first order logic (FOL), *trivialism*, the truth of all sentences, and *explosion*, the derivability of any sentence, are obtained using the rule *ex contradictione quodlibet* (ECQ): $A, \neg A \vdash B$. The classical justification for ECQ rests on the alleged evidence that no contradiction can be true, evidence rejected in paraconsistent theories, in particular by dialetheists, who hold that there are dialetheiae, i.e. propositions that are both true and false.[1] Indeed, dialetheism maintains the thesis that there are true contradictions, i.e. true sentences of form $(A \wedge \neg A)$, called *dialetheiae*. More generally, they call *dialetheia* any sentence that is both true and false. In an extensive series of papers and books (see for example, Priest [1979], Priest [2001], Priest [2002], Priest [2006a], Priest [2006b]), Priest claims that the paradoxical sentences obtained from self-reference are *dialetheiae*.[2]

In standard natural deduction of FOL, ECQ can be derived using reductio ad absurdum (RAA) and other apparently non-problematic rules. It is a standard derived rule of FOL. But if you hold that there are dialetheiae, in order to avoid trivialism, RAA should be immediately rejected. Unfortunately, banishing RAA is insufficient to avoid trivialism: Curry's paradox, from which trivialism follows, can be generated without using RAA, but with just *modus ponens* (MP) and the derived rule of Absorption, i.e. ABS: $(A \rightarrow (A \rightarrow B)) \vdash (A \rightarrow B)$. In order to save dialetheism from trivialism, Priest adopts in the Logic of Paradox (LP) (1979) the material conditional, for which he rejects the general validity of MP.

The crucial problem is whether *trivialism* can follow even from logical principles that are dialetheistically correct. In this paper I concentrate, specifically, on a notion that Priest himself introduced in *The Logic of Paradox* (1979), i.e. that of naïve proof, a notion amplified in his *Is Arithmetic Consistent?* (1994) and developed also in other texts (Priest [2006b]).

In *The Logic of Paradox*, Priest developed an argument, grounded in the notion of naïve proof, to the effect that Gödel's first incompleteness theorem would suggest

1 Priest uses the terms 'dialetheiae' and 'true contradictions' to indicate 'gluts'. propositions both true and false, a term coined by K. Fine in Fine [1975]. For an introduction to dialetheism, see e.g. Berto [2007].

2 A discussion on the same topic is in Beall [2009], Colyvan [2009], and Weber [2010]. For a short general introduction to the topic, see Murzi and Carrara [2015].

Massimilano Carrara: FISPPA Department – Section of Philosophy, Padua University

the presence of dialetheiae in the standard model of arithmetic. Chihara [1984], Shapiro [2002], and others Berto [2009], for example, criticised the argument. Much of the criticism was directed against the notion of naïve proof itself, in particular against the thesis that everything that is naively provable is true. Surely, if the notion of naïve proof is understood as embracing all proofs performed by real working mathematicians, as Priest seems to suggest, the thesis is hardly tenable. The aim of the paper is to show that from a notion of *naïve proof* – dialetheically acceptable – *trivialism* follows.

Finally, I would like to point out here that part of Sergio Galvan's research was on naïve proof and Gödel's incompleteness theorems.[3] I hope this paper can help others recognise the importance of these arguments and their implications for logic and its philosophy. In so doing, I hope to follow some steps of Sergio Galvan's research.

2 Curry's paradox and its arithmetical formalisation

Curry's paradox belongs to the family of so-called paradoxes of self- reference (or paradoxes of circularity). In short, the paradox is derived from natural-language from sentences like (a):

(a) If sentence (a) is true, then Santa Claus exists.

Suppose that the antecedent of the conditional in (a) is true, i.e. that sentence (a) is true. Then, by MP, Santa Claus exists. In this way, the consequent of (a) is proved under the assumption of its antecedent. In other words, we have proved (a). Finally, by MP, Santa Claus exists.

Of course, we could substitute any arbitrary sentence for 'Santa Claus exists', for example, that 'you will win the lottery', etc., which means that every sentence can be proved: from Curry's paradox, *trivialism* follows. Priest (1979, IV. 5) observes that, in a semantically closed theory, using MP and ABS

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

a version of Curry's paradox is derivable. In what follows, I reconstruct his argument in the language of first order arithmetic with a truth predicate.[4] Let L be the

3 The notion of naïve proof and Gödel's incompleteness theorems have been developed by Sergio Galvan in particular in Galvan 1983, Galvan 1992, pp. 183-202).

4 I follow here Carrara and Martino [2011], and Carrara et al. [2011].

language of first order arithmetic and N be its standard model. Extend L to the language L^* by introducing a new predicate T. With reference to a codification of the syntax of L^* by natural numbers, extend N to a model N^* of L^* by interpreting T as the truth predicate of L^*, so that, for all $n \in N$, $T(n)$ is true iff n is the code of a true sentence A of L^*, in symbols $n = \ulcorner A \urcorner$.

Of course, classically, such an interpretation is impossible, because the theory obtained by adding to Peano arithmetic (PA) the truth predicate for the extended language L^* with Tarski's biconditionals is inconsistent. Not so for dialetheism, where inconsistent models are accepted. But if one uses the classical rules of the conditional in natural deduction (from which ABS is derivable) and Tarski's scheme, i.e.:

$$T(\ulcorner A \urcorner) \leftrightarrow A$$

the model N^* turns out to be trivial. Let A be any sentence of L^*. By diagonalisation, there is a natural number k such that

$$k = \ulcorner T(\underline{k}) \to A \urcorner$$

We can derive A as follows:

1	(1)	$T(\underline{k}) \leftrightarrow \ulcorner T(\underline{k}) \to A \urcorner$	Tarski's schema
2	(2)	$T(\underline{k})$	Assumption
1, 2	(3)	$T(\underline{k}) \to A$	1, 2 MP
1, 2	(4)	A	2, 3 MP
1	(5)	$T(\underline{k}) \to A)$	2, 4 $\to$ Introduction
1	(6)	$T(\underline{k})$	1, 5 MP
1	(7)	At	5, 6 MP

Priest blocks this derivation in LP by rejecting the general validity of MP. According to him, this rule is not valid but *quasi-valid*, i.e. valid insofar as no dialetheia is involved. Priest (2006a), in order to reject MP, identifies, in the object language, $(A \to B)$ with $(\neg A \vee B)$. Then, the rejection proceeds as follows:

Proof. Suppose that A is a dialetheia; $(\neg A \vee B)$ is true even if B is not. In this case, if you infer B from A and $(A \to B)$, you get from true premises a not-true conclusion. This shows that MP may fail to preserve truth. Thus, the possibility of dialetheiae justifies the rejection of MP. qed

In the next sections I first introduce a notion of *naïve proof* (§3), then I argue (§4) how it is possible to obtain Curry's paradox using it, without adopting MP.

3 On naïve proofs

Let us consider *naïve* proof, a notion introduced by Priest in his *The Logic of Paradox* (1979) in order to argue that Gödel's first incompleteness theorem would suggest the presence of dialetheiae in the standard model of arithmetic. Priest describes the *naïve* notion of *proof* as follows:

> Proof, as understood by mathematicians (not logicians), is that process of deductive argumentation by which I establish certain mathematical claims to be true. In other words, suppose I have a mathematical assertion, say a claim of number theory, whose truth or falsity I wish to establish. I look for a proof or a refutation, that is a proof of its negation. [...] I will call the informal deductive arguments from basic statements *naïve proofs*. [Priest, 2006b, p. 40]

The alleged paradox should be suggested by the analogy of the familiar *informal proof* of Gödel's undecidable sentence G with the liar's paradox:

> As is clear to anyone who is familiar with Gödel's theorem, at its heart there lies a paradox. Informally the 'undecidable' sentence is the sentence 'this sentence is not provable'. Suppose that it is provable; then, since whatever is provable is true, it is not provable. Hence it is not provable. But we have just proved this. So it is provable after all (as well). (Priest [2006b], p. 237)

This argument is well known and widely discussed in the literature. Following Dummett, we can call it the *simple proof* argument (Dummett [1959]). It is worth noticing that the *simple proof* holds even dialetheically. Some people have maintained that this proof implicitly uses the consistency of PA, which, according to Gödel's second incompleteness theorem, is formally unprovable. Observe, however, passim that the proof at issue does not assume the consistency of PA, but by virtue of the fact that what is provable is true, dialetheically, consistency does not follow. Nor does the proof exploit *RAA*, but the tertium non datur, which dialetheically holds. The proof runs as follows:

Proof. G is provable or not provable. But if it is provable, then it is true and hence, as it says, (also) unprovable. In any case, it is unprovable and hence true. qed

Priest holds that the naïve notion of proof of an L-sentence is recursive so that the predicate 'P' of *naïve provability* is arithmetic. It follows that the relative Gödel's sentence G is a dialetheia. Priest's argument for the claim that the notion of naïve proof is recursive rests on the observation, supported by Dummett, that any mathematical proof of a sentence is recognisable as such. So, the argument goes, given a sentence A and a finite sequence p of formulas, one can decide whether p is a proof of A or not. It follows, by Church's thesis, that the relation between a proof and its conclusion is recursive. Priest's conclusion presupposes the following:

(i) naïve proofs form a well-determined set codifiable by a set S of natural numbers, and
(ii) there is a mechanical procedure for deciding whether a number belongs to S or not.

I am not going to discuss the evidence for either (i) and (ii). I would like to just consider naïve proof as a dialetheically acceptable notion and see what happens in terms of self-reference paradoxes, specifically in the case of Curry's paradox.

4 Naïve proofs and Curry's paradox

The new version of Curry's paradox[5] here proposed is obtained without making use of MP. I just make use of the notion of *naïve poof*. Consider the extension L' of the language L of first order arithmetic, obtained by introducing a new predicate $P(x)$. Extend the standard model N of arithmetic to the model N', where P is interpreted as *naïve provability* for the language L' (with reference to a numerical codification of the syntax of L'). More precisely, $P(\ulcorner A \urcorner)$ means: It is naïvely provable that A is true in N.

In *Naïve Proofs*, Priest observes, "It is analytic that whatever is naïvely provable is true. Naïve proof is just that sort of mathematical argument that establishes something as true. And since this is analytic, it is itself naively provable [...]" (Priest [2006b], p. 238). Moreover, he argues, "If something is naïvely proved then this fact itself constitutes a proof that A is provable" (Priest [2006b], p. 238).

On the basis of the above remarks, one can argue that:

(a) $P(\ulcorner A \urcorner) \to A$ is naively provable;
(b) If A is naïvely provable, then $P(\ulcorner A \urcorner)$ is naïvely provable.

Similarly, let us extend L to a language L^* by introducing a binary predicate $D(x, y)$. Then, we can extend the standard model N to the model N^* of L^*, where D is interpreted as the naïve deducibility relation for L^* (with reference to a codification of L^*). $D(x, y)$ means the following:

– y is naively deducible from x.

Or, in more explicit terms:

– There is a naïve proof that, assuming that x is true in N^*, leads to the conclusion that y is true in N^*.

5 A version of this paradox is in Carrara and Martino [2011] and Carrara et al. [2011].

Consider a natural *deduction* system. Rules of *elimination* and *introduction* for D, analogous to (a) and (b) stated above, are as follows:

(DE) From premises A and $D(\ulcorner A \urcorner, \ulcorner B \urcorner)$, one can derive B. The conclusion depends on all assumptions upon which the premises depend.

(DI) From premise B, depending on the unique assumption A, one can infer $D(\ulcorner A \urcorner, \ulcorner B \urcorner)$, discharging A.

Theorem. *From* (DE) *and* (DI), *trivialism follows.*

Proof. Let A be any L^*-sentence. By diagonalisation, we get a natural number k such that $k = \ulcorner D(\underline{k}, \ulcorner A \urcorner) \urcorner$. Using $\underline{k}$ as a name of $D(\underline{k}, \ulcorner A \urcorner)$, suppose that $\underline{k}$ is true. Since $\underline{k}$ says that A is deducible from $\underline{k}$ and deduction is sound, A is true. So we have proved A from the assumption $\underline{k}$. Hence $D(\underline{k}, \ulcorner A \urcorner)$, i.e. $\underline{k}$, is true. And, since deduction is sound, A is true. qed

A formal proof of A in natural deduction (where $\underline{k}$ is used again as a name of $D(\underline{k}, \ulcorner A \urcorner)$ is as follows.

1	(1)	$\underline{k}$	Assumption
1	(2)	$D(\underline{k}, \ulcorner A \urcorner)$	1, Identity
1	(3)	A	1,2 DE
	(4)	$D(\underline{k}, \ulcorner A \urcorner)$	1, 3 DI (discharging (1))
	(5)	$\underline{k}$	4, Identity
	(6)	A	4, 5 DE

Since A is arbitrary, N^* is trivial. But N^* differs from N only for the introduction of the relation of naïve deducibility: the arithmetical sentences of L are interpreted in N^* as in N. Therefore N is trivial as well.

5 Conclusion

In this paper, I took a notion of *naïve proof*, defended by Priest in his discussion of Gödel's theorem. I consider his characterisation of the notion *via* (a) and (b). By using it, a new version of Curry's paradox is proposed, obtained without making use of MP.

Acknowledgments

I would like to thank Enrico Martino because much of the theoretical work done in this paper was developed in cooperation with him, Ciro De Florio for inviting me to write a paper for this volume, for his patience, support and friendship, and, finally, an anonymous reviewer. I met Sergio Galvan when I was no longer a young boy (!). I was lucky to have had the opportunity to know him and to count on his friendship. It was a pity to have known him so late. I usually feel like one of David Foster Wallace's young fish, and I have the feeling that things would have been better for me if I had met Sergio earlier. Anyway, *this is water*.[6]

6 'There are these two young fish swimming along, and they happen to meet an older fish swimming the other way, who nods at them and says, "Morning, boys, how's the water?" And the two young fish swim on for a bit, and then eventually one of them looks over at the other and goes, "What the hell is water?"' (David Foster Wallace, *This is Water: Some Thoughts, Delivered on a Significant Occasion, about Living a Compassionate Life*. Little, Brown and Company, 2009, 1).