Deflationism and the Gödel Phenomena: Reply to Cieśliński

by

Neil Tennant*

Department of Philosophy
The Ohio State University
Columbus, Ohio 43210
email tennant.9@osu.edu

March 20, 2009

Abstract

We clarify how the requirement of conservative extension features in the thinking of various deflationists, and how this relates to another litmus claim, that the truth-predicate stands for a real, substantial property. We discuss how the deflationist can accommodate the result, to which Cieslinski draws attention, that non-conservativeness attends even the generalization that all logical theorems in the language of arithmetic are true. Finally we provide a four-fold categorization of various forms of deflationism, by reference to the two claims of conservativeness and substantiality. This helps to clarify the various possible positions in the deflationism debate.

In offering the following considerations on behalf of deflationism, the present author is continuing to act as an *amicus curiae*, rather than as a contending party.[1] For he is a substantialist about truth, believing that the truth of a sentence consists in the existence of a truth-maker (and that falsity

*For helpful comments on an earlier draft, the author is grateful to Salvatore Florio, Robert Kraut, Patrick Reeder, Kevin Scharp and Stewart Shapiro. Paul Horwich was kind enough to clarify aspect of his position in correspondence. Remaining defects are the author's sole responsibility.

[1]For the original brief, see Tennant 2002.

consists in the existence of a falsity-maker). Those philosophical convictions, however, do not turn on the issue of whether a deflationary construal of the truth-predicate is possible in light of the Gödel phenomena.

1 The assumption of conservative extension

For the reader not familiar with the debate at hand, it is worth emphasizing an assumption common to both parties: that it is criterial, for a theory of truth to be deflationist, that it be conservative over one's non-semantic theorizing. That is, no unprovable assertion or underivable inference in the language L of one's non-semantical theorizing becomes provable or derivable, respectively, upon the adjunction of one's theory of truth for L. This *assumption of conservative extension* (on behalf of deflationism) made its first explicit appearance in Horsten 1995, which examined Paul Horwich's minimalist theory of truth (Horwich 1990), 'widely regarded as the deflationist theory that has been developed in most detail and that has been given the clearest expression to date' (p. 175). Horsten writes at p. 183:

> The minimalist theory entails that a truth predicate should be *conservative* over a given theory that is stated without the truth predicate (or any other semantical notions).

He adds immediately (fn. 15):

> Maybe I am wrong here. But if I am, then I do not see what the neutrality of the notion of truth according to deflationism amounts to. In any case, this is an issue which I think should be addressed by defenders of deflationist theories of truth.

Later, in discussing Feferman's axiomatic system KF (which seeks to govern the behaviour of a truth predicate added to the language of Peano arithmetic), Horsten writes (p. 184):

> ... there are purely arithmetical sentences that are provable in KF, but not in Peano arithmetic. The notion of truth inflates the mathematical theory to which it is applied, and this seems to accord ill with *any* version of the deflationary theory of truth.

Shapiro 1998 appears to share the view that the deflationist ought to be making the assumption of conservative extension (p. 497):

> If truth/satisfaction is not substantial—as the deflationist contends—then we should not need to invoke truth in order to establish any new results not involving truth explicitly.

But in order to pin the assumption in question on the deflationist, the consequent of Shapiro's conditional needs to read 'we should not be able, by invoking truth [theory], to establish any results not involving truth explicitly'. For there are other ways—such as the truth-predicate-free reflection principles discussed in Tennant 2002—by means of which one can establish new results not involving truth explicitly.

Ketland 1999 is another author who attributes the assumption of conservative extension to the deflationist. He even offers, as the title for his formal Theorem 1 (p. 76), 'The conservativeness of deflationary truth theories'.

The assumption of conservative extension may, of course, be challenged by either a deflationist (such as Field 1999) or a non-deflationist.[2] In a private communication dated 29 November 2008, Horwich demurs on the imputation of conservativeness:

> . . . it doesn't in the slightest bother me that the minimal theory is not conservative (assuming it isn't); . . . perhaps a conservative theory of truth would, on that account, be "more deflationary" than mine (as would, perhaps, the view—which I reject—that there is no property of truth). But so what; why shouldn't the correct theory of truth be non-conservative?

Horwich also thought it appropriate to

> draw attention to the fact (a) that my account (pp. 23-5 of the 2nd edition) of what it takes for a theory of truth to be adequate does not include conservativeness, and (b) that I never allude to conservativeness—and, in particular, never suggest that my minimal theory is conservative.

[2] For a useful discussion, see Halbach 2001. But note that Halbach's quote on his p. 168 from Field 1999:

> [...] there is no need to disagree with Shapiro when he says "conservativeness is essential to deflationism"

is taken slightly out of context, and should not have been used to back the suggestion that Field 'has come close to confessing a commitment to conservativeness'. For Field goes on to disagree with Shapiro's claim that even 'full-blooded' extension (which allows the truth-predicate to occur in instances of mathematical induction) should, for the deflationist, be conservative.

The question whether a truth-theorist (deflationist or otherwise) should make the assumption of conservative extension was perhaps first raised by Sheard 1994 (p. 1053):

> Are conservative extension results desirable? That is, does the fact that a particular theory [involving the truth predicate] is a conservative extension of the base theory in a language without the truth predicate indicate that we have selected the "appropriate" additional axioms for truth itself, or should we expect the introduction of a theory of truth to provide more formal power with respect to the underlying domain?

So we see that the assumption of conservative extension has been questioned both in its own right, and as criterial for deflationism. For the purposes of the present discussion, however, the assumption will be in place. The task at hand is to examine the extent to which certain non-conservativeness results reported in Cieśliński 2008 might undermine the defence of the deflationist that was offered in Tennant 2002.

2 Cieśliński's results

Suppose one is dealing with an axiomatic system—call it S—such as Peano Arithmetic, subject to Gödel's Second Incompleteness Theorem.[3]

Cezary Cieśliński (2008) addresses the present author's question (in Tennant 2002) as to why the deflationist should be saddled with a commitment to the soundness claim

All S-theorems are true,

expressed by means of semantical vocabulary (the truth predicate). The problem with the soundness claim, for the deflationist, is that it extends S non-conservatively (since, as pointed out by Shapiro 1998, it affords a proof of Con_S, the sentence of arithmetic, involving no semantical vocabulary, that expresses the consistency of S). Cieśliński is not satisfied with the present author's suggestion that the deflationist could instead express the conviction that S is sound by adding to S the schematic principle

$$Pr_S(\overline{\varphi}) \rightarrow \varphi.$$

[3]Here we continue with the notational convention in Tennant 2002, which was adopted by Cieśliński 2008. But Cieśliński also brings into the discussion a new sense for 'S', which needs to be distinguished from this one. See the digression on notation, below.

This principle is formulated without using any semantical vocabulary. Its instances are obtained by replacing the placeholder φ by sentences in the language of arithmetic. Adding those instances to S produces the so-called soundness extension of S.

Cieśliński argues that this reflection strategy 'does not help the deflationist'; but that in a modified form, which he proceeds to develop, it does. To this end, Cieśliński proves

Theorem 1 $PA(S)^{-} + \forall\psi[Pr_{\emptyset}(\psi) \Rightarrow Tr(\psi)] \vdash \forall\psi[Pr_{PA}(\psi) \Rightarrow Tr(\psi)]$

The system $PA(S)^{-}$ conservatively extends S, and therefore might commend itself to the deflationist.

Digression on notation.[4] Here, $PA(S)^{-}$ is 'the theory obtained from PA by adding the usual Tarski's clauses [*sic*] as new axioms', but with 'the induction schema ... restricted to arithmetical instantiations only'.

Note that the occurrence of 'S' in '$PA(S)^{-}$' derives from a notation borrowed from Kotlarski 1986, in which S is a 'full satisfaction class', consisting of pairs of the form $\langle\varphi, a\rangle$, where φ is a formula of the language of arithmetic and a is a sequence of individuals respectively coordinated with its free variables. The individuals are drawn from the domain of whatever model the satisfaction class is 'for'. The gloss 'S is a full satisfaction class' is just a technical way of abbreviating the conjunction of clauses involved in the familiar Tarski-style definition of the relation S of satisfaction (and its special case of truth, which is satisfaction by the null sequence), for the language in question (here, the language of arithmetic).

At p. 532 Kotlarski defines the theory $PA(S)$ to be

> PA + S is a full satisfaction class + induction in $L_{PA} \cup \{S\}$

This obviously involves treating 'S' as though it is a new primitive two-place predicate. That allows one to take S-involving instances (i.e. formulae in the extended language $L_{PA} \cup \{S\}$) of the induction axiom-scheme. It also allows one (*via* appropriate arithmetization of syntax, and in a strong enough arithmetical theory) to define the notion 'S is a full satisfaction class'. Kotlarski goes on to observe

> Let PA^{-} be the theory in L_{PA} consisting of recursive definitions of addition and multiplication and the definition of inequality.

[4]In both our discussion and quotations, we shall use notational italics even when the author being quoted might have used roman.

> It is well known that PA^- is sufficiently strong to represent all recursive functions ..., so as to arithmetise the language. Thus PA^- allows us to state the definition of a full satisfaction class. Let T be the following theory:
>
> $$PA^- + S \text{ is a full satisfaction class} + \forall\varphi((PA \vdash \varphi) \rightarrow S(\varphi)).$$
>
> It is convenient to read the last sentence as "S makes all the theorems of PA true".

Thus for Kotlarski, $S(\varphi)$ is short for $S(\varphi, \emptyset)$ (where S is treated as the *binary* predicate that it is), which in turn could be rendered as $Tr(\varphi)$. The arithmetised definition of 'S is a full satisfaction class' will have clauses mimicking the well-known satisfaction theory of Tarskian form.[5]

So we see that 'S', at all its occurrences in this digression, would have been better rendered as, say, 'Sat', so as to distinguish it from the use of 'S' as a placeholder for a sufficiently strong theory of arithmetic.[6]

End of digression

---

[5]These clauses will actually be embedded conjuncts, since 'S is a full satisfaction class' is captured by a single sentence.

[6]Kotlarski goes on to prove (Theorem 2.2, p. 532) that the theory T as he defines it is identical to the theory Δ_0-$PA(S)$, which he defines as

$$PA + S \text{ is a full satisfaction class} + \Delta_0\text{-induction in } L_{PA} \cup \{S\}.$$

This is the only theorem in Kotlarski's paper that appears to be a candidate for the reference 'Theorem 2 (Kotlarski 1986)' in Cieśliński's paper. Yet Cieśliński's version has it that Kotlarski proved Δ_0-$PA(S)$ to be identical to

$$PA(S)^- + \forall\psi[Pr_{PA}(\psi) \Rightarrow Tr(\psi)].$$

Since $PA(S)^-$ is defined by Cieśliński to be PA with induction restricted to arithmetical formulae, plus the 'usual Tarski clauses' for S, it follows that the latter theory is, by definition,

$$PA^- + \text{arithmetical induction} + S \text{ is a full satisfaction class} + \forall\psi[Pr_{PA}(\psi) \Rightarrow Tr(\psi)].$$

It is not immediately obvious that this theory is identical to Kotlarski's theory T:

$$PA^- + S \text{ is a full satisfaction class} + \forall\varphi((PA \vdash \varphi) \rightarrow S(\varphi)) .$$

Obviously the former theory implies the latter, given that the truth-predicate Tr is a special case of S. For the converse, however, we need to show that the latter theory implies all arithmetical instances of induction. So let θ be such an instance. Since S is a full satisfaction class, we shall be able to derive the 'disquotational' principle $S(\theta) \Rightarrow \theta$. Clearly $PA \vdash \theta$. Hence, since $\forall\varphi((PA \vdash \varphi) \rightarrow S(\varphi))$, we have $S(\theta)$. Now, since S is a full satisfaction class, we can 'disquote' to conclude θ.

It turns out, however, that the present discussion does not depend on Cieśliński's Theorem 2. It is enough to focus on Cieśliński's Theorem 1.

Cieśliński bases his criticism of reflective deflationism on the requirement that

(**R**) The deflationist should have at his disposal a theory which proves the basic, sound instances of (*),

where (*) is the formal schema

$$\forall\psi[\alpha(\psi) \Rightarrow Tr(\psi)]$$

whose instances would be obtained by replacing $\alpha(x)$ by an arithmetical formula defining a set A of sentences whose acceptance we wish to express. Cieśliński claims that

> The reason behind (**R**) is that if it is not satisfied, the truth predicate still seems useless (contrary to what the deflationist claims).

And he goes on to say that the deflationist's philosophical views (about truth) would be undermined by his inability to explain our acceptance of 'examples of most basic, intuitive generalizations' (presumably of the form (*)).

3 Rebuttal

What sort of inability is in question here? Must the deflationist be able to explain our acceptance of

(i) *all* examples of [the] most basic, intuitive generalizations of the form (*)?

Or would it be enough if the deflationist were able to explain our acceptance of

(ii) at least *some* of these examples, even if not all of them?

We have already seen, in light of the Gödel phenomena, that the deflationist cannot lay claim to the following example of the form (*):

$$\forall\psi[Pr_S(\psi) \rightarrow Tr(\psi)]$$

on pain of having his use of the truth-predicate Tr produce a non-conservative extension of the system S.

The problem, for the deflationist, becomes one of maintaining conservativeness (of extension of the system S by truth-talk) while yet being able to explain the utility of a truth-predicate in terms of the generalizations that it affords its users—generalizations not readily expressible in any other way. The main concern of Tennant 2002 was to argue that the assertion of a soundness generalization about a consistent (and sufficiently strong) system S could take the form of a schematic reflection principle. And although such reflection extends the system S in the original arithmetical language, that fact is benign, for that is just what reflection does—it produces warrant for new claims, in the original language, that one did not possess before engaging in the reflection. The truth-predicate itself (if we have one) would be playing no role in generating these further insights. (Of course, it might later play a role in *expressing* these insights in a more economical form.)

So the question is whether one can accord the truth-predicate an interesting enough role in producing desired generalizations, without such use of the truth-predicate itself affording warrants for new claims in the original language. Now, Cieśliński claims, in fn.16, that neither Horwich nor the present author specifies any plausible candidates for alternative functions that the truth-predicate might have, other than its permitting us to formulate such generalizations as 'for every x, if x is a proposition of the form $\langle p \Rightarrow p\rangle$, then x is true'. But it is not hard to think up such alternative functions. How about 'For every sentence x, if the Pope utters x assertively, then x is true'? Or 'for every sentence x, if x is not true, then the negation of x is true'? For the deflationist's theory of truth can encompass (as theorems) not only the well-known Tarskian biconditionals[7] of the form

$$\varphi \leftrightarrow Tr(\overline{\varphi})$$

but also any results that follow from the basic rules of inference laid out in Tennant 1987, pp. 71–3, governing the distribution of truth over the logical operators. An example of such a rule is

$$\frac{Tr(\overline{\varphi}) \quad Tr(\overline{\psi})}{Tr(\overline{\varphi \& \psi})} \text{ where } \varphi,\ \psi \text{ are in the unextended language}$$

So the present author concludes, on behalf of the deflationist, that there is plenty of useful work to be done by a truth-predicate, short of so deploying it as to obtain non-conservative extensions of theories of ours that do not use it.

[7]For a deflationist such as Horwich, these biconditionals are taken as axiomatic. But it is also open to the deflationist to take as his deductive starting points the rules of truth-distribution mentioned here, and to derive the Tarskian biconditionals as theorems.

Cieśliński wishes, however, to be more venturesome on behalf of the deflationist. He sees in his Theorem 1 'an example of a basic generality of the required sort, not provable in any conservative truth theory'. And he suggests that, confronted with this, the present author's strategy would '[amount] to rejecting (**R**)'.

The trouble, however, is that, given the deflationist's commitment to conservativeness, Theorem 1 *cannot* be taken as 'an example of a basic generality of the required sort'. Required *by whom and for whom?*, one might ask. We have already seen that the deflationist cannot, and will not, use his truth-predicate to say that all S-theorems are true; for that would be to give up conservativeness. Why, then, should the deflationist be willing to assert any other alleged 'basic generality' involving truth, if that puts him in the same predicament? The deflationist is free to take a judicious path between the Scylla of saying or explaining *nothing* of interest about our use of the truth-predicate, and the Charybdis of saying *too much*, truth-theoretically, for the truth-predicate to be construable in a deflationary way. So it is up to the deflationist to make judicious choices of generalizations about truth, in such a way as to steer a middle course. He is already avoiding Scylla, by undertaking the commitments we have already mentioned. All we have to do, now, is ensure that he does not founder on Charybdis. Moreover, generalizations 'about truth' that he cannot afford to make must find some other, equally serviceable form of deflationarily licit expression (as is the case with reflection principles).

In this connection it is worth asking exactly why Cieśliński thinks that his Theorem 1 is indeed an example of a basic generality [about truth] of the sort he has in mind, and exactly why it is not provable in any conservative truth theory. The generality's not being provable in any conservative truth theory is enough for the deflationist to deny that he (the deflationist) ought to be able to prove it (rather than find a deflationarily licit, equally serviceable expression for it). Note that Theorem 1 states an *inferential connection*. From the conservative extension $PA(S)^-$ (in which the induction axiom-scheme is allowed no instances involving the satisfaction predicate S—hence no instances of its special case, the truth-predicate Tr) plus the generalization that all theorems of logic (in the language of arithmetic) are true ($\forall\psi[Pr_\emptyset(\psi) \Rightarrow Tr(\psi)]$), one can infer that all theorems of PA are true ($\forall\psi[Pr_{PA}(\psi) \Rightarrow Tr(\psi)]$). The latter claim, as we have already seen, is one that the deflationist must avoid making. Hence, if he commits himself to the conservative extension $PA(S)^-$ of PA, then he must *also* avoid committing himself to the most obvious regimentation of the generalization that all theorems of logic (in the language of arithmetic) are true

($\forall\psi[Pr_{\emptyset}(\psi) \Rightarrow Tr(\psi)]$). All that Theorem 1 teaches us is the interesting lesson (for the deflationist) that—should he wish to assert all of $PA(S)^{-}$—he will have to rely on the reflection strategy even in order to express his faith in 'the truth of' all theorems of the *logic* of the language of arithmetic. That is, he will have to express his faith by means of a reflection principle:

$$Pr_{\emptyset}(\varphi) \rightarrow \varphi.$$

This, however, is already at hand; for he is already committed to the reflection principle

$$Pr_{PA}(\varphi) \rightarrow \varphi,$$

and he can also prove

$$\forall\psi[Pr_{\emptyset}(\psi) \rightarrow Pr_{PA}(\psi)],$$

where the universal quantification is over all sentences ψ of the language of arithmetic.

Notwithstanding the interest of Cieśliński's Theorem 1, it does not appear to the present author to alter significantly the dialectical situation in which the Gödel phenomena have already placed the deflationist. The reflection strategy, eschewing use of a truth-predicate, is adequate to what is now seen as simply a wider task. The deflationist is not, for want of a substantive theory of truth, unable to express things that substantive truth-theorists happily express by means of their truth-predicate. The deflationist can still point to useful generalizations that the use of a deflationarily construed truth-predicate allows us to make. In particular, he can point out that he can still commit himself to all of $PA(S)^{-}$.

We have not yet been offered any clinching consideration, on behalf of the substantive truth-theorist, against the deflationist's position. It might then be thought that the best way forward, for the substantive truth-theorist, would be a pragmatic one: to point to the fruitfulness of truth-talk, and the unification of reflective insights that it affords. But that ground is already occupied by the 'robust deflationist' of Kraut 1993. So, the tentative conclusion is that the substantial character of truth will have to be vouchsafed by arguments establishing the intellectual appeal of, or theoretical need for, more than just 'saving the phenomena' of our truth-talk.

4 Further reflections

The considerations entered thus far suggest a way of re-configuring the debate over deflationism that might be helpful for the future direction of these discussions. It strikes the present author that there are two distinctions at work, which are worth clarifying, since they generate four distinct positions. One of the distinctions can be thought of as a logical one, and the other as metaphysical.

First, the logical distinction. Either one asserts

> *Truth-theory must be conservative over non-semantical theorizing*

(a claim that we have been treating as an assumption held in common by both parties to the current debate thus far); or one asserts its contradictory:

> *Truth-theory need not be conservative over non-semantical theorizing.*

Secondly, the metaphysical distinction. Either one asserts, in a realist, descriptivist, or factualist vein that

> *The truth-predicate stands for a real, substantial property*;

or one asserts its irrealist, projectivist, or anti-factualist contradictory:

> *The truth-predicate does not stand for a real, substantial property.*

Let us imagine that the making of the second distinction is not parasitic upon the first: that is, one's preferred metaphysical claim (about the substantiality of truth) is not a mere metaphorical picture wholly explicable in terms of one's preferred logical claim (about conservativeness). Let us also assume the converse: that is, one's preferred logical claim does not determine one's preferred metaphysical claim.

It then follows that each of the four cells of the box below can be, and indeed is, occupied by at least one theorist, as indicated:[8]

[8] With an eye to the following diagram: are there any Conservative Deflationists? The evidence suggests that there are. At p. 12 Horwich assures us that the theory of truth consists in 'an infinity of biconditionals of the form, $\langle$p$\rangle$ is true *iff* p'. That guarantees conservativeness (provided that the truth predicate is not allowed to feature in instances of the induction axiom schema of arithmetic). Moreover, although at p. 38 he writes

> ... it is not part of the minimalist conception to maintain that truth is not a property. On the contrary, 'is true' is a perfectly good English predicate— and ... one might well take this to be a conclusive criterion of standing for a property of *some* sort[,]

he goes on to downplay what this might entail:

> What the minimalist wishes to emphasize, however, is that truth is not a *complex* or *naturalistic* property but a property of some other kind. ... different kinds of property correspond to different roles that predicates play in our language.

	Truth-theory must be conservative over non-semantical theorizing	Truth-theory need not be conservative over non-semantical theorizing
The truth-predicate stands for a real, substantial property	Conservative Anti-Deflationist (Tennant)	Non-Conservative Anti-Deflationist (Shapiro)
The truth-predicate does not stand for a real, substantial property	Conservative Deflationist (Horwich*)	Robust Deflationist (Kraut) Non-Conservative Deflationist (Field)

Tennant and Kraut take issue with Shapiro in his opposition to the Conservative Deflationist. But they are not wholly allied with Horwich*. Tennant disagrees with Horwich* about whether the truth predicate expresses a substantial property. Kraut disagrees with Horwich* about whether a theory of truth must be conservative over non-semantical theorizing. Moreover—and ironically—Tennant and Kraut disagree on both fronts. Nevertheless, each provides an alternative to Horwich*'s Conservative Deflationist line.

In this classification, the present author is labeled a Conservative Anti-Deflationist. This is because, as an anti-realist, he maintains that the truth of a sentence φ consists in the existence of a truthmaker for φ. This is a substantial matter, something more than possession of a mere 'logical' property. Tennant 2002 was at pains to point out this important difference in view from the Conservative Deflationist, while contending nevertheless that on the *other* score—the question of conservativeness of truth-theory—the

It is reasonable to interpret Horwich's words here in a way that places him in the second row of our box. But, given the interpretive slack involved, and especially his recent demurral over conservativeness reported above, we shall err on the side of exegetical caution and call this representative Horwich*. The author is assured by Halbach (private communication) that he would not wish to be boxed with Horwich*.

Conservative Deflationist could be defended against the criticisms of Shapiro 1998 that were based on the Gödel phenomena. (The Non-Conservative Deflationist—see Field 1999—takes himself to be an unjustified target of those criticisms.)

The question now arises: how might, or why should, one who takes truth to be substantial nevertheless maintain that truth-theory should be (or even *can* be) conservative over one's non-semantical theorizing? The *amicus* brief on behalf of the Conservative Deflationist has stressed the importance of truth-predicate-free *reflection principles* as the legitimate source of those theoretical expansions that would otherwise have to be effected by means of theoretical principles involving a truth-predicate. We can attain the sought Gödelian insights—which threaten non-conservativeness—*without* recourse to a truth-predicate, and thereby avoid that threat altogether. Whatever generalizations about truth the Deflationist truth-theorist wishes to make, can be made without the threat of non-conservative extension.

The tendency to deploy a truth-predicate T rather than resort to T-free reflection principles can be explained by two considerations. First, Tarskian theorizing about truth (in formalized languages, such as that of arithmetic) preceded the account of reflection principles first put forward by Feferman 1962. So the preferred resort to principles involving a truth-predicate is a matter of history and habit. Secondly, reflection principles are of schematic form; and the temptation is strong to codify or express them as single sentences. *That* is where the truth-predicate reveals its utility. It is ideally suited for the 'making explicit'[9] of those theoretical commitments that would otherwise remain only implicit in the theorist's willingness to assent to (the potential infinity of) instances of a reflection principle, *seriatim.* As soon as that becomes one's preferred way of expressing the convictions arising from reflection on one's means of proof, then *of course* the newly available talk of truth will be non-conservative—for it merely 'makes explicit' the means whereby one effects the very non-conservative extension that one had originally sought!

References

Robert B. Brandom. *Making It Explicit: Reasoning, Representing, and Discursive Commitment.* Harvard University Press, Cambridge, MA, 1994.

[9]See Brandom 1994, 2000.

Robert B. Brandom. *Articulating Reasons: An Introduction to Inferentialism.* Harvard University Press, Cambridge, MA, 2000.

Cezary Cieśliński. Truth, Conservativeness and Provability. *Mind*, XXX (XXX):XXX–XXX, 2008.

Solomon Feferman. Transfinite recursive progressions of axiomatic theories. *Journal of Symbolic Logic*, 27(3):259–316, 1962.

Hartry Field. Deflating the Conservativeness Argument. *Journal of Philosophy*, 96(10):533–540, 1999.

Volker Halbach. How Innocent is Deflationism? *Synthese*, 126(1/2):167–194, 2001.

Leon Horsten. The Semantical Paradoxes, the Neutrality of Truth and the Neutrality of the Minimalist Theory of Truth. In P. Cortois, editor, *The Many Problems of Realism,* Vol. 3 of *Studies in the General Philosophy of Science*, pages 173–187. Tilburg University Press, Tilburg, 1995.

Paul Horwich. *Truth.* Basil Blackwell, Oxford, 1990.

Jeffrey Ketland. Deflationism and Tarski's Paradise. *Mind*, 108(429):69–94, 1999.

Henryk Kotlarski. Bounded Induction and Satisfaction Classes. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 32:531–544, 1986.

Robert Kraut. Robust Deflationism. *Philosophical Review*, 102(2):247–263, 1993.

Stewart Shapiro. Proof and Truth: Through Thick and Thin. *Journal of Philosophy*, XCV(10):493–521, 1998.

Michael Sheard. A Guide to Truth Predicates in the Modern Era. *Journal of Symbolic Logic*, 59(3):1032–1054, 1994.

Neil Tennant. *Anti-Realism and Logic: Truth as Eternal.* Clarendon Library of Logic and Philosophy, Oxford University Press, 1987.

Neil Tennant. Deflationism and the Gödel Phenomena. *Mind*, 111:551–582, 2002.