Deflationary truth: conservativity or logicality ?

Henri Galinon

[Draft of a paper published in the *Philosophical Quarterly*]

Henri Galinon

Deflationary truth: conservativity or logicality ?

Abstract: It has been argued in the literature that the deflationists' thesis about the dispensability of truth as an explanatory notion forces them to adopt a conservative theory of truth. I suggest that we should take more seriously the deflationist's claim that the notion of truth is akin to a *logical notion*. This claim casts some doubts on the adequacy of the conservativity requirement, while it also calls for further investigation to assess its philosophical plausibility.

1. What is deflationism ?

One central tenet of alethic deflationism, as defended e.g. by Horwich (1999) or Field (2001), is the thesis that the truth predicate has no explanatory power. It is rather, the deflationist claims, a purely expressive device. The distinction is suggestive, but it is not entirely clear where the boundary should be drawn between explanatory and non-explanatory notions, nor is it clear whether being non-explanatory is equivalent to being purely expressive. It is thus not surprising that deflationism's supporters and discontents seek to sharpen the point. In the following I want to discuss what is probably the best known attempt in this direction and argue rather for another perspective on the deflationist's claim.

2. Conservativity

The well-known attempt is due to Stewart Shapiro and Jeff Ketland. Shapiro (1998) and Ketland (1999) have suggested the following test against the deflationist's claim, which goes some ways as a clarification of the distinction :

> **Conservativity test**: the notion of truth is explanatory if an adequate theory of truth non-conservatively extends some non-semantic theory.

From that, the authors have argued against deflationism on the ground that the Tarskian truth-theoretical extension of Peano Arithmetic is adequate but non-conservative.[1]

As is well known, there is a slight logical subtlety behind this non-conservativity claim. The point is that the Tarskian truth-theoretic extension of PA (let us write T(PA)) , can be said to be non-conservative over PA only if PA is construed as a « schematic » theory. More precisely, let T(PA) be the Peano axioms extended by the Tarskian axioms for truth and the new substitution instances of the induction schema by formulas in the extended vocabulary of L_{PA+Tr} containing the truth predicate. And let $T^-(PA)$ be just the Peano axioms extended by the Tarskian axioms for truth (without the new instances of the induction scheme). Then we have : T(PA) is a non-conservative extension of PA, and $T^-(PA)$ is a conservative extension of PA.[2] Thus Shapiro and Ketland's claim is in fact that non-conservativity over PA of the Tarskian axioms for truth *and the new instances of the induction scheme* shows that truth is explanatory. I find this conclusion puzzling, and side with those who disagree[3]. But whereas some philosophers, especially Field (1999), are ready to endorse a cautious version of the conservativity test – a version that would distinguish clearly between extensions by axioms that are « essential to truth » and extensions involving altogether the new instances of the induction scheme -, I doubt that conservativity is suited at all for the purpose at hand, as I shall argue below.[4]

3. Logicality

[1] We say that a theory of truth is adequate if it accounts for the uses it has to account for. For Ketland and Shapiro, it is part of adequacy that the truth-theoretic extension of PA should allow one to prove that all the theorems of PA are true. I shall remain silent on the relevance of this requirement here.

[2] See e.g. Halbach (1999)

[3] In particular Field (1999) and Azzouni (1999).

[4] Let us note in passing that non-conservativity claims are traditionally involved in philosophical *indispensability* arguments within discussions of fictionalist or instrumentalist account of a notion, whereas the deflationist's claim about the truth predicate has *prima facie* no well-articulated connection to indispensability *simpliciter* : deflationism is rather explicitly, right from the start, a mixed claim involving explanatory dispensability and expressive indispensability. In the following, I shall argue for my interpretation of what this means and leave aside the assessment of the instrumentalist reading of deflationism.

My point is about what the deflationist is supposed to be claiming. Let us pause for a second and ask ourselves what kind of notion truth can be, that would be both explanatorily idle and indispensable for expressive purposes. Do we know of any notions that would fit the bill ? I think there are. Indeed it is even a relatively common philosophical view that logical notions are primary examples of such notions. This view of course dates back at least to the positivists and the logical atomism of Russell and Wittgenstein. Actually, contemporary deflationists themselves often phrase their thesis as the thesis that the truth predicate (or the notion of truth) is precisely a logical expression (a logical notion) or, alternatively, a predicate fulfilling only a logical role, something like a conjunction of a sort.[5] Conjunction, disjunction, quantifiers, identity, etc. are precisely examples of non-explanatory expressions that are just functional tools that the deflationist may have in mind: in old days parlance, they serve the purpose of articulating contents in discourse, as opposed to descriptive notions who bring new content into it.[6] I won't try to argue for the thesis that logical notions are not explanatory, although I find the view that logical notions are epistemically special attractive. My point here is simply to emphasize the explicit connection made by the deflationist between the purported non-explanatory nature of the notion of truth and its being akin to a logical notion, as opposed to an explanatory one. As a matter of fact, logicality is the main paradigm offered by the deflationist to clarify his point about the explanatory idleness of the notion of truth. Thus, following this line of thought, I think that deflationism about truth is best understood as the claim that truth is a logical notion rather than as a reductionist, instrumentalist or fictionalist claim about the notion of truth.

4. Conservativity and logicality

Now if we take the deflationist to claim that truth is a logical notion and that logical notions are expressive but non-explanatory notions, it seems that the conservativity test

---

[5] This is explicit for instance in Horwich (1998), p.2 and Field (2001). Field (2001), p.139 comes very close to claiming that the truth predicate, as understood by the deflationist, is « a logical predicate ».

[6] I should note here that Field himself may certainly not accept such a formulation, which fits better with the metaphysics of early logical atomism than with his professed physicalist methodology.

does not do justice to deflationism. Why ? Because it is known in various ways that logical notions may sometimes behave in the same way as the Tarskian notion of truth does, with respect to non-conservativity. For the purpose of comparison with the logical situation put forward by Shapiro and Ketland, consider first the following theory, PA_{At}, constituted by the Peano axioms with the restriction that only atomic sentences can be substituted for sentences letters in the induction schema. The « only » difference between PA and PA_{At} is thus that the logical constants are allowed to appear in formulas that instantiate the induction schema, whereas they are not allowed to in PA_{At}. On the face of it, the difference between PA and PA_{At} is thus analogous to the difference between T(PA) and T^-(PA), and consists in allowing or not allowing a given set of expressions, the usual logical constants in the first case and the truth predicate in the second case, to appear in sentences that instantiate the induction scheme. Now it happens that PA is not conservative over $PA_{At.}$.[7] What shall we conclude? It seems that we do not want to conclude from *that* that the connectives and quantifiers are explanatory notions. As Field (1999) has emphasized, it comes as no surprise that by adding some expressive resources to our language our schemas gain in proof-theoretical strength.

But « unexpected » non-conservativeness phenomena also arise in contexts where the specific interplay between the induction scheme and the increase in expressive means is not crucial. Such phenomena have been known for long and have been widely discussed by philosophers in the literature on logical constants in the tradition of proof-theoretic semantics[8] and elsewhere. A classical and well known example is the extension of intuitionistic logical rules in natural deduction by the rules for classical negation, which allow to prove Pierce's Law that $((P \to Q) \to P) \to P$. Another stimulating example is Dummett's introduction of ordinary rules for disjunction over a system already containing the ordinary rules for other connectives *and* rules for a non-standard, restricted, notion of disjunction.[9]

---

[7] See e.g. Hajek and Pudlak (1993), theorem 1.26 p.220 and 1.29 p. 221.

[8] See for instance Dummett (1991).

[9] See Dummett (1991), p.290. More specifically, the nonstandard disjunction here has the usual introduction rules of disjunction, whereas the elimination rule is like the standard elimination rule for disjunction except that it requires that there be no collateral assumptions in the minor premises.

Non-conservativity results in that case[10], whereas adding the very same ordinary rules for disjunction over the set of non-pathological rules only yields a conservative extension. Again, it seems clear that non-conservativity that obtains from adding the standard rules for disjunction over the pathological system, does not tell against the logicality of disjunction.[11] Outside the proof-theoretical semantics tradition, the logicists have of course held that higher-order arithmetic is logic, or that type-theoretical notions of any order are logical notions (Tarski 1986), despite repeated non-conservativity of the extensions as one climbs up the hierarchy of types. And whether or not one accepts logicism, the fact that full second-order Peano arithmetic is not conservative over first-order Peano arithmetic is generally hardly considered to adjudicate the issue about whether or not second-order logic is logic.

Do we have any reason to believe that the non-conservativity of the Tarskian extension of first-order PA is telling against the logicality of truth then ? And why should we regard this non-conservativity result as more significant than the conservativity of the Tarskian extension over Robinson's arithmetic Q, or over PA + ω-rule ? Why would non-conservativity over first-order PA be of any special significance for the notion of truth ? The point of the previous examples is to recall something quite familiar to those who have

---

[10] The rules for standard disjunction allow a detour to prove the - previously unprovable - distributivity of conjunction over the non-standard disjunction.

[11] A couple of remarks might be in order here. Dummett himself has taken similar arguments, based on a natural deduction formalisation of logic, to show that classical logic is not logic. Dummett's motivating idea was that a logical connective's introduction and elimination rules should be "self-justifying", which he took to imply conservativity. In response to Dummett, a number of philosophers have worked out new logico-philosophical framework allowing for full inferential justification of classical logic keeping with the conservativity requirement (for references see e.g. Restall 2005 and Rumfitt 2000). However, others have also questioned the implication from harmony to conservativity. Many philosophers, taking their inspiration from Gentzen, have held that the sought-after "self-justifying" property of "harmony" of the rules is a "local" property that hinges on the relationship between introduction and elimination rules only, or, in sequent formalism, between left and right introduction rules (see Schroeder-Heister's survey of proof-theoretic semantics (2014) for further variations). But such local notions of harmony do not generally imply conservativity (for instance in sequent formalism one also needs the so-called "subformula property"). Some philosophers have embraced the idea that the introduction of harmonious rules need not even imply the consistency of the resulting system (see e.g. Read 2000, 2010). The upshot of these remarks is that even in the proof-theoretic semantics tradition, the relationship between logicality and various conservativity and non-conservativity results is controversial. That much is sufficient for my purpose here.

studied logical constants: examples abound of conflicts between our judgement of logicality regarding an expression and the conservativity properties of the axioms governing the notion over a variety of base theories framed in various logical frameworks. Presumably, this is partly because the epistemic significance of non-conservativity is not univocal[12] - especially when the base theories under consideration are thought to be "deficient" or abnormal in some special way -, and partly because conceptions about logical notions are themselves to some extent a matter of controversy.

5. Deflationism, conservativity and logicality

In view of the various examples of the previous section, are we prepared to argue that non-conservativity shows that logical constants are *explanatory* rather than *expressive* devices ? Most of us, I guess, would resist the conclusion. But be that as it may, it remains that the conservativity test quoted above is in bad shape. Recall that the deflationist's paradigm for the truth predicate is that of logical expressions, or so I have suggested.[13] Then either the relevant target distinction to be captured by the conservativity test is the deflationist's one, but then the test seems to fail to achieve its purpose, or one maintains the test and declares truth to be explanatory on this ground, but then the argument fails in another way. For now what would really be at stake is whether logical constants generally are to be conceived as explanatory or not, and that would certainly require a principled discussion about the nature of logical notions and their function in discourse and thought – something which is conspicuously missing from the argument.

The conclusion to be drawn from these brief remarks is not, in my view, that the conservativity argument fails to refute deflationism. It is that the argument succeeds in

[12] A point that Gödel had already emphasized in different contexts, e.g. in his discussion of the extensions of set theory in Gödel 1947.

[13] Remark that this claim is different, and more cautious, than the claim that the T-sentences are logical truths. Assessing the latter claim would involve entirely different considerations pertaining to the occurrences of names for sentences in the T-sentences. See e.g. Cook 2014 for an argument and further references on the modal status of T-sentences. On the nature of the sentential names involved in the conceptually basic form of the T-sentences according to the deflationist - a key issue to adjudicate whether those T-sentences can be conceived are purely formal truth – I refer the reader to Field (2001), chap. 5.

putting pressure on the deflationist to clarify his distinction between explanatory and non - explanatory notions and, by the same token, if this is the path he is willing to take, between logical and non-logical notions. Much work has been done in the last fifty years on the distinction between logical and non-logical expressions, both in the proof-theoretic tradition (e.g. Prawitz (1969), Dummett (1991) and their followers) and in the semantic tradition (Tarski (1986), Sher (1991), Feferman (1999), and their followers). If we are not to stay content with vague claims regarding the « logical function » of the truth predicate or the « logical nature » of the concept of truth, and if we want to understand what this really means[14], a first step in the way of the proposed clarification would be to discuss how the deflationist's view that truth is a logical notion fits in the framework of these traditions. This is a work I postpone for another occasion.[15]

References

Cook, Roy T . 2012. « The T-schema is not a logical truth », *Analysis* 72 (2), pp. 231-239.

Dummett, Michael. 1991. *The logical basis of metaphysics.* Harvard University Press.

Feferman, Solomon. 1999. « Logic, Logics, Logicism ». *Notre Dame Journal of Formal Logic* 40 (1) :31-54

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

Field, Hartry. 2001. *Truth and the absence of facts.* Oxford University Press.

Gödel, Kurt. 1947. "What is Cantor's continuum problem ?". *The American Mathematical Monthly* 54 (9):515-525.

[14] It is to be noted that the usual remarks regarding the inter-definability of the truth predicate with substitutional quantifiers, or its « equivalence » with infinite conjunctions (over definable sets of sentences), are by no means by themselves sufficient arguments for the logical character of the notion of truth. The reason is that it is not clear at all that substitutional quantifiers or infinite conjunctions are themselves logical notions. More generally we expect from a principled criteria of logicality that it be grounded on an epistemological analysis of the nature of logical notions or logical truth and that it be able in principle to distinguish e.g. logical notions from mathematical ones.

[15] See [reference to a forthcoming work by the author]

Hajek, Petr, and Pavel Pudlak. 1993. *Metamathematics of first-Order Arithmetic*: Springer Verlag.

Halbach, Volker. 1999. "Conservative theories of classical truth." *Studia Logica* 62:353-370.

Horwich, Paul. 1999. *Truth*, 2nd edition, Oxford University Press.

Ketland, Jeffrey. 1999. "Deflationism and Tarski's Paradise." *Mind* 108:69-94.

McGee, Vann. 1996. "Logical operations". *Journal of Philosophical logic* 25:567-580.

Prawitz, Dag. 1965. *Natural deduction*. Almquist & Wicksell.

Read, Stephen. 2000. Harmony and Autonomy in Classical Logic. *Journal of Philosophical Logic* 29 (2):123-154

Read, Stephen. 2010. "General-Elimination Harmony and the Meaning of Logical Constants". *Journal of Philosophical Logic*, 39 (5):557-576

Restall, Greg. 2005. "Multiple conclusions" in P. Hajek, L. Valdés-Villanueva, and D. Westerstalh (eds.), *Logic, Methodology and Philosophy of Science : Proceedings of the Twelfth International Congress*, KCL Publications, 189-205.

Shapiro, Stewart. 1998. "Proof and Truth : Through Thick and Thin." *Journal of Philosophy* 95 (10):493-521.

Sher, Gila. 1991. *The bounds of logic.* MIT Press.

Schroeder-Heister, Peter. 2014. "Proof-Theoretic Semantics", *The Stanford Encyclopedia of Philosophy* (Summer 2014 Edition), Edward N. Zalta (ed.), forthcoming URL = <http://plato.stanford.edu/archives/sum2014/entries/proof-theoretic-semantics/>.

Tarski, Alfred. 1986. "What are logical notions ?". *History and Philosophy of Logic* 7:143-154.