THE REVIEW OF SYMBOLIC LOGIC
Volume 4, Number 2, June 2011

T-EQUIVALENCES FOR POSITIVE SENTENCES

CEZARY CIEŚLIŃSKI
Institute of Philosophy, The University of Warsaw

Abstract. Answering a question formulated by Halbach (2009), I show that a disquotational truth theory, which takes as axioms all positive substitutions of the sentential T-schema, together with all instances of induction in the language with the truth predicate, is conservative over its syntactical base.

§1. Introduction. Disquotational theories of truth can be based on the local or the uniform T-schema. Accordingly, there are two possible ways for the disquotationalist to proceed. The first option is to take as a starting point the schema:

(Tr-local) $Tr(\ulcorner\varphi\urcorner) \equiv \varphi$

and to declare as axioms all substitutions of (Tr-local) by sentences (possibly with the truth predicate) forming an appropriate recursive substitution class.

The second possibility is to use the (apparently) more comprehensive schema of uniform disquotation:

(Tr-uniform) $\forall x_1 \dots x_n[Tr(\ulcorner\varphi(x_1 \dots x_n)\urcorner) \equiv \varphi(x_1 \dots x_n)]$

As before, the set of truth-theoretical axioms will be then defined by choosing a recursive class of formulas to be substituted in (Tr-uniform).[1]

Apart from these truth-theoretical principles, disquotational truth theories will contain also axioms of their syntactical base theory. In what follows we will assume that Peano arithmetic plays this role. In the usual axiomatization of PA, all arithmetical substitutions of the induction schema qualify as axioms; henceforth we will assume that the disquotational theory extends PA also in the sense that it contains as axioms all instances of induction for the extended language, with the truth predicate.

The most basic variants of disquotational theories are obtained by taking just arithmetical sentences (or formulas) as substitution classes for (Tr-local) or (Tr-uniform). It is a well-known fact that these theories are very weak: both the uniform and the local version is a conservative extension of PA; they are also quite weak in proving truth-theoretical generalizations (see e.g., Halbach, 2001, p. 1960). This weakness motivates a search for other natural substitution classes, which would produce stronger theories, permitting us at the same time to avoid paradoxes.

Received: December 7, 2010

[1] The axioms of uniform disquotation have a clear interpretation in the arithmetical context. Expressions like "$\forall x Tr(\ulcorner\varphi(x)\urcorner)$" can be understood as "$\forall x\ Tr(sub(\ulcorner\varphi(x)\urcorner, name(x)))$," with "$sub$" standing for substitution, and "$name(x) = y$" being an arithmetical representation of a recursive function, which assigns to a number x its canonical name "$S \dots S(0)$" where the successor symbol "S" is repeated x times. In some other contexts axioms of uniform disquotation would not employ a notion of truth, but a binary satisfaction predicate "$Sat(x, y)$"; the axiom schema would then be: $\forall x_1 \dots x_n[Sat(\ulcorner\varphi(x_1 ... x_n)\urcorner, x_1 \dots x_n) \equiv \varphi(x_1 \dots x_n)]$.

doi:10.1017/S1755020311000116

The existence of such stronger consistent theories is beyond question. In fact every consistent recursively axiomatizable theory in the language with the truth predicate can be axiomatized just by choosing an appropriate substitution class for (Tr-local)—this follows from a result by McGee (1992), who showed that every sentence in the language of arithmetic extended with "Tr" is provably (in PA) equivalent with some substitution of (Tr-local); moreover, the method of finding for a given β a T-sentence provably equivalent to β is effective. However, from a philosophical point of view McGee's result doesn't give us much. We can't rest satisfied with discerning a particular substitution class for a given truth schema, if our only motivation for choosing this class amounts to the fact that it permits us to axiomatize a specific (nondisquotational) theory of our choice. The guiding intuition of the disquotationalist is rather that substitutions of T-schemata are epistemologically basic (perhaps obvious and unproblematic)—we do not accept them *because* we accept the axioms of some nondisquotational truth theory: if at all, the justification should proceed in the opposite direction. Of course in view of the paradoxes, the disquotationalist must restrict somehow his set of substitutions of the T-schemata. And I think we should grant him that much: avoiding the paradoxes should be treated as a permissible motivation for the disquotationalist, who tries to restrict the scope of the available substitution classes for the T-schemata.

Halbach (2009) formulates an interesting proposal which goes in this direction. After observing that paradoxical reasonings involve the application of truth to sentences containing a negative occurrence of the truth predicate (p. 788), he proposes the set of positive formulas (formulas in which every occurrence of "Tr" appears in the scope of an even number of negations) as a substitution class for (Tr-uniform). The resulting theory, denoted as PUTB (*positive uniform Tarski biconditionals*), is then showed to be quite strong indeed: by Halbach's theorem 5.1 (p. 792) it proves all arithmetical sentences derivable in Kripke–Feferman theory KF (in fact Halbach shows that KF and PUTB are arithmetically equivalent).[2] The truth predicate in PUTB is not compositional though—compositional axioms of KF are not theorems of PUTB (see Halbach, 2009, lemma 6.1 and below). Nevertheless, the disquotationalist has at his disposal an arithmetically strong truth theory, obtained from (Tr-uniform) by a legitimate choice of a suitable substitution class (the choice is motivated by the analysis of the way in which paradoxes are produced).

Halbach ends his paper with an open problem: he asks what happens if we drop (Tr-uniform) and build instead a theory taking as axioms all substitutions of (Tr-local) by positive sentences? In the present paper I give an answer to this question: (Tr-local) with a substitution class consisting of positive sentence only produces a theory which conservatively extends PA. The next sections contain a proof of this result.

§2. Notation and basics.

Throughout this paper following notation will be used:

- The symbols $\neg$, $\wedge$, $\vee$, $\exists$, $\forall$ for sentential connectives and quantifiers.
- L_{PA}, $Sent_{PA}$—arithmetical formulas and sentences.
- L_{Tr}, $Sent_{Tr}$—formulas and sentences of the language of arithmetic extended with "Tr."
- L^{+}_{Tr}, $Sent^{+}_{Tr}$—positive formulas and sentences (to be defined below).
- Ind_{φ}—induction for a formula φ.

[2] The classical paper Kripke (1975) describes Kripke's semantic construction; some years later Feferman (1991) presented an axiomatic theory, which came to be known as KF.

DEFINITION 2.1 *We define by simultaneous induction sets of formulas P_k and N_k ("Q" is a quantifier and "$\circ$" is either $\wedge$ or $\vee$):*

$$\begin{array}{llll} P_0 = L_{PA} \cup \{Tr(t) : t \in Tm\} & & N_0 = L_{PA} \cup \{\neg Tr(t) : t \in Tm\} & \\ P_{k+1} = P_k \cup \{\neg\alpha : \alpha \in N_k\} & & N_{k+1} = N_k \cup \{\neg\alpha : \alpha \in P_k\} & \\ \quad \cup \{\neg Qx\alpha : \alpha \in N_k\} & & \quad \cup\{\neg Qx\alpha : \alpha \in P_k\} & \\ \quad \cup \{\alpha \circ \beta : \alpha, \beta \in P_k\} & & \quad \cup\{\alpha \circ \beta : \alpha, \beta \in N_k\} & \\ \quad \cup \{Qx\alpha : \alpha \in P_k\} & & \quad \cup\{Qx\alpha : \alpha \in N_k\} & \end{array}$$

The set L^+_{Tr} of positive formulas is then specified as $\bigcup_{n \in N} P_n$.

Now we formulate the main theorem.

THEOREM 2.2 *Denote as PTB ("positive Tarski biconditionals") the theory: $PA \cup \{Tr(\ulcorner\varphi\urcorner) \equiv \varphi : \varphi \in Sent^+_{Tr}\} \cup \{Ind_\varphi : \varphi \in L_{Tr}\}$. Then PTB is conservative over PA.*

We are going to show that for an arbitrary finite $Z \subseteq PTB$ and for an arbitrary recursively saturated model M, M can be extended to a model of L_{Tr} in such a way as to make all sentences in Z true. Then if for some $\psi \in L_{PA}$ $PTB \vdash \psi$, ψ can be derived from some finite subset Z of PTB, and since every recursively saturated model of PA can be extended to a model of Z, we will have: ψ is true in every such model; therefore $PA \vdash \psi$.[3]

Below I introduce some basic definitions and facts to be used later.

DEFINITION 2.3 *We define a translation function $t(a, \varphi)$—for φ belonging to L_{Tr}, it gives as value an arithmetical formula with a parameter a. The function is defined by induction on the complexity of a formula belonging to the language with the truth predicate.*

- $t(a, \ulcorner t = s\urcorner) = \ulcorner t = s\urcorner$
- $t(a, Tr(t)) = \ulcorner t \in a\urcorner$
- $t(a, \neg\psi) = \neg t(a, \psi)$, *similarly for conjunction and disjunction*
- $t(a, \exists x\psi) = \exists x t(a, \psi)$, *similarly for a general quantifier.*

Expressions with "$\in$" should be understood as arithmetical formulas (possibly with parameters) used for the purposes of coding sets; for example, "$x \in a$" could be a formula "$p_x | a$," with p_x being the xth prime.[4]

FACT 2.4 *Let $d \in M$. Let $K = (M, T)$ with $T = \{a : M \models a \in d\}$. Then for every $\varphi \in L_{Tr}$, for every valuation v in M, we have:*

$$M \models t(d, \varphi)[v] \text{ iff } K \models \varphi[v]$$

Proof. The proof is by induction on the complexity of φ. If for example, $\varphi = Tr(t)$, then we have: $M \models t(d, Tr(t))[v]$ iff $M \models t \in d[v]$ iff $val^M(t, v) \in T$ iff $K \models Tr(t)[v]$. The proof of the other clauses is routine. □

FACT 2.5 *Let $M_1 = (M, A)$, $M_2 = (M, B)$ with A, B being subsets of M such that $A \subseteq B$. Then for every valuation v in M, for every $\varphi(x_1 \dots x_n) \in L^+_{Tr}$, we have: if $M_1 \models \varphi(x_1 \dots x_n)[v]$, then $M_2 \models \varphi(x_1 \dots x_n)[v]$.*

Proof. The proof consists in showing that every formula in L^+_{Tr} is logically equivalent with some *strictly positive* formula, that is, a formula in which no occurrence of

[3] On recursively saturated models and their properties, see Kaye (1991), especially pp. 148ff.

[4] On coded sets, see for example, Kaye (1991, p. 141).

"Tr" is negated. Then it is enough to check that every strictly positive formula satisfies Fact 2.5. □

§3. Proof of Theorem 2.2. We remind the reader, that a set $Z(x, a_1 \ldots a_n)$ of formulas with a free variable x and parameters $a_1 \ldots a_n$ from a model M is a *type* over M, if its every finite subset has a realization, that is, for every finite subset S of $Z(x, a_1 \ldots a_n)$ there is an $a \in M$ which makes all formulas in S true in M. Even though a type Z is thus finitely realized, it's quite possible in general that M doesn't contain a number a which makes all formulas in Z true (i.e., realizes Z as a whole). However, in recursively saturated models all recursive types are realized—and that is what will be used in the definition to follow.

DEFINITION 3.1 *Given a recursively saturated model M, we are going to define by induction a family of recursive types over M, a family of elements realizing these types and a family of models M_n which extend M to a model of L_{Tr}.*

1.
- $p_0(x) = \{\varphi \in x \equiv \varphi : \varphi \in Sent_{PA}\} \cup \{\forall w(w \in x \Rightarrow w \in Sent_{PA})\}$
- d_0 *realizes* $p_0(x)$
- $T_0 = \{a : M \models a \in d_0\}$
- $M_0 = (M, T_0)$

2.
- $p_{n+1}(x, d_n) = \{\varphi \in x \equiv t(d_n, \varphi) : \varphi \in Sent^+_{Tr}\} \cup \{\forall z(z \in d_n \Rightarrow z \in x)\} \cup \{\forall z(z \in x \Rightarrow z \in Sent^+_{Tr})\}$
- d_{n+1} *realizes* $p_{n+1}(x, d_n)$
- $T_{n+1} = \{a : M \models a \in d_{n+1}\}$
- $M_{n+1} = (M, T_{n+1})$

First, we are going to show that the above definitions are correct ones. The set $p_0(x)$ is obviously a type, so after choosing an element d_0, both T_0 and the model M_0 become uniquely determined.[5] Given d_n and M_n, we show that $p_{n+1}(x, d_n)$ is a type. Consider a finite subset Z of $p_{n+1}(x, d_n)$. Let "$\varphi_0 \in x \equiv t(d_n, \varphi_0) \ldots \varphi_i \in x \equiv t(d_n, \varphi_i)$" be all T-sentences in Z for which the formula on the right side of the equivalence symbol is true in the model, that is, for every $k \leq i M \models t(d_n, \varphi_k)$. Define a (nonstandard) number s in M as $d_n \cup \{\varphi_0 \ldots \varphi_i\}$. We claim that s realizes Z. From the construction of s, obviously $M \models \forall z(z \in d_n \Rightarrow z \in s)$ and also $M \models \forall z(z \in s \Rightarrow z \in Sent^+_{Tr})$. It remains to be shown that the condition "$d_n \subseteq s$" generates no conflict, that is, we must show that: $\forall \varphi \in d_n M \models \varphi \in s \equiv t(d_n, \varphi)$. This follows however from the fact that:

$$\forall \varphi \in d_n M \models t(d_n, \varphi)$$

For $n = 0$ this is obviously true ($t(d_0, \varphi)$ is just φ—a sentence true in M), so assume that $n = i + 1$. Fix $\varphi \in d_{i+1}$. Then $\varphi \in L^+_{Tr}$ and $M \models t(d_i, \varphi)$. By Fact 2.4, $M_i \models \varphi$ and by Fact 2.5 $M_{i+1} \models \varphi$. So again by Fact 2.4 $M \models t(d_{i+1}, \varphi)$; in other words: $M \models t(d_n, \varphi)$ as required.

Since for every n, d_n and M_n are well defined, we can formulate the following corollary to Fact 2.4.

[5] Both in the basic and the inductive step of the construction, the choice of an element realizing a given type is not unique. Since the number of choices to be made is infinite, a specific choice function f may be fixed with the assumption that on each level we specify our d_n as the value of f.

COROLLARY 3.2 $\forall\varphi \in Sent_{Tr}\forall n\, [M \models t(d_n, \varphi)$ *iff* $M_n \models \varphi]$.

Now we are ready to prove Theorem 2.2.

Proof of Theorem 2.2. Let Z be a finite subset of PTB. Given a recursively saturated model M, we will find an L_{Tr} extension of M which makes Z true. Let $A = \{Tr(\ulcorner\varphi_0\urcorner) \equiv \varphi_0 \ldots Tr(\ulcorner\varphi_k\urcorner) \equiv \varphi_k\}$ be a set of all T-sentences in Z. Fix n as the smallest natural number such that:

$$\forall i \leq k[M_n \models \varphi_i \vee \neg\exists l \in NM_l \models \varphi_i]$$

The existence of such a number follows from Fact 2.5 together with the observation that $T_0 \subseteq T_1 \subseteq T_2 \ldots$. Then we observe that $M_{n+1} \models Z$. Since T_{n+1} is parametrically definable in M, it is inductive. It remains to be checked that $\forall i \leq kM_{n+1} \models Tr(\ulcorner\varphi_i\urcorner) \equiv \varphi_i$. For $i \leq k$, we have:

$$M_n \models \varphi_i \vee \neg\exists l \in NM_l \models \varphi_i$$

Case 1: $M_n \models \varphi_i$. Then $M_{n+1} \models \varphi_i$ (because φ_i is positive); we have also: $M_{n+1} \models Tr(\varphi_i)$ (since $M_n \models \varphi_i$, we know by Corollary 3.2 that $M \models t(d_n, \varphi_i)$, so with φ_i being positive, $\varphi_i \in d_{n+1}$). Therefore $M_{n+1} \models Tr(\varphi_i) \equiv \varphi_i$.
Case 2: $\neg\exists l \in NM_l \models \varphi_i$. Then $M_{n+1} \not\models \varphi_i$, and also $M_{n+1} \not\models Tr(\varphi_i)$, because otherwise $M \models \varphi_i \in d_{n+1}$, so $M \models t(d_n, \varphi_i)$, therefore by Corollary 3.2 $M_n \models \varphi_i$, contrary to the assumption. Finally in this case again: $M_{n+1} \models Tr(\varphi_i) \equiv \varphi_i$.

In effect we showed that a recursively saturated model of PA can be always extended to a model of Z, which ends the proof. □

The above reasoning shows that recursively saturated models of PA can be always *locally* extended with respect to PTB—we can always find in them an interpretation of "Tr" which makes true an arbitrary finite subset of PTB. This is enough for establishing the conservativeness result. But can they be extended to models of the whole of PTB? The following theorem gives an answer to this question.

THEOREM 3.3 (expandability) *Let $L \subseteq L^+$ be finite languages, M a countable, recursively saturated model of arithmetic with language L, and T an L^+-theory consistent with $Th(M)$. If T has a recursive axiomatization, then M can be expanded to a model of T.*[6]

In the proof of Theorem 2.2 we showed in fact the consistency of PTB with $Th(M)$ for an arbitrary recursively saturated model M; in effect Theorem 3.3 permits us to obtain the following corollary:

COROLLARY 3.4 *Every countable recursively saturated model of arithmetic can be extended to a model of L_{Tr} in such a way that it satisfies:*

1. *All equivalences of the form "$Tr(\varphi) \equiv \varphi$" for $\varphi \in Sent^+_{Tr}$*
2. *Induction in the extended language (with "Tr").*

[6] See Smorynski (1981), cf. also Kossak & Schmerl (2006, pp. 14–15). Smorynski formulates this result as theorem 3.9, p. 278. In fact he proves a more general version—instead of "T has a recursive axiomatization" he uses a condition "some axiomatization of T is coded in M." Since every recursive set will be coded, the formulation used above is also correct.

At this moment I will make two additional comments.

Comment 1. All models M_n satisfy the condition "$Tr(\psi) \Rightarrow \psi$" for all $\psi \in L_{Tr}$, so the same proof establishes conservativeness of a theory containing not only true-positive biconditionals with induction, but also all instances (not just the positive ones) of the "Tr-out" schema.

Comment 2. A slightly modified construction gives a proof of a stronger result (in the formulation below $\vec{z}$ stands for a sequence of variables).

THEOREM 3.5 *Let* $T = PA \cup \{Tr(\varphi) \equiv \varphi : \varphi \in Sent^{+}_{Tr}\} \cup \{\forall\vec{z}[Tr(\varphi(\vec{z})) \Rightarrow \varphi(\vec{z})] : \varphi(\vec{z}) \in L_{Tr}\} \cup \{Ind_{\varphi} : \varphi \in L_{Tr}\}$. *Then* T *is conservative over PA.*

In the proof, the only real change is a different characterization of the set of types (cf. Definition 3.1). Fixing a model M and a nonstandard $a \in M$, we put:

- $p_0(x, a) = \{\forall\vec{z} < a[\varphi(\vec{z}) \in x \equiv \varphi(\vec{z})] : \varphi(\vec{z}) \in L_{PA}\} \cup \{\forall w[w \in x \Rightarrow \exists\varphi(\vec{z}) \in L_{PA} \exists\vec{s} < a\ w = \ulcorner\varphi(\vec{s})\urcorner\}$
- $p_0(x, d_n, a) = \{\forall\vec{z} < a[\varphi(\vec{z}) \in x \equiv t(d_n, \varphi(\vec{z}))] : \varphi(\vec{z}) \in L^{+}_{Tr}\} \cup \{\forall z[z \in d_n \Rightarrow z \in x\} \cup \{\forall w[w \in x \Rightarrow \exists\varphi(\vec{z}) \in L^{+}_{Tr} \exists\vec{s} < a\ w = \ulcorner\varphi(\vec{s})\urcorner\}$

with d_n and M_n defined exactly as before. The rest of the proof doesn't differ much from the previous one.

§4. Final remarks.

In contrast with the arithmetical case (with substitution classes for (Tr-local) and (Tr-uniform) being just arithmetical sentences and formulas), positive substitution classes generate two arithmetically different disquotational theories. From a philosophical point of view, Theorem 2.2 should be treated as a negative result. PUTB could be considered as a welcome tool by philosophers who are both disquotationalists and instrumentalists about truth. In other words: if you think that truth is mere disquotation, and at the same time you consider it a useful device for proving new arithmetical facts (without worrying too much about specific truth-theoretic content of your theory), then PUTB might be for you. It exemplifies nicely that there is no conflict between these two intuitions: principled disquotationalism (with a non ad hoc choice of the substitution class for (Tr-uniform)) can indeed be squared with proof-theoretic strength. But even so, the question about sentential disquotationalism still remains. Our negative result eliminates one possible candidate for the role of a non ad hoc substitution class for (Tr-local), endowing our theory with proof-theoretical strength. We are still left with the question whether some version of sentential disquotationalism qualifies as a tenable philosophical standpoint.

§5. Acknowledgements.

The work reported in this paper was supported by grant number N N101 170438 by the Polish Ministry of Science and Education (MNiSW). I would like to thank Volker Halbach and an anonymous referee of RSL for their helpful comments on the original version of the paper.

BIBLIOGRAPHY

Feferman, S. (1991). Reflecting on incompleteness. *Journal of Symbolic Logic*, **56**, 1–49.

Halbach, V. (2001). Disquotational truth and analyticity. *Journal of Symbolic Logic*, **66**, 1959–1973.

Halbach, V. (2009). Reducing compositional to disquotational truth. *Review of Symbolic Logic*, **2**, 786–798.

Kaye, R. (1991). *Models of Peano Arithmetic*. Oxford, UK: Clarendon Press, Oxford.

Kossak, R., & Schmerl, J. (2006). *The Structure of Models of Peano Arithmetic*. Oxford, UK: Clarendon Press, Oxford.

Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy*, **72**, 690–712.

McGee, V. (1992). Maximal consistent sets of instances of Tarski's schema (T). *Journal of Philosophical Logic*, **21**, 235–241.

Smorynski, C. (1981). Recursively Saturated Nonstandard Models of Arithmetic. *Journal of Symbolic Logic* **46**, 259–286.

INSTITUTE OF PHILOSOPHY
THE UNIVERSITY OF WARSAW
POLAND
E-mail: c.cieslinski@poczta.uw.edu.pl