Notre Dame Journal of Formal Logic
Advance publication◇

Core Tarski and Core McGee

Neil Tennant

Abstract We furnish a core-logical development of the Gödel numbering framework that allows metamathematicians to attain limitative results about arithmetical truth *without* incorporating a genuine truth predicate into the language in a way that would lead to semantic closure.

We show how Tarski's celebrated theorem on the arithmetical undefinability of arithmetical truth can be established using only core logic in both the object language and the metalanguage. We do so at a high level of abstraction, by augmenting the usual first-order language of arithmetic with a primitive predicate *Tr* and then showing how it cannot be a truth predicate for the augmented language.

McGee established an important result about consistent theories that are in the language of arithmetic augmented by such a "truth predicate" *Tr* and that use Gödel numbering to refer to expressions of the augmented language. Given the nature of his sought result, he was forced to use classical reasoning at the meta level. He did so, however, on the additional and tacit presupposition that the arithmetical theories in question (in the object language) would be closed under classical logic. That left open the dialectical possibility that a constructivist (or intuitionist) could claim not to be discomfited by the results, even if they were to "give a pass" on the unavoidably classical reasoning at the meta level. In this study we "constructivize" McGee's result, by presuming only core logic for the object language. This shows that the perplexity induced by McGee's result will confront the constructivist (or intuitionist) as well.

1 Preliminaries

1.1 Definitions and terminology

Definition 1 $\mathbb{N}$ is the standard model of arithmetic.
R is Raphael Robinson's only-infinitely-axiomatizable arithmetic.

Received June 22, 2023; accepted April 26, 2024
◇Final volume, issue, and page numbers to be assigned.
2020 Mathematics Subject Classification: 03F30, 03F40, 03F52, 03F99, 03B47
Keywords: Robinson arithmetic, schema T, liar, recursive functions, representability, core logic, classical core logic, consistency, fixed points
 10.1215/00294527-2024-0032

Q is Raphael Robinson's finitely axiomatizable arithmetic (which is slightly stronger than R).

L is the first-order language of arithmetic with extralogical primitives 0, s, $+$, $\times$.

$\underline{n}$ is the numeral for the number n (i.e., the term $ss\ldots s0$ with n occurrences of s).

Th($\mathbb{N}$*)* is the set of all sentences of L that are true in $\mathbb{N}$: $\{\phi \in L \mid \mathbb{N} \Vdash \phi\}$.

Tr is a monadic primitive predicate.

$\mathcal{L}$ is the first-order language that augments L with the monadic predicate *Tr*.

Γ, Δ, Ω are sets of $\mathcal{L}$-sentences.

Π, Σ, Ξ are proofs (in a logical system specified in the context).

ϕ, ψ, θ, χ are formulas or sentences in $\mathcal{L}$.

A, B, C are propositional placeholders.

A system g of Gödel numbering is presumed fixed for all expressions of the language $\mathcal{L}$. For any such expression E, $\ulcorner E \urcorner$ is the numeral for the Gödel number $g(E)$ of E.

A Tarskian biconditional is a sentence (in $\mathcal{L}$) of the form $Tr(\ulcorner\phi\urcorner) \leftrightarrow \phi$; and any sentence ϕ for which this is taken to hold we shall say is being regarded, or treated, as Tarskian.

Tarskian biconditionals are instances of the famous schema (T).

We abbreviate such instances $Tr(\ulcorner\phi\urcorner) \leftrightarrow \phi$ as $\tau(\phi)$.

The set of all Tarskian biconditionals is called $\mathcal{T}$. (This is a subset of $\mathcal{L}$.)

ε is an enumeration, presumed fixed, of all the Tarskian biconditionals in $\mathcal{L}$.

I is intuitionistic logic.

C is classical logic.

$\mathbb{C}$ is core logic.

$\mathbb{C}^+$ is classical core logic.

$\vdash_{\mathbf{I}}$ is deducibility in intuitionistic logic.

$\vdash$ is deducibility in core logic.

$\vdash_{\mathbb{C}^+}$ is deducibility in classical core logic.

1.2 The rules of core logic $\mathbb{C}$ and classical core logic $\mathbb{C}^+$ The reader familiar with the rules of the core systems and with basic results about those systems may advance to Section 2. For the reader not familiar with the core systems, their rules and basic results about them will be set out here.

The rules are stated rather more fastidiously than is usually the case in traditional presentations of natural deduction à la Gentzen and Prawitz.

In the graphic statements of natural deduction rules below, a box appended to the discharge stroke over an "assumption for the sake of argument" indicates that one must have made use of that assumption in deriving the subordinate conclusion, whereas a diamond indicates that "vacuous discharge" is permitted.

Note that vertically descending dots in the following graphic rules indicate that nontrivial proof work may be involved in descending from the assumptions indicated to the subordinate conclusion in question. The *absence* of such descending dots above a premise (which will be a major premise for an elimination) indicates that the premise in question stands proud, with no nontrivial proof work above it.

In the list of rules below, the introduction and elimination rules of natural deduction are stated on the left; the corresponding right and left logical rules of the sequent calculus are stated on the right.[1]

(¬I)
$$\dfrac{\begin{array}{c}\overline{\varphi}^{\,\square\,(i)}\\ \vdots\\ \bot\end{array}}{\neg\varphi}\,(i)$$

$(\neg_R)$
$$\dfrac{\varphi\,,\Delta:\bot}{\Delta:\neg\varphi}$$

(¬E)
$$\dfrac{\neg\varphi \quad \begin{array}{c}\vdots\\ \varphi\end{array}}{\bot}$$

$(\neg_L)$
$$\dfrac{\Delta:\varphi}{\neg\varphi,\Delta:\bot}$$

(∧I)
$$\dfrac{\begin{array}{c}\vdots\\ \varphi\end{array} \quad \begin{array}{c}\vdots\\ \psi\end{array}}{\varphi\wedge\psi}$$

$(\wedge_R)$
$$\dfrac{\Delta:\varphi \qquad \Gamma:\psi}{\Delta,\Gamma:\varphi\wedge\psi}$$

(∧E)
$$\dfrac{\varphi\wedge\psi \qquad \begin{array}{c}(i)\overline{\underbrace{\varphi\;,\;\psi}}^{\,\square\,}(i)\\ \vdots\\ \theta\end{array}}{\theta}\,(i)$$

where □ means at least one of φ, ψ must have been used as an assumption

$(\wedge_L)$
$$\dfrac{\Delta:\theta}{\varphi\wedge\psi,\Delta\setminus\{\varphi,\psi\}:\theta}$$

where $\Delta\cap\{\varphi,\psi\}\neq\emptyset$

(∨I)
$$\dfrac{\begin{array}{c}\vdots\\ \varphi\end{array}}{\varphi\vee\psi} \quad \dfrac{\begin{array}{c}\vdots\\ \psi\end{array}}{\varphi\vee\psi}$$

$(\vee_R)$
$$\dfrac{\Delta:\varphi}{\Delta:\varphi\vee\psi} \quad \dfrac{\Delta:\psi}{\Delta:\varphi\vee\psi}$$

(∨E)
$$\dfrac{\varphi\vee\psi \quad \begin{array}{c}\overline{\varphi}^{\,\square\,(i)}\\ \vdots\\ \theta/\bot\end{array} \quad \begin{array}{c}\overline{\psi}^{\,\square\,(i)}\\ \vdots\\ \theta/\bot\end{array}}{\theta/\bot}\,(i)$$

$(\vee_L)$
$$\dfrac{\varphi,\Delta:\theta/\bot \qquad \psi,\Gamma:\theta/\bot}{\varphi\vee\psi,\Delta,\Gamma:\theta/\bot}$$

(→I)(a)
$$\dfrac{\begin{array}{c}\overline{\varphi}^{\,\square\,(i)}\\ \vdots\\ \bot\end{array}}{\varphi\rightarrow\psi}\,(i)$$

(→I)(b)
$$\dfrac{\begin{array}{c}\overline{\varphi}^{\,\diamond\,(i)}\\ \vdots\\ \psi\end{array}}{\varphi\rightarrow\psi}\,(i)$$

$(\rightarrow_R)$(a)
$$\dfrac{\Delta,\varphi:\bot}{\Delta:\varphi\rightarrow\psi}$$

$(\rightarrow_R)$(b)
$$\dfrac{\Delta:\psi}{\Delta\setminus\{\varphi\}:\varphi\rightarrow\psi}$$

(→E)
$$\dfrac{\varphi\to\psi \qquad \begin{array}{c}\vdots\\ \varphi\end{array} \qquad \begin{array}{c}\overline{\ \square\ }^{(i)}\\ \psi\\ \vdots\\ \theta\end{array}}{\theta}\,(i)$$

(→$_L$)
$$\dfrac{\Delta : \varphi \qquad \psi, \Gamma : \theta}{\varphi\to\psi, \Delta, \Gamma : \theta}$$

(∃I)
$$\dfrac{\begin{array}{c}\vdots\\ \varphi^x_t\end{array}}{\exists x\varphi}$$

(∃$_R$)
$$\dfrac{\Delta : \varphi^x_t}{\Delta : \exists x\varphi}$$

(∃E)
$$\dfrac{\exists x\varphi\text{ⓐ} \qquad \begin{array}{c}\overline{\ \square\ }^{(i)}\\ \underbrace{\text{ⓐ}\ldots\varphi^x_a\ldots\text{ⓐ}}\\ \vdots\\ \psi\text{ⓐ}\end{array}}{\psi}\,(i)$$

(∃$_L$)
$$\dfrac{\varphi^x_a, \Delta : \psi}{\exists x\varphi, \Delta : \psi}\ \text{ⓐ}$$

(∀I)
$$\dfrac{\begin{array}{c}\text{ⓐ}\\ \vdots\\ \varphi\end{array}}{\forall x\varphi^a_x}$$

(∀$_R$)
$$\dfrac{\Delta : \varphi}{\Delta : \forall x\varphi^a_x}\ \text{ⓐ}$$

(∀E)
$$\dfrac{\forall x\varphi \qquad \begin{array}{c}{}^{(i)}\overline{\quad}\ \cdots\square\cdots\ \overline{\quad}^{(i)}\\ \underbrace{\varphi^x_{t_1}\ ,\ \ldots\ ,\ \varphi^x_{t_n}}\\ \vdots\\ \theta\end{array}}{\theta}\,(i)$$

(∀$_L$)
$$\dfrac{\varphi^x_{t_1}\ ,\ \ldots\ ,\ \varphi^x_{t_n}, \Delta : \theta}{\forall x\varphi, \Delta : \theta}$$

The foregoing rules form the system $\mathbb{C}$ of *core logic*. We obtain the system $\mathbb{C}^+$ of *classical core logic* by adding the following classical rule of dilemma (natural deduction form on the left; sequent calculus form on the right).

(*Dil*)
$$\dfrac{\begin{array}{c}\overline{\ \square\ }^{(i)}\\ \varphi\\ \vdots\\ \psi\end{array} \qquad \begin{array}{c}\overline{\ \square\ }^{(i)}\\ \neg\varphi\\ \vdots\\ \psi/\bot\end{array}}{\psi}\,(i)$$

$$\dfrac{\Delta, \varphi : \psi \qquad \Gamma, \neg\varphi : \psi/\bot}{\Delta, \Gamma : \psi}$$

1.3 Earlier results about core logic $\mathbb{C}$ and classical core logic $\mathbb{C}^+$ that will find application in this study The reader not completely familiar with the core systems might find the following results of some use in tracking our subsequent discussion. Proofs can be found in Tennant [16] and in the various papers cited therein.

Metatheorem 1 $\left.\begin{array}{r}\Delta \nvdash_{\boldsymbol{I}} \bot \\ \Delta \vdash_{\boldsymbol{I}} \varphi\end{array}\right\} \Rightarrow \Delta \vdash \varphi .$

Metatheorem 2 (Cut admissibility for [classical] core proof)
There is an effective method [,] *that transforms any two [classical] core proofs*

$$\begin{array}{cc} \Delta & \chi, \Gamma \\ \Pi & \Sigma \\ \chi & \theta \end{array} \qquad \textit{(where } \chi \notin \Gamma \textit{ and } \Gamma \textit{ may be empty)}$$

into a [classical] core proof $[\Pi, \Sigma]$ of θ or of $\bot$ *from (some subset of)* $\Delta \cup \Gamma$.

The square bracketing indicates that Metatheorem 2 holds when it is read uniformly with "core" and when it is read uniformly with "classical core."

The following corollaries of Metatheorem 2 hold both with $\vdash$ read as $\vdash_{\mathbb{C}}$ throughout, and with $\vdash$ read as $\vdash_{\mathbb{C}^+}$ throughout.

Corollary 1 (Admissibility of cut for absurdity, or "cut for ⊥")
If $\Delta \vdash \varphi$ *and* $\Gamma, \varphi \vdash \bot$*, then* $\Delta, \Gamma \vdash \bot$.

Corollary 2 (Admissibility of cut on consistent premises)
If $\Delta \nvdash \bot$ *and* $\Delta \vdash \Gamma$ *and* $\Gamma \vdash \psi$*, then* $\Delta \vdash \psi$.

2 Two-Level Descriptors

We use the descriptor

$$\left[\frac{\mathcal{S}_1}{\mathcal{S}_2}\right]$$

to characterize reasoning at the meta level using logical system $\mathcal{S}_1$, about either the object-level logical system $\mathcal{S}_2$ itself or theories at the object level that are closed under deducibility in $\mathcal{S}_2$.

Some well-known results can be described as follows. Gödel's proof of the completeness of classical first-order logic **C** was of kind

$$\left[\frac{\mathbf{C}}{\mathbf{C}}\right].$$

The core logician's reworking of Henkin's method of proof for this result would be of kind

$$\left[\frac{\mathbb{C}^+}{\mathbb{C}^+}\right],$$

and we would be inclined to call it Core Henkin. And the part of it that delivers the Gödel–Glivenko–Gentzen theorem would be of kind

$$\left[\frac{\mathbb{C}^+}{\mathbb{C}}\right],$$

Kripke's proof of the completeness of intuitionistic logic **I** with respect to his possible-worlds frame semantics was of kind

$$\left[\frac{\mathbf{C}}{\mathbf{I}}\right].$$

The core logician's reworking of this result would be of kind

$$\left[\frac{\mathbb{C}^+}{\mathbb{C}}\right]$$

and we would be inclined to call it Core Kripke.

Veldman [18] and De Swart [6] produced intuitionistic proofs at the meta level of the completeness of intuitionistic first-order logic. These are results of kind

$$\left[\frac{\mathbf{I}}{\mathbf{I}}\right].$$

The core logician's reworking of this result, if it were ever to come to pass, would be of kind

$$\left[\frac{\mathbb{C}}{\mathbb{C}}\right]$$

and we would be inclined to call it Core de Swart–Veldman.

Our own studies Tennant [14] and [15] were of kind

$$\left[\frac{\mathbb{C}}{\mathbb{C}^{+}}\right].$$

In our later study Tennant [17] we conducted an investigation of type

$$\left[\frac{\mathbb{C}}{\mathbf{C}}\right].$$

There we were concerned to establish, using only core logic $\mathbb{C}$ at the meta level, Göel's first incompleteness theorem for suitably strong and consistent arithmetical theories closed under **C** (hence the title "Core Gödel"). The incompleteness theorem is at its strongest, of course, when the theory-closure at the object level is by means of (what would appear to be) the strongest possible logic, namely, classical logic **C**. The investigation could equivalently be described as one of kind

$$\left[\frac{\mathbb{C}}{\mathbb{C}^{+}}\right],$$

since the theories at the object level were assumed to be consistent, and classical core logic $\mathbb{C}^{+}$ proves the same consequences as does classical logic **C** from any consistent set of sentences.

Restricting the logic for closure of one's theorizing from classical logic to core logic is neither idiosyncratic nor philosophically irrelevant. After all, intuitionistic logic (hence also, this author would add, core logic) suffices for the metalinguistic derivations of all instances of Tarski's T-schema from his original recursive definition of satisfaction and truth.[2] Hence, core logic suffices to establish the *material adequacy* of Tarski's theory of truth (for any first-order language). This is a result of kind

$$\left[\frac{\mathbb{C}}{\mathbb{C}}\right].$$

Perhaps it would merit the title "Core-ur-Tarski." It should be pointed out, though, that when Tarski proved bivalence—that for every sentence φ of the object language, either φ is true or $\neg\varphi$ is true—he *had* to resort to using strictly classical reasoning at the meta level.[3] This could charitably be described as a result of kind

$$\left[\frac{\mathbb{C}^{+}}{\mathbb{C}^{+}}\right].$$

One can see that labeling a result as Core NN implies that an important result of NN is being "core-ified" at either the meta level or the object level (perhaps both). At which level this innovation is effected will depend on the kind of result NN proved.

The reader will see that in this study Core Tarski and Core McGee are so-called for slightly different reasons. Tarski's result about the indefinability of arithmetical truth is core-ified at *both* the meta level *and* the object level. McGee's result about the extraordinary power of arithmetical instances of the T-schema can be core-ified *only* at the object level (without, though, detracting from the philosophical interest it commands—in this regard it would be comparable with Core Kripke). What unifies our two chosen instances here of Core NN into a single study is their use of the same metamathematical materials and results. They deal with sufficiently strong and consistent subtheories of arithmetic; they appeal to representability of recursive relations; they make use of fixed points; and they explore the behavior of an elusive predicate *Tr*—whose two letters hint at the same philosophically important concept—in the arithmetical setting.

3 Formal Deducibilities in Core Logic

We shall be appealing to the fact that in core logic, CUT for $\vdash$ is admissible when its concluding sequent is either of the form $\Theta \vdash \chi$ with Θ consistent, or of the form $\Theta \vdash \emptyset$ (i.e., $\Theta : \bot$).

Bear in mind that any sequent provable in intuitionistic logic has a subsequent provable in core logic. We can therefore help ourselves to intuitionistic proofs in the object language, secure in the knowledge that the core logician can obtain results just as strong (if not even stronger).

Lemma 1 *Suppose that the sequents* $\Gamma, A : \neg A$, *and* $\Gamma, \neg A : A$ *are core-provable. Then* Γ *is core-inconsistent.*

Proof Suppose that we have both

$$(1) \quad \Gamma, A \vdash \neg A \qquad \text{and} \qquad (2) \quad \Gamma, \neg A \vdash A.$$

We shall show that we would then be able construct a core proof of the inconsistency of Γ. Our demonstration of this will be metalogical, using core logic in the metalanguage, and it will be formulated by means of core-deducibility statements.

Supposition (1) and the fact that $\neg A, A \vdash \bot$ ensure, by CUT FOR $\bot$, that $\Gamma, A \vdash \bot$. If $\Gamma \vdash \bot$, then we are done. Otherwise, $\Gamma, A \vdash \bot$ is true by virtue of a proof that uses A as an undischarged assumption. In this case it follows by $\neg$I in the object language that

$$(3) \quad \Gamma \vdash \neg A.$$

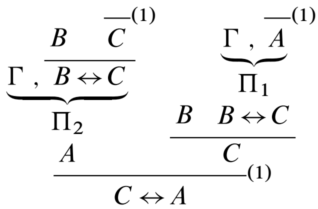
From supposition (2) along with (3) it follows by CUT ADMISSIBILITY FOR CORE PROOF that

$$\text{either} \qquad (i) \quad \Gamma \vdash A \qquad \text{or} \qquad (ii) \quad \Gamma \vdash \bot.$$

We proceed with proof by cases. In case (ii) we are done. It remains to consider only case (i). So suppose that $\Gamma \vdash A$. Then $\Gamma, \neg A \vdash \bot$. This, with (3), ensures by CUT FOR $\bot$ that $\Gamma \vdash \bot$. Now we are completely done. □

Lemma 2 *Suppose* $\Gamma, B \nvdash \bot$ *and that* $\Gamma, A \vdash B \leftrightarrow C$ *(by core proof* Π_1*, say); and* $\Gamma, B \leftrightarrow C \vdash A$ *(by core proof* Π_2*, say). Then* $\Gamma, B \vdash C \leftrightarrow A$.

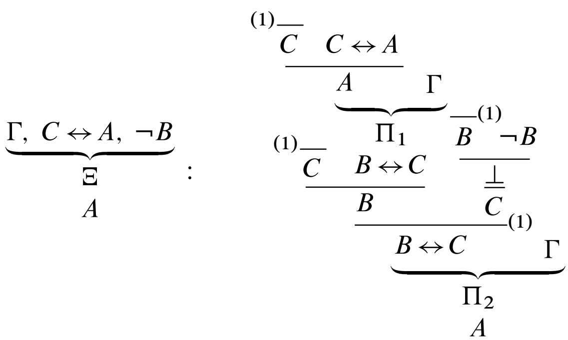
Proof We can construct the following intuitionistic proof.



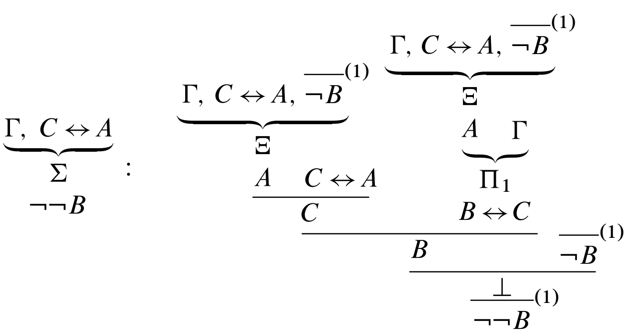
This proof shows that $\Gamma, B \vdash_{\mathbf{I}} C \leftrightarrow A$. Since by main supposition $\Gamma, B \nvdash \bot$, it follows by Metatheorem 1 that $\Gamma, B \vdash C \leftrightarrow A$. □

Lemma 3 *Suppose* $\Gamma, C \leftrightarrow A \nvdash \bot$ *and, as for Lemma 2, that* $\Gamma, A \vdash B \leftrightarrow C$ *(by core proof* Π_1*, say) and* $\Gamma, B \leftrightarrow C \vdash A$ *(by core proof* Π_2*, say). Then* $\Gamma, C \leftrightarrow A \vdash \neg\neg B$*.*

Proof First we construct the following intuitionistic proof Ξ.



Using Ξ, we can now form the following intuitionistic proof Σ.



Thus we have

$$\Gamma, C \leftrightarrow A \vdash_{\mathbf{I}} \neg\neg B.$$

Since by main supposition

$$\Gamma, C \leftrightarrow A \nvdash \bot,$$

it follows by Metatheorem 1 that

$$\Gamma, C \leftrightarrow A \vdash \neg\neg B.$$

□

Note that the classical logician could apply classical reductio at the final step labeled (1) in the last displayed proof, so as to infer the conclusion B. But the constructivist cannot do this. The constructivist has to rest content with $\neg\neg B$.

The main suppositions of consistency in Lemmas 2 and 3 will be sustained or fulfilled in all subsequent contexts in which those lemmas find application. This will be the case with the proofs of Lemmas 6, 7, 8, Corollary 5, and Lemma 9, all of which lead to Theorem 3 (constructivized McGee).

4 Discussion of Method

The metatheorem that Q affords representability of recursive functions is constructive at both the object level and the meta level. That is, the turnstiles of the interdeducibilities that are called for in the representation of recursive functions can all be read as deducibility in core logic $\mathbb{C}$; and the meta-level reasoning involved in establishing the representability theorem can also be carried out in $\mathbb{C}$. This was established in Tennant [17].

In their classic monograph [8], Tarski, Mostowski, and R. Robinson established the representability of all recursive functions in R (R. Robinson's well-known infinitely axiomatized proper subtheory of his finitely axiomatized theory Q). In doing so they worked with J. Robinson's definition of recursive functions that she gave in her paper Robinson [5]. For the version of representability with Q in place of R, but with Gödel's definition of general recursive functions, see Boolos and Jeffrey [1] and Tennant [9]. That all the extant formal explications of computable functions due to Turing, Church, Gödel, Kleene, and others are coextensive is well known, and it is what makes the celebrated Church–Turing thesis highly plausible.

Inspection of the informally rigorous proofs of all the results mentioned in the previous paragraph reveals them to be completely constructive, and therefore fully formalizable in intuitionistic (hence also in core) logic. We conjecture that the equivalence of J. Robinson's definition of general recursive functions with Gödel's definition of the same will likewise admit of completely constructive proof. Indeed, if this were not the case, then her (J. Robinson's) definition would be a singular outlier. We note the general emphatic claim of Odifreddi [4]:

> It should be noted that *the equivalence proofs among different notions of computability are effective* … . Effectiveness means that for any pair of notions there is a recursive function that, given the code of a recursive function relative to one notion, produces a code of the same recursive function relative to the other notion. (p. 101)

If our conjecture holds, then all of our subsequent discussion of matters modulo Q could be systematically strengthened so as to be matters modulo the slightly weaker theory R.

We shall proceed, however, with all our deducibility statements being modulo Q rather than modulo R, in order to ensure that we satisfy the classically captious. The reader nevertheless has the assurance that if and when our conjecture is rigorously established, one would be able to revert to expressing all the deducibilities modulo R rather than Q throughout our subsequent discussion. The "philosophical loss" of proceeding thus (in an abundance of caution) modulo Q rather than modulo R is not at all significant. Q is already a "weak enough" arithmetical theory to make the "core version" of the main results below arresting.

Theorem 1 (Representability) *For every recursive function $f(\vec{x})$ there is an L-formula $\varphi(\vec{x}, y)$ such that for all $\vec{n}$*

$$\varphi(\underline{\vec{n}}, a) \;\dashv_{\mathsf{Q}}\vdash\; a = \underline{f(\vec{n})}$$

with the parameter a not occurring in $\varphi(\vec{x}, y)$.

Definition 2 A binary formula $\delta(x, y)$ is a diagonal for Δ just in case for every unary formula $\psi(x)$ we have

$$\delta(\ulcorner\psi\urcorner, a) \ \dashv_{\Delta}\vdash \ a = \ulcorner\psi(\ulcorner\psi\urcorner)\urcorner,$$

with a parametrical.

Lemma 4 Q *has a diagonal.*

Proof Suppose # is a coding of unary formulas and ♮ is a coding of sentences. The mapping

$$n \mapsto \natural\big[\#^{-1}(n)(\underline{n})\big]$$

is effective. By Church's thesis it is recursive. Hence by the representability theorem there is a formula $\delta(x, y)$ such that for all n we have

$$\delta(\underline{n}, a) \ \dashv_{\mathsf{Q}}\vdash \ a = \underline{\natural\big[\#^{-1}(n)(\underline{n})\big]}.$$

If ψ is a unary formula, let $\ulcorner\psi\urcorner = \underline{\#\psi}$, and if θ is a sentence, let $\ulcorner\theta\urcorner = \underline{\natural\theta}$. Then for all unary formulas ψ we have

$$\delta(\underline{\#\psi}, a) \ \dashv_{\mathsf{Q}}\vdash \ a = \underline{\natural\big[\#^{-1}(\#\psi)(\underline{\#\psi})\big]},$$

that is,

$$\delta(\ulcorner\psi\urcorner, a) \ \dashv_{\mathsf{Q}}\vdash \ a = \ulcorner\psi(\ulcorner\psi\urcorner)\urcorner. \qquad \square$$

For any theory that affords numeral-wise representability of all recursive functions, the following important corollary holds. And the reasoning from representability to this corollary is, like the reasoning that establishes representability in the first place, thoroughly constructive, hence formalizable in core logic.

Corollary 3 (Fixed points) *For every unary formula* $\psi(x)$ *there is a sentence* θ *(called a fixed point for* ψ*) that is uniquely determined relative to one's chosen system of Gödel numbering, and is such that*

$$\theta \dashv_{\mathsf{Q}}\vdash \psi(\ulcorner\theta\urcorner),$$

that is,

$$\mathsf{Q}, \theta \vdash \psi(\ulcorner\theta\urcorner)$$

and

$$\mathsf{Q}, \psi(\ulcorner\theta\urcorner) \vdash \theta.$$

Proof This is a well-known corollary of numeral-wise representability of recursive functions in Q. We shall prove it as follows.

Let $\delta(x, y)$ be the diagonal for Q guaranteed by Lemma 4.

Let the unary formula $\chi(x)$ be

$$\forall y\big(\delta(x, y) \to \psi(y)\big).$$

Let θ be $\chi(\ulcorner\chi\urcorner)$, that is,

$$\forall y\big(\delta(\ulcorner\chi\urcorner, y) \to \psi(y)\big).$$

We now have to show that

$$\theta \dashv_{\mathsf{Q}}\vdash \psi(\ulcorner\theta\urcorner).$$

First we show that $\mathsf{Q}, \theta \vdash \psi(\ulcorner\theta\urcorner)$.

By Lemma 4 there is a parametric core proof

$$\frac{\underbrace{\mathsf{Q}, a = \ulcorner\chi(\ulcorner\chi\urcorner)\urcorner}_{\Pi(a)}}{}\quad \delta(\ulcorner\chi\urcorner, a)$$

Substituting $\ulcorner\chi(\ulcorner\chi\urcorner)\urcorner$ for a in $\Pi(a)$ we obtain the core proof

$$\underbrace{\mathsf{Q}, \overline{\ulcorner\chi(\ulcorner\chi\urcorner)\urcorner = \ulcorner\chi(\ulcorner\chi\urcorner)\urcorner}}_{\Pi(\ulcorner\chi(\ulcorner\chi\urcorner)\urcorner)} \quad \delta(\ulcorner\chi\urcorner, \ulcorner\chi(\ulcorner\chi\urcorner)\urcorner)$$

Now we can construct the core proof[4]

$$\dfrac{\underbrace{\mathsf{Q}, \overline{\ulcorner\chi(\ulcorner\chi\urcorner)\urcorner = \ulcorner\chi(\ulcorner\chi\urcorner)\urcorner}}_{\Pi(\ulcorner\chi(\ulcorner\chi\urcorner)\urcorner)} \atop \delta(\ulcorner\chi\urcorner, \ulcorner\chi(\ulcorner\chi\urcorner)\urcorner) \qquad \dfrac{\theta\text{, i.e.,} \atop \forall y(\delta(\ulcorner\chi\urcorner, y) \to \psi(y))}{\delta(\ulcorner\chi\urcorner, \ulcorner\chi(\ulcorner\chi\urcorner)\urcorner) \to \psi(\ulcorner\chi(\ulcorner\chi\urcorner)\urcorner)}}{\psi\big(\ulcorner\chi(\ulcorner\chi\urcorner)\urcorner\big), \atop \text{i.e., } \psi(\ulcorner\theta\urcorner)}$$

So we have shown that $\mathsf{Q}, \theta \vdash \psi(\ulcorner\theta\urcorner)$.
Second we show that $\mathsf{Q}, \psi(\ulcorner\theta\urcorner) \vdash \theta$.

By Lemma 4 once again, but in the converse direction, there is a parametric core proof

$$\underbrace{\mathsf{Q}, \delta(\ulcorner\chi\urcorner, a)}_{\Sigma(a)} \quad a = \ulcorner\chi(\ulcorner\chi\urcorner)\urcorner$$

Now we can construct the core proof

$$\dfrac{\dfrac{\dfrac{\underbrace{\mathsf{Q}\ \overline{\delta(\ulcorner\chi\urcorner, a)}^{(1)}}_{\Sigma(a)} \atop a = \ulcorner\chi(\ulcorner\chi\urcorner)\urcorner \qquad \psi(\ulcorner\theta\urcorner),\ \text{i.e.,} \atop \psi(\ulcorner\chi(\ulcorner\chi\urcorner)\urcorner)}{\psi(a)}}{\delta(\ulcorner\chi\urcorner, a) \to \psi(a)}{}^{(1)}}{\forall y\big(\delta(\ulcorner\chi\urcorner, y) \to \psi(y)\big), \atop \text{i.e., } \theta}$$

So we have shown that $\mathsf{Q}, \psi(\ulcorner\theta\urcorner) \vdash \theta$. □

Corollary 3 can be glossed as saying that the fixed point θ for a predicate ψ "says of itself" (modulo Q, and via the chosen system of Gödel numbering) that it has the property expressed by ψ. Note that the construction that produces θ ensures that ψ is a subformula within θ. Also, θ contains an occurrence of the augmenting predicate *Tr* if and only if ψ does.

Note that neither R nor Q contains any sentences involving *Tr*. $\mathcal{L}$-extensions of either of these two theories might contain sentences that involve the predicate *Tr*; but also might not contain any (except for sentences containing *Tr* that are logically true, such as, for example, $\forall x(Tr(x) \to Tr(x))$).[5] Examples of consistent $\mathcal{L}$-extensions of R, none of whose sentences contains *Tr* are: Robinson's arithmetic Q; Peano arithmetic PA; PA + $\mathrm{Con}_{\mathrm{PA}}$; PA + $\neg\mathrm{Con}_{\mathrm{PA}}$; and so on. We can even add Th($\mathbb{N}$) to this list.

We are now in a position, with enough preliminary details laid out, to embark on our "twin study" of philosophically important results of Tarski and of McGee about *Tr* that require, for their development, the foregoing remarks and results about representability of recursive functions in sufficiently strong and consistent theories of arithmetic, the existence of fixed point for unary formulas, and so on.

An important point to appreciate at the outset is that the predicate *Tr* that augments L is *not* being treated as a truth predicate. The question to be addressed in the two main parts of this study (Sections 5 and 6) is: *Could* it be?

5 Tarski

5.1 On possible interpretations of *Tr* To be sure, *Tr* will be featuring in biconditionals of the form $Tr(\ulcorner\phi\urcorner) \leftrightarrow \phi$, some (but not all) of which can be members of the (consistent) sets of $\mathcal{L}$-sentences, or $\mathcal{L}$-theories, with which we shall be concerned.

Recall that L is the unaugmented language of first-order arithmetic based on 0, s, $+$, and $\times$, and therefore does *not* contain any sentence involving the predicate *Tr*. The predicate *Tr* is, as it were, uninterpreted from the point of view of the theorist using only the language L. But one could provide an interpretation for *Tr* in the standard model $\mathbb{N}$ by assigning to *Tr* as its extension any set of natural numbers (including the empty set!). The question is only: Would whatever predicate-extension is chosen for *Tr* allow one to construe $Tr(\ulcorner\phi\urcorner)$ as expressing the *truth* of ϕ, on the understanding that $\ulcorner\phi\urcorner$ denotes ϕ?

Indeed, one very full such extension for *Tr* would be the set of all Gödel numbers of sentences true in $\mathbb{N}$. The standard model $\mathbb{N}$, with this extension supplied for the predicate *Tr*, would make true every Tarskian biconditional

$$Tr(\ulcorner\phi\urcorner) \leftrightarrow \phi$$

for sentences ϕ in L. And the $\mathcal{L}$-theory

$$\mathrm{Th}(\mathbb{N}) \cup \{Tr(\ulcorner\phi\urcorner) \leftrightarrow \phi \mid \phi \in L\}$$

would be consistent. No "paradox" would arise. *The same* holds true if one were to assign to *Tr* the set of all Gödel numbers of sentences *false* in $\mathbb{N}$. But in this devious (yet, still, consistent) case the predicate *Tr* would be better understood as expressing *falsity* rather than *truth*. (And the "Tarskian biconditionals" would in fact be true.)

The only context in which insuperable problems arise (for an interpretation of *Tr*) is where the sentences ϕ on the right-hand sides of Tarskian biconditionals *are allowed to contain occurrences of Tr itself.*

5.2 Back to fixed-point considerations From the standpoint of the $\mathcal{L}$-user, the Gödel-coding function g is firmly determined and settled upon.[6] It ensures (see Corollary 3) that *every* predicate (unary formula) $\psi(x)$ in $\mathcal{L}$ will have a fixed point θ, in the sense that

$$\theta \dashv_{\mathsf{Q}}\vdash \psi(\ulcorner\theta\urcorner).$$

We have already observed that θ contains an occurrence of *Tr* if and only if ψ does. The question then immediately arises: Can such θ, for whatever pertinent *Tr*-involving formula $\psi(x)$ happens to be chosen, consistently feature on the right-hand side of a Tarskian biconditional? That is, can one consistently assert, against the background of the coding-affording and consistent theory Q, the biconditional $Tr(\ulcorner\theta\urcorner) \leftrightarrow \theta$ — so that for the (presumed consistent) theory $\mathsf{S} = \mathsf{Q} \cup \{Tr(\ulcorner\theta\urcorner) \leftrightarrow \theta\}$, one has

$$\theta \dashv_{\mathsf{S}}\vdash Tr(\ulcorner\theta\urcorner)\ ?$$

The answer to this question is negative. It is impossible, given Q, for every sentence of $\mathcal{L}$ to be Tarskian. This is established by the following theorem.

Theorem 2 (Tarski) *It is not the case that for every $\mathcal{L}$-predicate $\psi(x)$, its fixed point θ can core-consistently be taken,* modulo Q*, as an instance of schema-T.*

Proof Let $\psi(x)$ be $\neg Tr(x)$. The fixed point θ for *this* choice of $\psi(x)$ will feature thus:

$$\theta \dashv_{\mathsf{Q}}\vdash \neg Tr(\ulcorner\theta\urcorner).$$

Now suppose for reductio that θ can core-consistently be taken, modulo Q, as an instance of schema-T. It would then follow that we would have, for the core-consistent theory $\mathsf{S} = \mathsf{Q} \cup \{Tr(\ulcorner\theta\urcorner) \leftrightarrow \theta\}$, that

$$\theta \dashv_{\mathsf{S}}\vdash Tr(\ulcorner\theta\urcorner),$$

and consequently also

$$Tr(\ulcorner\theta\urcorner) \dashv_{\mathsf{S}}\vdash \neg Tr(\ulcorner\theta\urcorner).$$

But now recall Lemma 1: if the sequents $\Gamma, A : \neg A$, and $\Gamma, \neg A : A$ are core-provable, then Γ is core-inconsistent. Take S for Γ, and $Tr(\ulcorner\theta\urcorner)$ for A. The reasoning in Lemma 1 showed that S would then be core-inconsistent, contradicting our assumption for reductio. Theorem 2 follows. □

In summary: by Corollary 3 we *do* have

$$\theta \dashv_{\mathsf{Q}}\vdash \neg Tr(\ulcorner\theta\urcorner);$$

so by Theorem 2 we *cannot* have

$$\theta \dashv_{\mathsf{S}}\vdash Tr(\ulcorner\theta\urcorner).$$

Tarski's schema-T cannot have as instances all fixed-point sentences for *Tr*-involving predicates. This is the enduring legacy of the liar in the era of Gödel numbering. Note, however, that the sentence θ here has to contain an occurrence of *Tr*.

Corollary 4 (Tarski) *Arithmetical truth is not definable by any arithmetical predicate.*

Proof Any such arithmetical predicate $\chi(x)$ would have to satisfy all instances of schema-T; and its negation would admit of a fixed point. Now apply Theorem 2 with $\chi(x)$ in place of $Tr(x)$. □

Note that this is not a logico-semantic *paradox*. It is a straightforward impossibility result: it is impossible to *define Tr(x)* by means of a unary formula $\chi(x)$ in the *Tr*-augmented language $\mathcal{L}$ of arithmetic so that it "behaves like a truth predicate" in accordance with Tarski's famous material adequacy condition on a theory of truth—that is, so that for every sentence θ of that language we have

$$\theta \dashv_{\mathsf{Q}}\vdash Tr(\ulcorner\theta\urcorner).$$

5.3 Arithmetical impossibility vs. logico-semantic paradox Theorem 2 and Corollary 4 reveal an impossibility—no pun intended—at the very core of our conceptual scheme. This is because the inconsistency established is *core* inconsistency, not simply (as in the usual treatments) *classical* inconsistency. Every core-inconsistent set is obviously classically inconsistent. But the converse does not, *at first order*, hold.[7] At first order there are classically inconsistent sets that are not core-inconsistent. An example is the singleton $\{\neg\forall x(Fx \vee \neg Fx)\}$. Since it is "easier" for a set of first-order sentences to be core-consistent than it is for it to be classically inconsistent, Theorem 2 arguably has a philosophical edge over the version that Tarski actually gave us, in which all the reasoning at the object level is taken, by default, to be classical reasoning.

That the difficulty with the concept of truth "lies deep" in the way brought out by core-logical analysis of its indefinability in arithmetic is of a piece (in a way meriting further investigation) with the fact (so this author contends) that *one needs only core-logical reasoning* in order to reveal the paradoxicality of any set of sentences whose inconsistency presents, intuitively, as a logico-semantic paradox. By the author's proof-theoretic criterion of paradoxicality, the disproofs involved with genuine paradoxes of this kind cannot be brought into normal form.[8] Their reduction sequences—in pursuit of normal form—do not terminate after finitely many steps. This is engendered by explicitly treating the truth predicate as a *logical* predicate, subject to introduction and elimination rules that appear to be in harmony. The nonnormalizability of the resulting disproofs—all of which are furnished in *core* logic—that are associated with logico-semantic paradoxes engendered in this way is the symptom, in this context (where the truth predicate is taken to be a *logical* predicate) of the same deep problem whose manifestation, in the context of attempting to frame an arithmetical *definition* of truth-in-arithmetic, is the straightforward, nonparadoxical, core-inconsistency of any attempt to do so.

6 McGee

McGee [3] proved an interesting and deep result (his Theorem 1 on p. 237) about the existence of maximally consistent sets of instances of schema (T), and how they reveal the extraordinary power of having a "truth predicate" *Tr* as part of the object language—which McGee took to be the first-order language of arithmetic, augmented by the primitive predicate *Tr* (remember we are calling this language $\mathcal{L}$). The "having" of a "truth predicate" was a matter of having *Tr* feature in Tarskian biconditionals—not all of them (on pain of inconsistency, as shown in Section 5), but rather in *as many as possible of them*, short of such inconsistency. Details will emerge presently.

McGee proved his metatheorem using classical logic **C**. It concerned consistent theories at the object level that contained R and were themselves, also, closed under classical logic. So his investigation was of type

$$\left[\frac{\mathbf{C}}{\mathbf{C}}\right], \text{ or, equivalently, } \left[\frac{\mathbb{C}^{+}}{\mathbf{C}}\right].$$

We say "equivalently" for the same reason as before: classical core logic $\mathbb{C}^{+}$ proves the same consequences as **C** does from any consistent set of sentences.[9] In this part of our study, by contrast, the investigation is of type

$$\left[\frac{\mathbf{C}}{\mathbb{C}}\right], \text{ or, equivalently, } \left[\frac{\mathbb{C}^{+}}{\mathbb{C}}\right].$$

We are allowing ourselves full classical logic **C** (equivalently, for the reason already expounded, classical core logic $\mathbb{C}^{+}$) to prove an analog, for the constructivist, of McGee's Theorem 1 (to be stated below). As already remarked, McGee's (classical) metatheorem concerned classically closed theories at the object level. We, by contrast, want to show that the metatheorem goes through virtually to the same effect for theories at the object level that are closed only under core logic $\mathbb{C}$. Thus the discombobulation (or philosophical puzzlement) effected by McGee's original result can afflict the core logician just as acutely (at least, to "within an epsilon"—which, in this context, is a double-negation).

In McGee's statement of his original theorem 1 (and accordingly in its subsequent proof) there is consideration of consistent $\mathcal{L}$-extensions S of Robinson arithmetic R. As is well known, R is an infinitely axiomatized arithmetical theory (in the language of arithmetic *unsupplemented*, of course, by the predicate *Tr*), which suffices for the representation of all recursive functions. Here is an exact statement of the original theorem:

> Let Δ be an S-consistent set of sentences of $\mathcal{L}$. Then there is a set Γ of instances of (T) such that (1) all the members of Δ are S-entailed by Γ, (2) Γ is S-consistent, (3) any set of instances of (T) which properly includes Γ is S-inconsistent, and (4) $\Gamma \cup$ R is a complete first–order theory [in $\mathcal{L}$—NT]. (p. 237 infra)

McGee provoked his reader to consider

> ... what our response to the liar paradox would look like if it were developed under this constraint alone[:] ... that it not restrict [Tarski's Schema] (T) more severely than necessary (p. 236)

and he developed the question further:

> [T]his is such an important constraint that it is worthwhile to study its effects by asking what our response to the liar paradox would look like if it were developed *under this constraint alone*, without any other considerations. (p. 236; emphasis added)

McGee then went on to state the philosophically surprising conclusion (for the adherent of *classical* logic, at least, who claims to be committed to a version of minimalism) that

> ... the mere desire to preserve as many instances of (T) as possible will give us too little to go on in constructing a consistent alternative to the naive theory of truth. (p. 237)

Our overarching concern in this section of our study is to determine whether the thinker who insists on dealing with consistent object-language theories closed only under *intuitionistic* logic (equivalently: under *core* logic) can be similarly discomfited. And we shall determine that this is indeed the case.

McGee's proof of his theorem contains the following brief passage (p. 238), consideration of which, at this stage, does not require an explanation of the sentence B_ϕ (though explanation will come in due course).

> $$\ldots\ B_\phi \leftrightarrow (\phi \leftrightarrow Tr(\ulcorner B_\phi \urcorner))$$
>
> is a theorem of R. It follows by truth-functional logic that
>
> $$\phi \leftrightarrow (Tr(\ulcorner B_\phi \urcorner) \leftrightarrow B_\phi)$$

is a theorem of R.

This appeal to "truth-functional logic" is of course an appeal to *strictly classical* logic—which McGee was assuming as the default logic for the object language.[10] $A \leftrightarrow (B \leftrightarrow C)$ logically implies $B \leftrightarrow (C \leftrightarrow A)$ in *classical* logic, but not in intuitionistic logic.

McGee was not concerned with the possibility of constructivizing his result, in the sense of assuming that the *object* language is governed by a constructive logic (even if, perforce, one needs a degree of classicism in establishing his result in the *meta*language). Without any interest in or concern about this possibility, no special attention was paid to either the left-to-right direction of the second biconditional in the quote above:

$$\phi \rightarrow (Tr(\ulcorner B_\phi \urcorner) \leftrightarrow B_\phi)$$

or its right-to-left direction:

$$(Tr(\ulcorner B_\phi \urcorner) \leftrightarrow B_\phi) \rightarrow \phi.$$

McGee appeared to leave open the dialectical possibility that a constructivist (or intuitionist) could claim not to be discomfited by his result, *if* it could somehow be shown that its proof went through only because of the "strict classicism" that had been presumed for the logic of the object language.

We proceed now to close off that dialectical possibility. We investigate how best to constructivize McGee's result, so that semantic closure (in the sense of allowing *nestings* of *Tr*) is as powerfully discomfiting for the constructivist at the object level as it is for the classicist. In the course of so doing we shall pay particular attention to the difference between the two directions of the biconditional.

The outcome of this investigation is that one has further confirmation, from an interestingly different angle, of the present author's contention (for which, see Tennant [16], and our earlier discussion at the end of Section 5.3) that the deductive reasoning at the object level that is involved in establishing the paradoxicality of any (sets of) sentences in a semantically closed object language *in which Tr is to play the logical role of a truth predicate* can always be carried out in core logic. The *core* reasoning in this case reveals the same baffling surprises, to a closely similar extent, that McGee's *classical* reasoning uncovered for any attempt to maximize the bounty of Tarskian biconditionals that one might aspire to garner in one's hunt to approximate, as best one can, the material adequacy of one's "theory of *Tr*" in the arithmetical setting.

6.1 The expansion method

Lemma 5 (Expansions) *Let Φ be a compact property of sets of sentences—that is, Φ holds of a set Ω if and only if Φ holds of every finite subset of Ω. Let γ be a countable enumeration of sentences, of order type ω. Suppose that $\Phi(\Omega)$. Define*

$$\begin{aligned}
\Omega_0 &=_{df} \Omega \\
\Omega_{n+1} &=_{df} \begin{cases} \Omega_n \cup \{\gamma_n\} \textit{ if } \Phi(\Omega_n \cup \{\gamma_n\}) \\ \Omega_n \textit{ otherwise} \end{cases} \\
\Omega^\gamma &=_{\text{df}} \bigcup_n \Omega_n .
\end{aligned}$$

Then $\Phi(\Omega^\gamma)$.

Proof At the outset we point out the classicism involved in the foregoing inductive definition of Ω^{γ}. The definition's inductive step requires one, at stage $n+1$, to determine whether the property Φ is enjoyed by the set $\Omega_n \cup \{\gamma_n\}$, where Ω_n has been "produced" by stage n. If Φ is not an effectively decidable property, then the definition is not constructive. And in this setting, where Φ involves determination of *consistency at first order*, it is definitely *not* effectively decidable. The classicism at this point is unavoidable.

We proceed, then, with a classical proof of the sought result. First we show by mathematical induction that $\forall m \Phi(\Omega_m)$.

By main supposition, we have $\Phi(\Omega_0)$. This accomplishes the basis for the induction. From the inductive step in the definition it is clear that if $\Phi(\Omega_n)$, then $\Phi(\Omega_{n+1})$. Hence by mathematical induction we have

$$\forall m \Phi(\Omega_m). \qquad (*)$$

Now suppose for reductio that

$$\neg\Phi(\Omega^{\gamma}).$$

Since Φ is compact, it follows by *classical* logic again (at the meta level) that there is some finite subset—Ω', say—of Ω^{γ} such that $\neg\Phi(\Omega')$. By definition of Ω^{γ} as $\bigcup_n \Omega_n$, it follows that some stage Ω_k, say, includes Ω'. So we have

$$\neg\Phi(\Omega')\ ;\ \Omega' \text{ finite};\ \Omega' \subseteq \Omega_k.$$

Since Φ is compact, we now have

$$\neg\Phi(\Omega_k),$$

contrary to (*). By classical reductio, discharging the *reductio* assumption $\neg\Phi(\Omega^{\gamma})$, it follows that

$$\Phi(\Omega^{\gamma}).$$ [11] □

Usually the compact property Φ is consistency of some kind.[12] In this study our choice of the property Φ of sets of $\mathcal{L}$-sentences will be consistency with Q. (See the comments in Section 4.) Note also that the enumeration γ can be of sentences in some restricted class that includes Ω. For example, Ω could be a set of Tarskian biconditionals, and γ could be an enumeration of Tarskian biconditionals. The important point is that relative to a particular enumeration γ presumed given, Ω^{γ} is uniquely determined.

In due course we shall use an enumeration of Tarskian biconditionals in an application of the foregoing expansion method. The choice of such an enumeration is of course arbitrary. There are uncountably many of them. This should be borne in mind as one further source (in our core logical reconstruction of McGee's result), over and above the contribution of the initial set Ω itself, of the multiplicity of maximal consistent sets of instances of the T-schema that McGee emphasizes in his paper.[13]

6.2 Constructivizing McGee's result The constructivization of McGee's results—in the sense of treating the closure of theories in the object language as effected constructively—can definitely go through with Q taking the place occupied by R in McGee's seminal study.[14] Note that since R is a subtheory of Q we are entitled to reprise all of McGee's deducibilites modulo R as deducibilites modulo Q.

We turn now to our investigation of what the core logician *can* accomplish within the setting that McGee so resourcefully created, which he did by means of a fixed-point construction using the cleverly chosen predicate $\phi \leftrightarrow Tr(x)$.

Definition 3 For any sentence ϕ, let B_ϕ be the “fixed point” (modulo Q) for the unary formula $\phi \leftrightarrow Tr(x)$.

Note that B_ϕ is a notation that we have taken over from McGee, to make comparison with his paper easier. B_ϕ is not to be confused with our propositional placeholder B.

The fixed point B_ϕ “says of itself” (modulo the theory Q) that it is true if and only if ϕ. For consider again what Corollary 3 provides:

$$\theta \dashv_{\mathsf{Q}}\vdash \psi(\ulcorner\theta\urcorner).$$

By Corollary 3, if we take $\phi \leftrightarrow Tr(x)$ for the predicate $\psi(x)$ and take B_ϕ for the fixed-point sentence θ, we have

$$B_\phi \;\dashv_{\mathsf{Q}}\vdash\; \phi \leftrightarrow Tr(\ulcorner B_\phi\urcorner).$$

In McGee’s hands, in the classical setting (in the object language) this entails

$$\phi \;\dashv_{\mathsf{Q}}\vdash\; B_\phi \leftrightarrow Tr(\ulcorner B_\phi\urcorner).$$

We shall in due course, however, see that for the core logician, by contrast, it yields (from Lemmas 6 and 7) “only”

$$\neg\neg\phi \;\dashv_{\mathsf{Q}}\vdash\; B_\phi \leftrightarrow Tr(\ulcorner B_\phi\urcorner)$$

—at least, in the right-to-left direction.

Definition 4 For any sentence ϕ, let $\beta(\phi)$ be the Tarskian biconditional $\tau(B_\phi)$, that is,

$$Tr(\ulcorner B_\phi\urcorner) \leftrightarrow B_\phi.$$

Tarskian biconditionals of this form will be called **fixed-point** Tarskian biconditionals.

Note that fixed-point Tarskian biconditionals are quite a rare breed among Tarskian biconditionals generally. Tarskian biconditionals such as

$$\begin{array}{l} Tr(\ulcorner 0=s0\urcorner) \leftrightarrow 0=s0 \\ Tr(\ulcorner \neg 0=s0\urcorner) \leftrightarrow \neg 0=s0 \\ Tr(\ulcorner \mathrm{Con}_{\mathrm{PA}}\urcorner) \leftrightarrow \mathrm{Con}_{\mathrm{PA}} \\ Tr(\ulcorner \neg\mathrm{Con}_{\mathrm{PA}}\urcorner) \leftrightarrow \neg\mathrm{Con}_{\mathrm{PA}} \\ \quad\vdots \end{array}$$

are definitely not fixed-point Tarskian biconditionals. But the Tarskian biconditionals

$$\begin{array}{l} Tr(\ulcorner B_{0=s0}\urcorner) \leftrightarrow B_{0=s0} \\ Tr(\ulcorner B_{\neg 0=s0}\urcorner) \leftrightarrow B_{\neg 0=s0} \\ Tr(\ulcorner B_{\mathrm{Con}_{\mathrm{PA}}}\urcorner) \leftrightarrow B_{\mathrm{Con}_{\mathrm{PA}}} \\ Tr(\ulcorner B_{\neg\mathrm{Con}_{\mathrm{PA}}}\urcorner) \leftrightarrow B_{\neg\mathrm{Con}_{\mathrm{PA}}} \\ \quad\vdots \end{array}$$

definitely are.

Definition 5 For any set Δ of sentences, let $\beta\Delta$ be the set $\{\beta(\phi) \mid \phi \in \Delta\}$.

Recall Definition 3, which defines B_ϕ as the fixed point for the unary formula $\phi \leftrightarrow Tr(x)$, so that we have, for any ϕ in $\mathcal{L}$,

$$B_\phi \;\dashv_{\mathrm{Q}}\vdash\; \phi \leftrightarrow Tr(\ulcorner B_\phi \urcorner).$$

Definition 6 By the *special substitution* we shall mean the following substitution for the placeholders Γ, A, B, and C in Lemmas 2 and 3:

$$\begin{array}{cccc} \Gamma & A & B & C \\ \downarrow & \downarrow & \downarrow & \downarrow \\ \mathrm{Q} & B_\phi & \phi & Tr(\ulcorner B_\phi \urcorner) \end{array} .$$

The substituend ϕ that is involved in the special substitution is itself a placeholder for $\mathcal{L}$-sentences. We shall see in due course that upon applying the special substitution to Lemmas 2 and 3, their consistency suppositions will be satisfied.

Lemma 6 *Suppose* $\mathrm{Q}, \phi \nvdash \bot$. *Then* $\mathrm{Q}, \phi \vdash \beta(\phi)$*—that is,*

$$\mathrm{Q}, \phi \;\vdash\; Tr(\ulcorner B_\phi \urcorner) \leftrightarrow B_\phi .$$

Proof By Corollary 3 (fixed points) we have

$$\mathrm{Q}, B_\phi \vdash \phi \leftrightarrow Tr(\ulcorner B_\phi \urcorner)$$

and

$$\mathrm{Q}, \phi \leftrightarrow Tr(\ulcorner B_\phi \urcorner) \vdash B_\phi .$$

The result follows by Lemma 2, using the special substitution. □

So: modulo Q, any sentence ϕ consistent with Q core-implies the Tarskian biconditional for the fixed-point sentence that "says of itself" that it is true if and only if ϕ.

Lemma 7 *Suppose* $\mathrm{Q}, \beta(\phi) \nvdash \bot$. *Then* $\mathrm{Q}, \beta(\phi) \vdash \neg\neg\phi$ *; that is,*

$$\mathrm{Q}, Tr(\ulcorner B_\phi \urcorner) \leftrightarrow B_\phi \vdash \neg\neg\phi \;.$$

Proof Once again, by Corollary 3 (fixed points) we have

$$\mathrm{Q}, B_\phi \vdash \phi \leftrightarrow Tr(\ulcorner B_\phi \urcorner)$$

and

$$\mathrm{Q}, \phi \leftrightarrow Tr(\ulcorner B_\phi \urcorner) \vdash B_\phi .$$

The result follows by Lemma 3, once again using the special substitution. □

So: modulo Q, the Tarskian biconditional for the fixed-point sentence (that "says of itself" that it is true if and only if ϕ is the case) core-implies the double negation of ϕ—provided only that that Tarskian biconditional is consistent with Q.

Lemma 8 *Suppose* $\mathrm{Q}, \Delta \nvdash \bot$. *Then* $\mathrm{Q}, \beta\Delta \nvdash \bot$.

Proof For (constructive) reductio ad absurdum, suppose that $\mathrm{Q}, \beta\Delta \vdash \bot$. Thus for some finite $\Delta' \subseteq \Delta$ we have $\mathrm{Q}, \beta\Delta' \vdash \bot$. Take any ϕ in Δ. Clearly $\mathrm{Q}, \phi \nvdash \bot$. By Lemma 6 we therefore have $\mathrm{Q}, \phi \vdash \beta(\phi)$. Hence by (multiple, but only finitely many applications of) CUT for $\bot$, it follows that $\mathrm{Q}, \Delta' \vdash \bot$, whence also $\mathrm{Q}, \Delta \vdash \bot$, contrary to main supposition. Thus $\mathrm{Q}, \beta\Delta \nvdash \bot$. □

Corollary 5 *Suppose* $\mathrm{Q}, \Delta \nvdash \bot$. *Then* $\mathrm{Q}, (\beta\Delta)^\varepsilon \nvdash \bot$.

Proof Immediate by Lemmas 5 and 8. □

Lemma 9 *Suppose* $\mathsf{Q}, \Delta \nvdash \bot$. *Suppose* γ *is an enumeration of all Tarskian biconditionals. Then for every sentence* ψ *we have either* $\mathsf{Q}, (\beta\Delta)^\gamma \vdash \neg\neg\psi$ *or* $\mathsf{Q}, (\beta\Delta)^\gamma \vdash \neg\psi$.

Proof Suppose γ is an enumeration of all Tarskian biconditionals. Let ψ be an arbitrary sentence. Now make the main supposition:

$$\mathsf{Q}, \Delta \nvdash \bot.$$

By Lemma 8 we have

$$\mathsf{Q}, \beta\Delta \nvdash \bot.$$

In other words, we have $\Phi(\beta\Delta)$, where Φ is the compact property of Q-consistency. Hence by Lemma 5 we have

$$\mathsf{Q}, (\beta\Delta)^\gamma \nvdash \bot. \qquad (\dagger)$$

Now suppose for the sake of argument that

$$\mathsf{Q}, (\beta\Delta)^\gamma, \psi \vdash \bot.$$

By (†) it is clear that this deducibility statement would have to be witnessed by a (dis)proof that uses ψ as an assumption. By the rule $\neg$I of core logic (which does not allow vacuous discharge) it would then follow that

$$\mathsf{Q}, (\beta\Delta)^\gamma \vdash \neg\psi.$$

Suppose further, for the sake of argument, that

$$\mathsf{Q}, (\beta\Delta)^\gamma, \neg\psi \vdash \bot.$$

By CUT FOR $\bot$ it would follow that

$$\mathsf{Q}, (\beta\Delta)^\gamma \vdash \bot.$$

But this contradicts (†).

It therefore follows by classical logic at the meta level that

$$\text{either} \quad \mathsf{Q}, (\beta\Delta)^\gamma, \psi \nvdash \bot \quad \text{or} \quad \mathsf{Q}, (\beta\Delta)^\gamma, \neg\psi \nvdash \bot.$$

We now proceed to investigate each of these cases.
Case (i): $\mathsf{Q}, (\beta\Delta)^\gamma, \psi \nvdash \bot$.
Clearly $\mathsf{Q}, \psi \nvdash \bot$. By Lemma 6 we therefore have

$$\mathsf{Q}, \psi \vdash \beta(\psi)\,.$$

Suppose for reductio that

$$\mathsf{Q}, (\beta\Delta)^\gamma, \beta(\psi) \vdash \bot.$$

By CUT FOR $\bot$ it would follow that

$$\mathsf{Q}, (\beta\Delta)^\gamma, \psi \vdash \bot,$$

contradicting our case assumption. We therefore conclude by reductio that

$$\mathsf{Q}, (\beta\Delta)^\gamma, \beta(\psi) \nvdash \bot.$$

Clearly, then,

$$\mathsf{Q}, (\beta\Delta)^\gamma \nvdash \bot.$$

By Lemma 7 we therefore have

$$\mathsf{Q}, (\beta\Delta)^\gamma \vdash \neg\neg\psi\,.$$

Hence by $\vee$-Introduction we have

$$\text{either}\quad Q, (\beta\Delta)^{\gamma} \vdash \neg\neg\psi \quad\text{or}\quad Q, (\beta\Delta)^{\gamma} \vdash \neg\psi .$$

Case (ii): $Q, (\beta\Delta)^{\gamma}, \neg\psi \nvdash \bot$. Clearly, then,

$$Q, \neg\psi \nvdash \bot.$$

By Lemma 6 we therefore have

$$Q, \neg\psi \vdash \beta(\neg\psi).$$

Suppose for reductio that

$$Q, (\beta\Delta)^{\gamma}, \beta(\neg\psi) \vdash \bot.$$

By CUT FOR $\bot$ it would follow that

$$Q, (\beta\Delta)^{\gamma}, \neg\psi \vdash \bot,$$

contradicting our case assumption. We therefore conclude by reductio that

$$Q, (\beta\Delta)^{\gamma}, \beta(\neg\psi) \nvdash \bot.$$

Clearly, then,

$$Q, (\beta\Delta)^{\gamma} \nvdash \bot.$$

By Lemma 7 we therefore have

$$Q, (\beta\Delta)^{\gamma} \vdash \neg\neg\neg\psi .$$

Since $Q, (\beta\Delta)^{\gamma} \nvdash \bot$ and $\neg\neg\neg\psi \vdash \neg\psi$, we now have by Corollary 2 (CUT ON CONSISTENT PREMISES) that

$$Q, (\beta\Delta)^{\gamma} \vdash \neg\psi .$$

Hence by $\vee$-Introduction we have

$$\text{either}\quad Q, (\beta\Delta)^{\gamma} \vdash \neg\neg\psi \quad\text{or}\quad Q, (\beta\Delta)^{\gamma} \vdash \neg\psi .$$

That concludes our exploration of each case, arriving at the same disjunctive conclusion in each. It now follows (by proof by cases) that

$$\text{either}\quad Q, (\beta\Delta)^{\gamma} \vdash \neg\neg\psi \quad\text{or}\quad Q, (\beta\Delta)^{\gamma} \vdash \neg\psi . \qquad \square$$

Theorem 3 (Constructivized McGee) *Suppose* $Q, \Delta \nvdash \bot$.
Let ε be a fixed enumeration of all Tarskian biconditionals in $\mathcal{L}$.
Then

1. *for every $\phi \in \Delta$ we have $Q, (\beta\Delta)^{\varepsilon} \vdash \neg\neg\phi$; and*
2. *for every sentence ψ we have either $Q, (\beta\Delta)^{\varepsilon} \vdash \neg\neg\psi$ or $Q, (\beta\Delta)^{\varepsilon} \vdash \neg\psi$.*

Note that the title of this Theorem—"Constructivized McGee"—adverts to the fact that in our version of McGee's result the object-linguistic deducibilities involved are all core deducibilities.

Proof Suppose $Q, \Delta \nvdash \bot$. By Lemma 8, we have $Q, \beta\Delta \nvdash \bot$.
Ad (1): Suppose $\phi \in \Delta$. Clearly $Q, \beta(\phi) \nvdash \bot$. By Lemma 7 we therefore have

$$Q, \beta(\phi) \vdash \neg\neg\phi .$$

But $\beta(\phi) \in \beta\Delta \subseteq (\beta\Delta)^{\varepsilon}$. Hence

$$Q, (\beta\Delta)^{\varepsilon} \vdash \neg\neg\phi .$$

Ad (2): This is Lemma 9. $\square$

6.3 Some final thoughts Theorem 3 says that if Δ is Q-consistent, then the expanded Q-consistent set $(\beta\Delta)^\varepsilon$ *of Tarskian biconditionals* has extraordinary strength:
(1) modulo Q, $(\beta\Delta)^\varepsilon$ core-implies $\neg\neg\Delta$; and
(2) modulo Q, the set $(\beta\Delta)^\varepsilon$ is, constructively, "almost complete."

To attain McGee's original THEOREM 1, all one need do is classicize the logic of the object language by appending, say, the rule of double negation elimination (so as to obtain classical core logic $\mathbb{C}^+$). That makes $(\beta\Delta)^\varepsilon$ imply Δ modulo Q, and also makes $(\beta\Delta)^\varepsilon$ classically complete, modulo Q. For then we have, for every sentence ψ,

$$\text{either} \quad \mathsf{Q}, (\beta\Delta)^\varepsilon \vdash_{\mathbb{C}^+} \psi \quad \text{or} \quad \mathsf{Q}, (\beta\Delta)^\varepsilon \vdash_{\mathbb{C}^+} \neg\psi.$$

One possible choice for Δ is $\emptyset$. Part (1) of Theorem 3 is then trivial; but part (2) in this case says that for every sentence ψ we have

$$\text{either} \;\; \mathsf{Q}, \emptyset^\varepsilon \vdash \neg\neg\psi \;\; \text{or} \;\; \mathsf{Q}, \emptyset^\varepsilon \vdash \neg\psi.$$

Marching down the list of all Tarskian biconditionals, starting empty-handed, considering them one-by-one in the order dictated by ε, and putting them into one's bag if and only if doing so preserves the Q-consistency of its contents, one ends up in the infinite limit with a *complete* consistent arithmetical $\mathcal{L}$-theory $\emptyset^\varepsilon$ extending Q. In particular, $\emptyset^\varepsilon$ will core-decide every sentence in L, the arithmetical language free of *Tr*.

Another possible choice for Δ is Th($\mathbb{N}$) itself. Note that because of Lemma 6 (for all ϕ core-consistent with Q, we have $\mathsf{Q}, \phi \vdash \beta(\phi)$) and the truth of Q, the set β(Th($\mathbb{N}$)) consists only of truths. The set $(\beta(\text{Th}(\mathbb{N})))^\varepsilon$ must, however, on pain of inconsistency, fail to contain *all* Tarskian biconditionals, regardless of the enumeration ε.

We sometimes have to pinch ourselves to be reminded that $(\beta(\text{Th}(\mathbb{N})))^\varepsilon$ is just a set of *Tarskian biconditionals* involving the predicate *Tr*—which, if *Tr really were* a truth predicate, would all be regarded (especially by the Horwichian minimalist) as "trivially true."

It is then surprising, if one is a constructivist, to learn that this set $(\beta(\text{Th}(\mathbb{N})))^\varepsilon$ of supposed trivialities, in conjunction with the very weak arithmetical theory Q, logically implies the whole "constructive content" $\neg\neg$Th($\mathbb{N}$) of the set Th($\mathbb{N}$) of classical arithmetical truths; and just as surprising, if one is a classicist, to learn that, again in conjunction with Q, it logically implies the whole set Th($\mathbb{N}$) itself. How can apparently trivial truths about truth do *that*?

One final thought in closing: on the universally accepted assumption that Th($\mathbb{N}$) is consistent, the set $(\beta(\text{Th}(\mathbb{N})))^\varepsilon$ is of course not effectively enumerable; for, if it were, one could effectively filter out from such an enumeration just those sentences in it that do not contain *Tr*, thereby achieving an effective enumeration of Th($\mathbb{N}$), contrary to Gödel's first incompleteness theorem.

Notes

1. The reader will find these rules in Tennant [16], pp. 161–164.

2. See Tennant [11], pp. 130 and 198. (What the present author called intuitionistic relevant logic at the time of that work he now calls core logic.) Note that the set of all Tarskian

biconditionals does not force the predicate *Tr* to satisfy the principle of bivalence, that is, for all φ, either $Tr(\varphi)$ or not-$Tr(\varphi)$.

3. See Tarski [7]. The bivalence result is Theorem 2, p. 197. The reader will find that strictly classical reasoning intrudes in Tarski's proof of Lemma B on p. 198, on which his Theorem 2 depends.

4. In order to limit sideways spread, we use here the serial forms of ∀E and of →E rather than their parallelized forms.

5. Thanks are owed here to an anonymous referee for pointing out this possible exception. The caveat is necessary if we take the "theories" in question to be logically closed, rather than sets of axioms that are not logically closed.

6. We avoid saying "fixed" here, to avoid any confusion with the fixed points that the function g will afford.

7. In propositional logic, every classically inconsistent set *is* core-inconsistent.

8. The proof-theoretic criterion of paradoxicality was put forward in Tennant [10] and refined in Tennant [12]. See also Tennant [13].

9. Note that an analogous observation holds in the constructive case: core logic $\mathbb{C}$ and intuitionistic logic **I** prove the same consequences from any consistent set of sentences. Note further that these pairs of systems ($\mathbb{C}$ and **I**, and $\mathbb{C}^+$ and **C**) also prove the same inconsistencies.

10. In reprising this part of McGee's argument, Cieśliński [2], p. 697, likewise assumes that it is classical logic that is being used for closure.

11. Note that the classicality of reductio at the final step here is tied to our use of Φ as a "primitive placeholder" for the property with respect to which the expansion is being carried out. In the context of our present concerns, however, Φ will be instantiated as some sort of *consistency* of the sets being constructed. It would therefore be regimentable as $\Omega_i \nvdash \bot$ for the i-th stage of expansion, and as $\Omega^\gamma \nvdash \bot$ for the final result Ω^γ of the expansion. This would enable us to take $\Omega^\gamma \vdash \bot$ as our reductio assumption for this final step, and to use the *constructive* rule ¬-I to draw our final conclusion $\Omega^\gamma \nvdash \bot$. The *other* points of "classicality" pointed out above, however, do not admit of any such constructive work-around.

12. Φ could also be the property of being a conservative extension of some base theory in a particular sublanguage. See the use made of this kind of Φ by Cieśliński [2], in Theorem 1 on p. 698. It asserts the existence of maximal conservative extensions of PA, in a language extending the language of arithmetic by including a truth predicate.

13. Thanks to Steven Dalglish for making me aware of the expository need to stress this point explicitly; and thanks also to an anonymous referee for pointing out that the initial set Ω itself is a separate source of the multiplicity in question.

14. But as pointed out in Section 4 the plausible conjecture lies close at hand that we can do the same without (in an abundance of caution) thus substituting the slightly stronger

theory Q for R. If our conjecture holds, then the upshot would be that all our subsequent 'modulo Q' turnstiles could be 'modulo R' turnstiles that could be read as deducibility in core logic.

References

[1] Boolos, G. S., and R. C. Jeffrey, *Computability and Logic*, Cambridge University Press, London-New York, 1974. Zbl 0298.02003. MR 0342374. 9

[2] Cieśliński, C., "Deflationism, conservativeness and maximality," *Journal of Philosophical Logic*, vol. 36 (2007), no. 6, pp. 695–705. MR 2349879. DOI 10.1007/s10992-007-9057-z. 23

[3] McGee, V., "Maximal consistent sets of instances of Tarski's schema (T)," *Journal of Philosophical Logic*, vol. 21 (1992), pp. 235–41. Zbl 0773.03003. MR 1172215. DOI 10.1007/BF00260929. 14

[4] Odifreddi, P., *Classical Recursion Theory: The Theory of Functions and Sets of Natural Numbers, Vol. 1*, vol. 125 of *Studies in Logic and the Foundations of Mathematics*, Elsevier, Amsterdam-New York, 1989. MR 0982269. 9

[5] Robinson, J., "General Recursive Functions," *Proceedings of the American Mathematical Society*, vol. 1 (1950), no. 6, pp. 703–18. Zbl 0041.15101. MR 0038912. DOI 10.2307/2031973. 9

[6] De Swart, H., "Another intuitionistic completeness proof," *Journal of Symbolic Logic*, vol. 41 (1976), no. 3, pp. 644–62. Zbl 0342.02013. MR 0485329. DOI 10.2307/2272042. 6

[7] Tarski, A., "The concept of truth in formalized languages," pp. 152–278 in *Logic, Semantics, Metamathematics* edited by J. H. Woodger, Clarendon Press, Oxford, 1956 [first published in Polish, 1933]. MR 0078296. 23

[8] Tarski, A., A. Mostowski, and R. M. Robinson, *Undecidable Theories*, North Holland, Amsterdam, 1953. Zbl 0053.00401. MR 0244048. 9

[9] Tennant, N., *Natural Logic*, Edinburgh University Press, Edinburgh, 1978. Zbl 0483.03001. MR 0569938. 9

[10] Tennant, N., "Proof and paradox," *Dialectica*, vol. 36 (1982), pp. 265–96. Zbl 0495.03036. MR 0720601. DOI 10.1111/j.1746-8361.1982.tb00820.x. 23

[11] Tennant, N., *Anti-Realism and Logic: Truth as Eternal*, Clarendon Library of Logic and Philosophy, Oxford University Press, Oxford, 1987. 22

[12] Tennant, N., "On paradox without self-reference," *Analysis*, vol. 55 (1995), pp. 199–207. Zbl 0943.03573. MR 1351564. DOI 10.2307/3328581. 23

[13] Tennant, N., "A new unified account of truth and paradox," *Mind*, vol. 123 (2015), pp. 571–605. Zbl 1316.03005. MR 3413349. DOI 10.1093/mind/fzu179. 23

[14] Tennant, N., "Cut for classical core logic," *Review of Symbolic Logic*, vol. 8 (2015), no. 2, pp. 236–56. Zbl 1337.03082. MR 3347997. DOI 10.1017/S1755020315000088. 6

[15] Tennant, N., "The relevance of premises to conclusions of core proofs," *Review of Symbolic Logic*, vol. 8 (2015), no. 4, pp. 743–84.. Zbl 1375.03078. MR 3423959. DOI 10.1017/S1755020315000040. 6

[16] Tennant, N., *Core Logic*, Oxford University Press, Oxford, 2017. Zbl 1386.03001. MR 3701409. DOI 10.1093/oso/9780198777892.001.0001. 4, 16, 22

[17] Tennant, N., "Core Gödel," *Notre Dame Journal of Formal Logic*, vol. 64 (2023), no. 1, pp. 15–59. Zbl 1541.03163. MR 4564835. DOI 10.1215/00294527-2022-0033. 6, 9

[18] Veldman, W., "An intuitionistic completeness theorem for intuitionistic predicate logic," *Journal of Symbolic Logic*, vol. 41 (1976), no. 1, pp. 159–66. Zbl 0355.02018. MR 0409124. DOI 10.2307/2272955. 6

Acknowledgment

The author is extremely grateful to an anonymous referee for incisive and constructive suggestions. They led to significant improvements in providing background, motivation, and unification of results.

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