Notre Dame Journal of Formal Logic
Volume 34, Number 4, Fall 1993

The Gupta-Belnap Systems $S^{\#}$ and S^{*} are not Axiomatisable

PHILIP KREMER

Abstract Anil Gupta and Nuel Belnap's *The Revision Theory of Truth* presents revision theoretic systems of circular definitions. Part I of the present paper shows that the revision theories $\mathbf{S}^{\#}$ and $\mathbf{S}^{*}$ are not axiomatisable. Part II refines this result. Among other things, Part II shows that there is a strong relationship between revision theories and the theory of inductive definitions. This relationship is exploited to show that $\mathbf{S}^{\#}$ and $\mathbf{S}^{*}$ (and all "plausible" revision theories of circular definitions) are of complexity at least Π^1_2.

1 Introduction ***The Revision Theory of Truth*** Gupta and Belnap [2] treats "____ is true" as a predicate of sentences. (Strictly speaking, "is true" can be meaningfully applied to non-sentences. "Tracy is true" is well-formed and false. "'Snow is white' is true" is well-formed and true.) Furthermore, [2] takes the corresponding concept, *truth*, to be a *circular* concept: in the definition of "is true", the expression "is true", which is the definiendum, appears in the definiens.

More precisely, truth is defined by the set of partial definitions of the form

'p' is true $=_{\mathrm{Df}} p$

where p ranges over the sentences of the language. The definition of truth insofar as truth applies to "snow is white" is not circular. The pertinent definition is

'snow is white' is true $=_{\mathrm{Df}}$ snow is white,

and the definiendum ("is true") does not occur in the definiens. But the definition of truth insofar as truth applies to "what Tracy says is true" *is* circular. The pertinent definition is

'what Tracy says is true' is true $=_{\mathrm{Df}}$ what Tracy says is true,

and the definiendum ("is true") *does* occur in the definiens.

Before considering the special behaviour of truth, [2] develops general semantic theories of circularly defined concepts: revision theories $\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_n, \ldots$ and $\mathbf{S}^{\#}$

Received April 23, 1993; revised June 28, 1993

and $\mathbf{S}^*$. It provides axiomatisations (in a sense) of the $\mathbf{S}_n$, but not of $\mathbf{S}^\#$ and $\mathbf{S}^*$. More precisely, it provides sound and complete calculuses $\mathbf{C}_n$ for the $\mathbf{S}_n$ such that, for every set $\mathfrak{D}$ of definitions and for every formula B, B is a theorem of $\mathbf{C}_n$ (relative to $\mathfrak{D}$) iff B is valid (on $\mathfrak{D}$) in $\mathbf{S}_n$. But it provides no such calculuses for $\mathbf{S}^\#$ and $\mathbf{S}^*$. Indeed, regarding $\mathbf{S}^\#$, [2] explicitly leaves the following open problem:

> Is a [sound and] complete calculus for $\mathbf{S}^\#$ possible? If not, what is the complexity of the theorems of $\mathbf{S}^\#$ (relative to that of $\mathfrak{D}$)? (p. 185, Problem 5D.7)

A similar open problem exists for $\mathbf{S}^*$.

The present paper largely closes these open problems. We provide a finite set, $\mathfrak{D}$, of definitions, and we show that the following sets are not recursively enumerable (indeed, the set of true arithmetical sentences is recursively embeddable into them):

$\{A: A \text{ is valid on } \mathfrak{D} \text{ in } \mathbf{S}^\#\}$, and
$\{A: A \text{ is valid on } \mathfrak{D} \text{ in } \mathbf{S}^*\}$.

(See the Main Result, Corollary 10.)

The Main Result suffices to establish that there is no complete calculus for either $\mathbf{S}^\#$ or $\mathbf{S}^*$. Suppose that there were a complete calculus for $\mathbf{S}^\#$. Then, since our set $\mathfrak{D}$ is finite, we could use the calculus to recursively enumerate $\{A: A$ is valid on $\mathfrak{D}$ in $\mathbf{S}^\#\}$. But this cannot be done. Similarly for $\mathbf{S}^*$.

The Main Result also partially answers the second question in [2]'s Problem 5D.7. Suppose we are looking for a minimal complexity, C, such that for all sets $\mathfrak{D}$ of definitions the complexity of the set of valid sentences (relative to that of $\mathfrak{D}$) is at most C. Then C is at least the complexity of true arithmetic (which is Δ^1_1).

The present paper is divided into two parts. Part I (Sections 2–5) establishes the Main Result. (Part I assumes much the same background as does [2].) Part II (Sections 6–10) refines the Main Result. (In Part II, Sections 6, 7, and 10 assume the same background as Part I, and Sections 8 and 9 assume some familiarity with the analytic hierarchy.) Section 6 sharpens the Main Result by using a single circular definition rather than the two circular definitions of Part I. Section 7 generalises the Main Result to other "plausible" revision theories. Section 8 investigates the relationship between revision theories and the theory of inductive definitions—a relationship which can be brought out by generalising the construction used in Part I. This relationship helps us improve upon the Main Result: the lower bound for the complexity of $\mathbf{S}^\#$ and $\mathbf{S}^*$ is raised to Π^1_2 (Corollary 17). Section 9 proves the main lemma used in proving Corollary 17. And Section 10 discusses [2]'s theories of truth, $\mathbf{T}^\#$ and $\mathbf{T}^*$, which are based on $\mathbf{S}^\#$ and $\mathbf{S}^*$. Throughout this paper, we assume all of the terminology and notation of [2] with one exception: where [2] uses script S (for revision sequences), we use bold-italic S ($\boldsymbol{S}$).

Part I The Main Result

2 Preliminaries Let L be the first-order language of arithmetic, with the following nonlogical constants: a name, 0; a unary function symbol, $'$; binary function symbols, + and ×; and a binary relation symbol, <. Let $\mathfrak{D}$ be the set consisting of the following two definitions:

$$Gx =_{\mathrm{Df}} [\exists z \forall y(y < z \leftrightarrow Gy) \,\&\, \forall y(y < x \rightarrow Gy)]$$

$$Hx =_{\mathrm{Df}} [\forall y(Gy \rightarrow Gy') \,\&\, \forall y \forall z((Gz \,\&\, y < z) \rightarrow Gy) \,\&\, \exists y Gy \,\&\, Gx]$$
$$\vee \; [[\exists y(Gy \,\&\, {\sim}Gy') \vee \exists y \exists z(Gz \,\&\, y < z \,\&\, {\sim}Gy) \vee {\sim}\exists y Gy] \,\&\, Hx].$$

The upshot of these definitions is made explicit in Section 3, where we note the behaviour of the revision rule $\delta_{\mathfrak{D},\mathrm{M}}$ and of revision sequences for $\delta_{\mathfrak{D},\mathrm{M}}$, when M is in a particular class of models soon to be specified.

First, let AX be the sentence which says that < is a strict linear order, that 0 is the <-smallest element, and that x' is indeed the immediate successor of x (see [2], Example 5A.17, p. 154):

$$\forall x \sim (x < x) \,\&\, \forall x \forall y \forall z((x < y \,\&\, y < z) \rightarrow x < z)$$
$$\&\, \forall x \forall y(x < y \vee x = y \vee y < x) \,\&\, \forall x \sim (x < 0)$$
$$\&\, \forall x(x < x') \,\&\, \forall x \forall y(x < y \rightarrow (y = x' \vee x' < y)).$$

We say that a model M of L is an AX-model iff it satisfies AX. (A good example is the set of natural numbers with the standard interpretation of the non-logical constants.) We are interested in the behaviour of $\delta_{\mathfrak{D},\mathrm{M}}$ (and of its revision sequences) when M is an AX-model. In our discussion below, if $\mathrm{M} = \langle \mathrm{D}, \mathrm{I} \rangle$ is a model of L, we write, for $\mathrm{d}, \mathrm{d}_1, \mathrm{d}_2 \in \mathrm{D}$,

d'	for $\mathrm{I}(')(\mathrm{d})$;
$\mathrm{d}_1 < \mathrm{d}_2$	for $\langle \mathrm{d}_1, \mathrm{d}_2 \rangle \in \mathrm{I}(<)$; and
0	for I(0).

Finally, given a hypothesis, h, we define $\mathrm{h}(G)$ to be $\{\mathrm{d} \in \mathrm{D} : \mathrm{h}(G,\mathrm{d}) = \mathbf{t}\}$, and $\mathrm{h}(H)$ to be $\{\mathrm{d} \in \mathrm{D} : \mathrm{h}(H,\mathrm{d}) = \mathbf{t}\}$.

Before we move on to the behaviour of $\delta_{\mathfrak{D},\mathrm{M}}$, some definitions.

Definition 1 Suppose that $\mathrm{M} = \langle \mathrm{D}, \mathrm{I} \rangle$ is a model of L. Then $\mathrm{S} \subseteq \mathrm{D}$ is *'-complete* iff $(\forall \mathrm{d} \in \mathrm{S})(\mathrm{d}' \in \mathrm{S})$. We say $\mathrm{S} \subseteq \mathrm{D}$ is an *initial segment* iff $(\forall \mathrm{d}_1 \in \mathrm{S})(\forall \mathrm{d}_2 \in \mathrm{D})$ (if $\mathrm{d}_2 < \mathrm{d}_1$ then $\mathrm{d}_2 \in \mathrm{S}$). For every $\mathrm{d}_1, \mathrm{d}_2 \in \mathrm{D}$, $[\mathrm{d}_1, \mathrm{d}_2] = \{\mathrm{d} \in \mathrm{D} : \mathrm{d}_1 \leq \mathrm{d} \leq \mathrm{d}_2\}$. For every n, $\mathrm{d}^{(\mathrm{n})} = \mathrm{d}''^{\cdots\prime}$, where $'$ occurs n times. Finally, $\mathbb{N}_\mathrm{M}$ is the smallest '-complete set containing 0. (That is, $\mathbb{N}_\mathrm{M} = \{0, 0', 0'', 0''', \ldots\}$.)

3 Behaviour of $\delta_{\mathfrak{D},\mathrm{M}}$ and of revision sequences for $\delta_{\mathfrak{D},\mathrm{M}}$

We prove Lemmas 2–5 in Section 4.

Lemma 2 *Suppose that* $\mathrm{M} = \langle \mathrm{D}, \mathrm{I} \rangle$ *is an AX-model and that* h *is a hypothesis. Then*

(i) *if* $\mathrm{h}(G) = [0, \mathrm{d}]$ *then* $\delta_{\mathfrak{D},\mathrm{M}}(\mathrm{h})(G) = [0, \mathrm{d}']$;
(ii) *if* $\mathrm{h}(G) = \varnothing$ *then* $\delta_{\mathfrak{D},\mathrm{M}}(\mathrm{h})(G) = \{0\}$; *and*
(iii) *otherwise,* $\delta_{\mathfrak{D},\mathrm{M}}(\mathrm{h})(G) = \varnothing$.

Lemma 3 *Suppose that* $\mathrm{M} = \langle \mathrm{D}, \mathrm{I} \rangle$ *is an AX-model and that* h *is a hypothesis. Then*

(i) *if* $\mathrm{h}(G)$ *is a non-empty ′-complete initial segment, then* $\delta_{\mathfrak{D},\mathrm{M}}(\mathrm{h})(H) = \mathrm{h}(G)$; *and*
(ii) *otherwise*, $\delta_{\mathfrak{D},\mathrm{M}}(\mathrm{h})(H) = \mathrm{h}(H)$.

Lemma 4 *Suppose that* $\mathrm{M} = \langle \mathrm{D},\mathrm{I}\rangle$ *is an AX-model and that* $\boldsymbol{S}$ *is a revision sequence for* $\delta_{\mathfrak{D},\mathrm{M}}$. *Then*

(i) $\boldsymbol{S}_{\omega}(G)$ *is a non-empty, ′-complete initial segment*;
(ii) $\boldsymbol{S}_{\omega+1}(G) = \varnothing$;
(iii) $\boldsymbol{S}_{2\omega}(G) = \mathbb{N}_{\mathrm{M}}$;
(iv) $(\forall\alpha > \omega)(\boldsymbol{S}_{\alpha}(G) \subseteq \mathbb{N}_{\mathrm{M}})$;
(v) *for every limit ordinal*, λ, $\boldsymbol{S}_{\lambda+\omega}(G) = \mathbb{N}_{\mathrm{M}}$; *and*
(vi) *for every limit ordinal*, λ, $\boldsymbol{S}_{\lambda+\omega+1}(G) = \varnothing$.

Lemma 5 *Suppose that* $\mathrm{M} = \langle \mathrm{D},\mathrm{I}\rangle$ *is an AX-model and that* $\boldsymbol{S}$ *is a revision sequence for* $\delta_{\mathfrak{D},\mathrm{M}}$. *Then*

(i) $\boldsymbol{S}_{2\omega+1}(H) = \mathbb{N}_{\mathrm{M}}$; *and*
(ii) $(\forall\alpha > 2\omega)(\boldsymbol{S}_{\alpha}(H) = \mathbb{N}_{\mathrm{M}})$.

Corollary 6 *Suppose that* $\mathrm{M} = \langle \mathrm{D},\mathrm{I}\rangle$ *is an AX-model. Then, if* h *is recurring for* $\delta_{\mathfrak{D},\mathrm{M}}$, *then* $\mathrm{h}(H) = \mathbb{N}_{\mathrm{M}}$.

Proof: By Lemma 5(ii).

4 Proofs of Lemmas 2–5 For these proofs, we introduce four abbreviations. (Notice that D, below, is logically equivalent to $\sim C$.)

A abbreviates $\exists z\forall y(y < z \leftrightarrow Gy)$;
Bx abbreviates $\forall y(y < x \rightarrow Gy)$;
C abbreviates $[\forall y(Gy \rightarrow Gy') \;\&\; \forall y\forall z((Gz \;\&\; y < z) \rightarrow Gy) \;\&\; \exists yGy]$; and
D abbreviates $[\exists y(Gy \;\&\; \sim Gy') \vee \exists y\forall z(Gz \;\&\; y < z \;\&\; \sim Gy) \vee \sim\exists yGy]$.

So our set $\mathfrak{D}$ of definitions (form Section 2) is

$$Gx =_{\mathrm{Df}} [A \;\&\; Bx]$$

$$Hx =_{\mathrm{Df}} [C \;\&\; Gx] \vee [D \;\&\; Hx].$$

Proof of Lemma 2: (i) Suppose that $\mathrm{h}(G) = [0,\mathrm{d}]$. Then A is true in $\mathrm{M} + \mathrm{h}$. Furthermore, in $\mathrm{M} + \mathrm{h}$, Bx is true of all and only the members of $[0,\mathrm{d}']$. (ii) Suppose that $\mathrm{h}(G) = \varnothing$. Then A is true in $\mathrm{M} + \mathrm{h}$. Furthermore, in $\mathrm{M} + \mathrm{h}$, Bx is true of 0 and only 0. (iii) Suppose that $\mathrm{h}(G) \neq \varnothing$ and that for every $\mathrm{d} \in \mathrm{D}$, $\mathrm{h}(G) \neq [0,\mathrm{d}]$. Then A is false in $\mathrm{M} + \mathrm{h}$. So, in $\mathrm{M} + \delta_{\mathfrak{D},\mathrm{M}}(\mathrm{h})$, Gx is false of every $\mathrm{d} \in \mathrm{D}$.

Proof of Lemma 3: (i) Suppose that $\mathrm{h}(G)$ is a non-empty ′-complete initial segment. Then D is false in $\mathrm{M} + \mathrm{h}$. So $(D \;\&\; Hx)$ is false in $\mathrm{M} + \mathrm{h}$, of every $\mathrm{d} \in \mathrm{D}$. Furthermore, C is true in $\mathrm{M} + \mathrm{h}$. So, in $\mathrm{M} + \mathrm{h}$, $(C \;\&\; Gx)$ is true of $\mathrm{d} \in \mathrm{D}$ just in case Gx is true of d. So Hx is true of d in $\mathrm{M} + \delta_{\mathfrak{D},\mathrm{M}}(\mathrm{h})$ just in case Gx is true of d in $\mathrm{M} + \mathrm{h}$. (ii) Suppose that $\mathrm{h}(G)$ is *not* a non-empty ′-complete initial segment. Then $(C \;\&\; Gx)$ is false, in $\mathrm{M} + \mathrm{h}$, of every $\mathrm{d} \in \mathrm{D}$. Furthermore, D is true in $\mathrm{M} + \mathrm{h}$. So, $(D \;\&\; Hx)$ true of d in $\mathrm{M} + \mathrm{h}$ just in case Hx is true of d in $\mathrm{M} + \mathrm{h}$. So, Hx is true of d in $\mathrm{M} + \delta_{\mathfrak{D},\mathrm{M}}(\mathrm{h})$ just in case Hx is true of d in $\mathrm{M} + \mathrm{h}$.

Proof of Lemma 4: (i) We consider three cases. (a) $\mathcal{S}_0(G) = [0,d]$. In this case, by Lemma 2(i), for every $n > 0$, $\mathcal{S}_n(G) = [0,d^{(n)}]$. Let $D_d = [0,d] \cup \{d^{(n)}: n$ is a natural number$\}$. Notice that for every $c \in D$, $\langle G,c\rangle$ is stably **t** in $\mathcal{S} \upharpoonright \omega$ iff $c \in D_d$; and $\langle G,c\rangle$ is stably **f** in $\mathcal{S} \upharpoonright \omega$ iff $c \notin D_d$. So, since $\mathcal{S}_\omega$ coheres with $\mathcal{S} \upharpoonright \omega$ ([2], Definitions 5C.2 and 5C.3), $\mathcal{S}_\omega(G) = D_d$, which is a non-empty, $'$-complete initial segment. (b) $\mathcal{S}_0(G) = \varnothing$. In this case, by Lemma 2(ii) and 2(i), for every $n > 0$, $\mathcal{S}_n(T) = [0,0^{(n-1)}]$. By an argument like that in case (a), $\mathcal{S}_\omega(G) = \mathbb{N}_M$, which is a non-empty, $'$-complete initial segment. (c) $\mathcal{S}_0(G) \neq \varnothing$ and for every $d \in D$, $\mathcal{S}_0(G) \neq [0,d]$. Then, by Lemma 2(ii), $\mathcal{S}_1(G) = \varnothing$. So, by Lemma 2(ii) and Lemma 2(i), for every $n > 1$, $\mathcal{S}_n(G) = [0,0^{(n-2)}]$. So, by an argument like that in case (a), $\mathcal{S}_\omega(G) = \mathbb{N}_M$, which is a non-empty, $'$-complete initial segment. (ii) This follows from Lemma 4(i) and Lemma 2(iii). (iii) This is proved by an argument like that for Lemma 4(i)(c). (iv) This is proved by induction on α. The base case $(\omega + 1)$ is by Lemma 4(ii). If α is a successor ordinal then the inductive hypothesis and Lemma 2 do the trick. If α is a limit ordinal then the induction hypothesis and the coherence requirement on revision sequences do the trick. (v) This is proved by an argument like that for Lemma 4(i) (where we consider three cases for $\mathcal{S}_\lambda$ corresponding to the three cases for $\mathcal{S}_0$), $\mathcal{S}_{\lambda+\omega}(G)$ is a non-empty, $'$-complete initial segment. By Lemma 4(iv), $\mathcal{S}_{\lambda+\omega}(G) \subseteq \mathbb{N}_M$. So $\mathcal{S}_{\lambda+\omega}(G) = \mathbb{N}_M$. (vi) This follows from Lemma 4(v) and Lemma 2(iii).

Proof of Lemma 5: (i) This follows from Lemma 4(iii) and Lemma 3(i). (ii) This is proved by induction on α. The base case is just Lemma 5(i). The limit case depends on the inductive hypothesis and on the coherence requirement on revision sequences. For the successor case, suppose that $\alpha = \beta + 1$, and consider two subcases. (a) $\mathcal{S}_\beta(G)$ is a non-empty $'$-complete initial segment. Then, by Lemma 4(iv), $\mathcal{S}_\beta(G) = \mathbb{N}_M$. So, by Lemma 2(i), $\mathcal{S}_\alpha(H) = \mathcal{S}_\beta(G) = \mathbb{N}_M$. (b) $\mathcal{S}_\beta(G)$ is not a non-empty $'$-complete initial segment. Then, by Lemma 3(ii), $\mathcal{S}_\alpha(H) = \mathcal{S}_\beta(H) = \mathbb{N}_M$ (by the inductive hypothesis).

5 Proof of the main result Let $N = \langle \omega, I_\omega\rangle$ be the standard model of arithmetic. N is clearly a model of AX. N is also a model of the sentence PA^- where PA^- is the sentence formed by universally closing all of the Peano axioms other than the axioms of induction, and then conjoining them. Indeed, we have the following. (We omit the proof.)

Lemma 7 *Suppose that* $M = \langle D,I\rangle$ *is an AX-model such that* (i) $\mathbb{N}_M = D$; *and* (ii) M *is a model of* PA^-. *Then* M *is isomorphic to* N.

Our main lemma is Lemma 8.

Lemma 8 *For every formula B of L*,

$N \models B$ *iff* $\models^{\mathfrak{D}}_{*}[(AX \,\&\, PA^- \,\&\, \forall x Hx) \rightarrow B]$ *iff* $\models^{\mathfrak{D}}_{\#}[(AX \,\&\, PA^- \,\&\, \forall x Hx) \rightarrow B]$.

Proof:

(i) $N \models B \Rightarrow\ \models^{\mathfrak{D}}_{*}[(AX \,\&\, PA^- \,\&\, \forall x Hx) \rightarrow B]$: Suppose that $N \models B$. Suppose also that $\not\models^{\mathfrak{D}}_{*}[(AX \,\&\, PA^- \,\&\, \forall x Hx) \rightarrow B]$. Then, there is a model $M = \langle D,I\rangle$ of L, and a recurring hypothesis h such that $M + h \models (AX \,\&\, PA^- \,\&\, \forall x Hx)$ and $M + h \not\models B$. Since h is recurring, $h(H) = \mathbb{N}_M$, by Corollary 6. So, since

$M + h \models \forall x Hx$, $D = \mathbb{N}_M$. So, by Lemma 7, M is isomorphic to N. So $M \models B$. So $M + h \models B$, which contradicts $M + h \not\models B$.

(ii) $\models^{\mathfrak{D}}_{*}[(AX \,\&\, PA^- \,\&\, \forall x Hx) \to B] \Rightarrow \models^{\mathfrak{D}}_{\#}[(AX \,\&\, PA^- \,\&\, \forall x Hx) \to B]$: See RTT, Theorem 5D.22 (p. 191).

(iii) $\models^{\mathfrak{D}}_{*}[(AX \,\&\, PA^- \,\&\, \forall x Hx) \to B] \Rightarrow N \models B$: Suppose that $\models^{\mathfrak{D}}_{*}[(AX \,\&\, PA^- \,\&\, \forall x Hx) \to B]$. Notice first that $\mathbb{N}_N = \omega$. Now let h be any hypothesis which is recurring for $\delta = \delta_{\mathfrak{D},N}$. (Such a hypothesis exists by [2], Theorem 5C.7(i), p. 170.) So there is a natural number p such that $N + \delta^p(h) \models [(AX \,\&\, PA^- \,\&\, \forall x Hx) \to B]$. Now, since $N \models AX \,\&\, PA^-$, and since G and H do not occur in $(AX \,\&\, PA^-)$, $N + \delta^p(h) \models AX \,\&\, PA^-$. Furthermore, since h is recurring for δ, $\delta^p(h)$ is recurring for δ. So, by Corollary 6, and by the fact that $\mathbb{N}_N = \omega$, $N + \delta^p(h) \models \forall x Hx$. And so $N + \delta^p(h) \models B$. Finally, since neither G nor H is in the vocabulary of B, $N \models B$, as desired.

Before we state the Main Result (Corollary 10), we give a definition which makes it precise.

Definition 9 Suppose that $A \subseteq X$ and $B \subseteq Y$, where each of X and Y is either the set of natural numbers or the set of formulas of a language whose syntax can be recursively arithmetised. A is *recursively embeddable* in B iff, for some 1–1 recursive function $f : X \to Y$, and for all $x \in X$, $x \in A$ iff $f(x) \in B$.

Corollary 10 (The Main Result) *The set of true arithmetic sentences is recursively embeddable in* $\{A: A$ is valid on $\mathfrak{D}$ in $\mathbf{S}^{\#}\}$ *and in* $\{A: A$ is valid on $\mathfrak{D}$ on $\mathbf{S}^{*}\}$.

Proof: Let f be the following function on the set of formulas of L:

$$f(B) = [(AX \,\&\, PA^- \,\&\, \forall x Hx) \to B].$$

Then, by Lemma 8, B is a truth of arithmetic iff $f(B)$ is valid on $\mathfrak{D}$ in $\mathbf{S}^{\#}$ iff $f(B)$ is valid on $\mathfrak{D}$ in $\mathbf{S}^{*}$.

As noted in Section 1, the non-axiomatisability of $\mathbf{S}^{\#}$ and $\mathbf{S}^{*}$ (and indeed that their complexity is at least Δ^1_1) follows from Corollary 10.

Part II Refinements

6 Using one definition instead of two Though our set $\mathfrak{D}$ of definitions contains circular definitions for *two* predicates, we can alter our proof so that the Main Result applies to the $\mathbf{S}^{\#}$ and $\mathbf{S}^{*}$ theories of a *single* definition (of a unary predicate). The trick is to define a unary predicate F whose behaviour on the odd numbers does the job of G and whose behaviour on the even numbers does the job of H. In order to make the distinction between even and odd numbers precise, we assume that we are working in models of $(AX \,\&\, PA^-)$ rather than in models of AX. We use two abbreviations:

(x is even) abbreviates $(\exists w)(x = (0'' \times w))$; and
(x is odd) abbreviates $\sim(\exists w)(x = (0'' \times w))$.

The definition of Fx is:

$$
\begin{aligned}
Fx =_{\mathrm{Df}} & [(x \text{ is odd}) \,\&\, \exists z \forall y (y < z \,\&\, (y \text{ is odd}) \leftrightarrow Fy) \\
& \quad \&\, \forall y (y < x \,\&\, (y \text{ is odd}) \rightarrow Fy)] \\
& \vee\ [(x \text{ is even}) \,\&\, \forall y (Fy \,\&\, (y \text{ is odd}) \rightarrow Fy'') \\
& \qquad \&\, \forall y \forall z ((Fz \,\&\, y < z \,\&\, (z \text{ is odd}) \,\&\, (y \text{ is odd})) \rightarrow Fy) \\
& \qquad \&\, \exists y ((y \text{ is odd}) \,\&\, Fy) \,\&\, Fx'] \\
& \vee\ [(x \text{ is even}) \,\&\, [\exists y ((y \text{ is odd}) \,\&\, Fy \,\&\, {\sim}Fy'') \\
& \qquad\qquad \vee \exists y \exists z ((y \text{ is odd}) \,\&\, (z \text{ is odd}) \,\&\, Fz \,\&\, y \\
& \qquad\qquad < z \,\&\, {\sim}Fy) \vee {\sim}\exists y ((y \text{ is odd}) \,\&\, Fy)] \,\&\, Fx].
\end{aligned}
$$

The analogue to Lemma 8 is the following. For every formula B of L,

$$
\begin{aligned}
& \mathrm{N} \models B \text{ iff} \\
& \quad \models_{*}^{\mathfrak{D}} [(AX \,\&\, PA^{-} \,\&\, \forall x ((x \text{ is even}) \rightarrow Fx)) \rightarrow B] \text{ iff} \\
& \quad \models_{\#}^{\mathfrak{D}} [(AX \,\&\, PA^{-} \,\&\, \forall x ((x \text{ is even}) \rightarrow Fx)) \rightarrow B].
\end{aligned}
$$

So the complexity of the $\mathbf{S}^{\#}$ and $\mathbf{S}^{*}$ theories of a *single* definition of (a unary predicate) is at least that of arithmetic (which is Δ_1^1).

7 Other revision theories $\mathbf{S}^{\#}$ and $\mathbf{S}^{*}$ are not the only plausible revision theories for sets of circular definitions. And (as pointed out by Gupta in correspondence) our results apply to any plausible such theory, since our proofs do not rely on the special features of $\mathbf{S}^{\#}$ and $\mathbf{S}^{*}$.

The idea guiding $\mathbf{S}^{*}$, the simplest revision theory, is this (where we fix a model M and a set $\mathfrak{D}$ of definitions): the *valid* sentences are those which come out stably true in every On-long (where On is the class of ordinals) revision sequence (for $\delta_{\mathfrak{D},\mathrm{M}}$). More precisely, given the α^{th} stage in a revision sequence $\boldsymbol{S}$, we can use the hypothesis $\boldsymbol{S}_\alpha$ to evaluate the truth values of the sentences. This results in an *evaluation* sequence of assignments of truth-values to sentences. So the idea guiding $\mathbf{S}^{*}$ is this: the *valid* sentences are those which come out stably true in every On-long evaluation sequence.

We can liberalise the idea guiding $\mathbf{S}^{*}$ (see [2], p. 168). One way is to restrict our attention to some subfamily of the On-long revision sequences—those considered somehow well-behaved. And so we might not insist that, in order for a sentence to be valid, the sentence to stably true in *all* evaluation sequences, but only in the well-behaved ones. Another strategy is not to restrict the family of revision sequences, but to weaken the insistence that, in order to be valid, a sentence be *stably* true (in all On-long evaluation sequences). In $\mathbf{S}^{\#}$, for example, the valid sentences are those which come out *nearly* stably true ([2], Definition 5C.5, p. 169) in every On-long evaluation sequence.

However we liberalise the idea guiding $\mathbf{S}^{*}$, this much seems certain: being *stably true* in *all* evaluation sequences ought to be a *sufficient* (if not a necessary) condition for validity. This motivates clause (5) of Definition 11, below. The other clauses are motivated by more general concerns. We do not provide a precise definition of the concept of a *revision theory* **S**. We assume that **S** is based on On-long revision sequences, and is guided by a liberalised version of idea guiding $\mathbf{S}^*$. And we assume that *some* definition has been given for "A is valid on $\mathfrak{D}$ in M in $\mathbf{S}$", and that *A is valid on $\mathfrak{D}$ in* $\mathbf{S}$ iff, for every model M (of the language L) A is valid on $\mathfrak{D}$ in M in $\mathbf{S}$.

Definition 11 A revision theory $\mathbf{S}$ is *plausible* iff, for every model M (of the original language) and every set $\mathfrak{D}$ of definitions and every sentence A (of the original language extended with the definienda):

(1) if A is valid on $\mathfrak{D}$ in M in $\mathbf{S}$ then $\sim A$ is not valid on $\mathfrak{D}$ in M in $\mathbf{S}$;
(2) if A and B are valid on $\mathfrak{D}$ in M in $\mathbf{S}$ then $(A \mathbin{\&} B)$ is valid on $\mathfrak{D}$ in M in $\mathbf{S}$;
(3) if A is valid on $\mathfrak{D}$ in M in $\mathbf{S}$ and $(A \rightarrow B)$ is a theorem of classical logic then B is valid on $\mathfrak{D}$ in M in $\mathbf{S}$;
(4) if A contains no definienda and A is classically validated by M then A is valid on $\mathfrak{D}$ in M in $\mathbf{S}$; and
(5) if A is valid on $\mathfrak{D}$ in M in $\mathbf{S}^*$ then A is valid on $\mathfrak{D}$ in M in $\mathbf{S}$.

Theorem 12 *If* $\mathbf{S}$ *is a plausible revision theory then the complexity of* $\mathbf{S}$ (*indeed, the complexity of the* $\mathbf{S}$ *theory of a single definition of a unary predicate*) *is at least* Δ^1_1.

Remark 13 We can identify a revision theory $\mathbf{S}$ with the corresponding three-place validity relation $\mathbf{S}(A,\mathfrak{D},\mathrm{M})$ = "A is valid on $\mathfrak{D}$ in M in $\mathbf{S}$". Indeed, we can take *any* three-place relation $\mathbf{S} = \mathbf{S}(A,\mathfrak{D},\mathrm{M})$ to be a theory of circular definitions, even if it is in no interesting sense a *revision* theory. If $\mathbf{S} = \mathbf{S}(A,\mathfrak{D},\mathrm{M})$ satisfies conditions (1)–(5) of Definition 11, we say that $\mathbf{S}$ is a *plausible* theory of circular definitions. Given a theory $\mathbf{S} = \mathbf{S}(A,\mathfrak{D},\mathrm{M})$ of circular definitions, we say that *A is valid on $\mathfrak{D}$ in* $\mathbf{S}$ iff, for every model M, $\mathbf{S}(A,\mathfrak{D},\mathrm{M})$. The complexity of $\mathbf{S}$ (in the sense we are interested in) is the following: the minimal complexity C such that, for all sets $\mathfrak{D}$ of definitions, the complexity of $\{A$: *A is valid on $\mathfrak{D}$ in* $\mathbf{S}\}$ (relative to that of $\mathfrak{D}$) is at most C. The complexity of the $\mathbf{S}$ theory of a single definition of a unary predicate is the following: the minimal complexity C such that, for all sets $\mathfrak{D} = \{\boldsymbol{D}\}$ where $\boldsymbol{D}$ defines a unary predicate, the complexity of $\{A$: *A is valid on $\mathfrak{D}$ in* $\mathbf{S}\}$ is at most C. (We can similarly define the complexity of the $\mathbf{S}$ theory of finitely many definitions, or of finitely many definitions of unary predicates.) Theorem 12 goes through for any *plausible* theory $\mathbf{S}$ of circular definitions, even if $\mathbf{S}$ is in no interesting sense a *revision* theory. Indeed, all our theorems below concerning plausible *revision* theories of circular definitions can be generalised to plausible theories of circular definitions.

8 Revision-theoretic definitions, and inductive and co-inductive definitions

Gupta has pointed out (in correspondence) that our construction can be generalised to show that every set or relation that is inductively (or co-inductively) definable in a given model is revision-theoretically definable in that model. We can take advantage of this to improve our lower bound for the complexity of $\mathbf{S}^{\#}$ and $\mathbf{S}^*$ (and any other plausible revision theory) by raising it from Δ^1_1 to Π^1_2 (Corollary 17—the hard work is in proving Lemma 16, which we do in Section 9).

Gupta defines a "translation" of any inductive definition into a pair of revision theoretic definitions which have the effect of the original inductive definition. Suppose that we fix a model, $\mathrm{M} = \langle \mathrm{D,I} \rangle$. Consider a single definition, $\boldsymbol{D}$, which is *positive* in the sense of [2], Definition 5D.24 (i.e., all the occurrences of the definiendum in the definiens are positive):

$$Hx =_{\mathrm{Df}} A(x,H).$$

Since $\boldsymbol{D}$ is positive, $\delta_{\{\boldsymbol{D}\},\mathrm{M}}$ is a monotone operator on the space of hypotheses for H, where these hypotheses are ordered as follows: $\mathrm{h}_1 \leq \mathrm{h}_2$ iff $(\forall \mathrm{d} \in \mathrm{D})$(if $\mathrm{h}_1(H,\mathrm{d}) = \mathbf{t}$ then $\mathrm{h}_2(H,\mathrm{d}) = \mathbf{t}$). Under these circumstances, in the system $\mathbf{S}_{\mathrm{i}}$ of inductive definitions ([2], p. 193) the interpretation of H is given by the least fixed point of $\delta_{\{\boldsymbol{D}\},\mathrm{M}}$. The extension thus assigned to H is *inductively defined* by $\boldsymbol{D}$. Also, a sentence A is *inductively valid* (or *valid in* $\mathbf{S}_{\mathrm{i}}$) *on* $\boldsymbol{D}$ *in* M (notation: $\mathrm{M} \models_{\mathrm{i}}^{\boldsymbol{D}} A$) iff A is true in the model which extends M by giving H the least fixed point interpretation. Finally a sentence A is *inductively valid* (or *valid in* $\mathbf{S}_{\mathrm{i}}$) *on* $\boldsymbol{D}$ (notation: $\models_{\mathrm{i}}^{\boldsymbol{D}} A$) iff, for every model M of the original language, A is inductively valid on $\boldsymbol{D}$ in M. (See [2], Definition 5D.25, p. 193.)

Let X be the set inductively defined by $\boldsymbol{D}$. Gupta notes that we can define X in any plausible revision theory as follows. For any unary predicate, F, let

(F is sound) abbreviate $\forall x(Fx \to A(x,F))$;
(F is replete) abbreviate $\forall x(A(x,F) \to Fx)$; and
(F is a fixed point) abbreviate (F is sound and replete).

Let $\mathfrak{D}$ be the set consisting of the following two circular definitions:

$$Gx =_{\mathrm{Df}} [(G \text{ is sound}) \mathbin{\&} {\sim}(G \text{ is replete}) \mathbin{\&} A(x,G)]$$

$$Hx =_{\mathrm{Df}} [(G \text{ is a fixed point}) \mathbin{\&} Gx] \vee [{\sim}(G \text{ is a fixed point}) \mathbin{\&} Hx].$$

The revision process for these definitions yields for H the same interpretation as the original definition $\boldsymbol{D}$, where $\boldsymbol{D}$ is understood as an inductive definition. More precisely, for every revision sequence $\boldsymbol{S}$ for $\delta_{\mathfrak{D},\mathrm{M}}$ and for every $\mathrm{d} \in \mathrm{D}$,

$\mathrm{d} \in \mathrm{X}$ iff $\langle H,\mathrm{d} \rangle$ is stably $\mathbf{t}$ in $\boldsymbol{S}$, and $\mathrm{d} \notin \mathrm{X}$ iff $\langle H,\mathrm{d} \rangle$ is stably $\mathbf{f}$ in $\boldsymbol{S}$.

And so, in any plausible revision theory, Hx strongly defines X (in the sense analogous to that of [2], Definition 5D.18).

Gupta also defines a translation of any *co*-inductive definition into a pair of revision theoretic definitions which have the effect of the original co-inductive definition. We can define a system $\mathbf{S}_{\mathrm{ci}}$ of co-inductive definitions analogously to the system $\mathbf{S}_{\mathrm{i}}$ of inductive definitions, and we can define $\mathbf{S}_{\mathrm{ci}}$ validity analogously to $\mathbf{S}_{\mathrm{i}}$ validity. In $\mathbf{S}_{\mathrm{ci}}$ the extension of H is given by the *greatest* fixed point of $\delta_{\{\boldsymbol{D}\},\mathrm{M}}$, and we say that this extension is *co*-inductively defined by $\boldsymbol{D}$.

Suppose that Y is this extension. Gupta notes that we can strongly define Y in any plausible revision theory with the use of the following set $\mathfrak{D}'$ of definitions:

$$Gx =_{\mathrm{Df}} [(G \text{ is sound}) \vee {\sim}(G \text{ is replete}) \vee A(x,G)]$$

$$Hx =_{\mathrm{Df}} [(G \text{ is a fixed point}) \mathbin{\&} Gx] \vee [{\sim}(G \text{ is a fixed point}) \mathbin{\&} Hx].$$

The revision process for these definitions yields for H the same interpretation as original definition $\boldsymbol{D}$, now understood as a co-inductive definition. More precisely, for every revision sequences $\boldsymbol{S}$ for $\delta_{\mathfrak{D}',\mathrm{M}}$ and for every $\mathrm{d} \in \mathrm{D}$,

$\mathrm{d} \in \mathrm{Y}$ iff $\langle H,\mathrm{d} \rangle$ is stably $\mathbf{t}$ in $\boldsymbol{S}$, and $\mathrm{d} \notin \mathrm{Y}$ iff $\langle H,\mathrm{d} \rangle$ is stably $\mathbf{f}$ in $\boldsymbol{S}$.

And so, in any plausible revision theory, Hx strongly defines Y (in the sense analogous to that of [2], Definition 5D.18).

The Gupta translations immediately yield the following:

Theorem 14 *If* S *is a plausible revision theory, then the complexity of the* S *theory of finitely many definitions of unary predicates (and hence the complexity of* S*) is at least that of the* $S_i(S_{ci})$ *theory of a single definition of a unary predicate.*

So far we have defined S_i-validity when a single positive definition $\boldsymbol{D}$ is given, but [2] defines S_i-validity given any (finite or infinite) positive set $\mathfrak{D}$ of definitions. We could analogously define S_{ci}-validity given any positive set of definitions. Now the Gupta translations can be generalised so as to apply not just to single positive definitions of unary predicates, but to any positive set of definitions. These generalised Gupta translations assign two revision theoretic definitions to each (co-)inductive definition in $\mathfrak{D}$. These generalised Gupta translations yield two results (Theorem 15 and Corollary 17), three remarks (Remarks 18, 19 and 21), an open problem (Problem 20), and a conjecture (Conjecture 22).

Theorem 15 *Suppose that* S *is a plausible revision theory. Then the complexity of* S *is at least the complexity of* $S_i(S_{ci})$. *Furthermore, the complexity of the* S *theory of finitely many definitions is at least the complexity of the* $S_i(S_{ci})$ *theory of finitely many definitions.*

Lemma 16 *The complexity of the* S_i *theory of finitely many definitions is at least* Π^1_2.

Proof: See Section 9.

Corollary 17 *Suppose that* S *is a plausible revision theory. Then the complexity of the* S *theory of finitely many definitions (and hence the complexity of* S*) is at least* Π^1_2.

Remark 18 Aldo Antonelli and Vann McGee have independently sketched (very similar) proofs that Π^1_2 is an upper as well as a lower bound for the complexity of $S^\#$ and S^*. The idea is to produce, in the language of second-order arithmetic, a Π^1_2 formula with two free variables $\mathfrak{D}$ and B, which says that the formula B is a consequence of the set $\mathfrak{D}$ of definitions. A Löwenheim-Skolem argument like that for [2], Theorem 5C.15, is needed to show that the only domain we need is the domain of natural numbers. Antonelli is currently working on the details of this. Such a result is probably not generalisable to all plausible revision theories.

Remark 19 [2] explicitly compares $S^\#$ and S^* to S_i, and notes a number of S_i validities which are not $S^\#$ or S^* validities. The conclusion is that "$S^\#$ and S^* are *not*... generalizations of the system of inductive definitions" (p. 193). In light of the Gupta translations, this conclusion must be rethought. For anything we can define inductively (or co-inductively, for that matter) we can strongly define in $S^\#$ and in S^* (and in any other plausible revision theory).

Problem 20 Is every set which is definable with a *single* inductive definition also definable with a *single* revision theoretic definition (rather than a pair of them)?

Remark 21 The Gupta translations are completely general: they are well-defined whether or not the original set of definitions is positive. Also, there has been independent research on non-monotone inductive definitions. This suggests the following.

Conjecture 22 There is an interesting relationship between the behaviour of a non-monotone definition, understood inductively, and the behaviour of its Gupta translate (in $\mathbf{S}^{\#}$ and in $\mathbf{S}^{*}$).

We also note another way, suggested by revision theoretic semantics, of extending the theory of inductive definitions to non-positive sets of inductive definitions. (Similar remarks apply to co-inductive definitions.) For positive sets, $\mathfrak{D}$, of definitions, $\mathbf{S}_{\mathrm{i}}$ can be thought of as a plausible revision theory, based on the following strategy for liberalising the idea guiding $\mathbf{S}^{*}$ (see Section 6): we consider only the $\varnothing$-beginning revision sequences, i.e., those which begin by assigning the extension $\varnothing$ to each definiendum. Thus restricting our attention to a subfamily of revision sequences does not require the set of definitions to be positive. And so, given any model M and *any* set $\mathfrak{D}$ of definitions, we extend $\mathbf{S}_{\mathrm{i}}$ as follows: a sentence A is *valid in* $\mathbf{S}_{\mathrm{i}}$ *on* $\mathfrak{D}$ *in* M, iff A is stably true in every $\varnothing$-beginning revision sequence for $\delta_{\mathfrak{D},\mathrm{M}}$.

Conjecture 23 There are interesting points of contact between $\mathbf{S}_{\mathrm{i}}$ (so extended) and the theory of non-monotone inductive definitions.

9 Proving Lemma 16 (A sketch of this proof was provided by Yiannis Moschovakis in e-mail correspondence.) For the purposes of this section, L is the first-order language of arithmetic (as in Section 2). If $\mathrm{n} \in \omega$, $\underline{\mathrm{n}}$ is the term of L got by appending n copies of $'$ to the constant 0. We consider three new unary predicates, Y, H, and Q. Given a subset S of $\{Y, H, Q\}$, $L \cup \mathrm{S}$ is the language got by enriching L with the predicates in S. We let M range over models of L, and N is the standard model of arithmetic (as in Section 4). Given an L-model $\mathrm{M} = \langle \mathrm{D}, \mathrm{I} \rangle$, Y, H, and Q range over subsets of D. We generally assume that Y, H, and Q are interpretations of Y, H, and Q. Given an L-model $\mathrm{M} = \langle \mathrm{D}, \mathrm{I} \rangle$, $\mathrm{M} + \mathrm{Y} = \langle \mathrm{D}, \mathrm{I}' \rangle$ is the model of $L \cup \{Y\}$ which interprets the constants of L as does M, and which assigns Y to Y. (Similarly for M + H, M + Y + H, M + Y + H + Q, etc.) Finally, when we give our set $\mathfrak{D}$ of inductive definitions, these definitions will be given for models M + Y of the language $L \cup \{Y\}$.

We prove Lemma 16 by showing that, for every Π^1_2 set $\pi \subseteq \omega$, there is a finite positive set $\mathfrak{D}$ of definitions (over the language $L \cup \{Y\}$ and with definienda H and Q) such that π is recursively embeddable in $\{A : A$ is valid on $\mathfrak{D}$ in $\mathbf{S}_{\mathrm{i}}\}$. So suppose that $\pi \subseteq \omega$ is Π^1_2. Then, for some second-order Π^1_1 relation $\mathrm{P} \subseteq \omega \times \mathcal{P}(\omega)$ (where $\mathcal{P}(\omega)$ is the power set of ω), $\pi = \{\mathrm{n} \in \omega : (\forall \mathrm{Y} \subseteq \omega)(\langle \mathrm{n}, \mathrm{Y} \rangle \notin \mathrm{P})\}$. We now state Moschovakis's "Abstract Kleene Theorem", regarding second-order Π^1_1 relations (see Moschovakis [3], Theorem 8A1, p. 132).

Theorem 24 (Moschovakis) *Every* Π^1_1 *second-order relation on a countable acceptable structure is inductive.*

Without defining the terms in this theorem, we note its upshot for P. There is a formula $A = A(x, Y, Q)$ of $L \cup \{Y, Q\}$ in which Q occurs positively and which is such that, for every $\mathrm{Y} \subseteq \omega$, the set $\mathrm{Q(Y)} = \{\mathrm{n} \in \omega : \langle \mathrm{n,Y} \rangle \in \mathrm{P}\}$ is inductively defined in the model N + Y by the definition:

$$Qx =_{\mathrm{Df}} A(x, Y, Q).$$

(See Moschovakis [3], Chapter 6, Section 6A.)

Let the set $\mathfrak{D}$ of definitions consist of:

$$Hx =_{\mathrm{Df}} (x = 0 \vee \exists y(Hy \,\&\, x = y'))$$

$$Qx =_{\mathrm{Df}} A(x, Y, Q).$$

For every $\mathrm{M} = \langle \mathrm{D,I} \rangle$ and for every $\mathrm{Y} \subseteq \mathrm{D}$, the least fixed point of $\delta_{\mathfrak{D},\mathrm{M+Y}}$ assigns to H the extension $\mathbb{N}_\mathrm{M}$ (Definition 1, Section 2). Also, for every $\mathrm{Y} \subseteq \omega$, the least fixed point of $\delta_{\mathfrak{D},\mathrm{N+Y}}$ assigns to H the extension ω and assigns to Q the extension Q(Y). And so we have the following:

Lemma 25 For every Y *and for every* $\mathrm{n} \in \omega$, $\mathrm{N} + \mathrm{Y} \models_\mathrm{i}^{\mathfrak{D}} {\sim}Q\underline{\mathrm{n}}$ *iff* $\langle \mathrm{n,Y} \rangle \notin \mathrm{P}$.

The core of the proof of Lemma 16 is in Lemma 26.

Lemma 26 *For every sentence B of* $L \cup \{Y, H, Q\}$,

$$\models_\mathrm{i}^{\mathfrak{D}} ((AX \,\&\, PA^- \,\&\, \forall xHx) \rightarrow B) \textit{ iff } (\forall \mathrm{Y} \subseteq \omega)(\mathrm{N} + \mathrm{Y} \models_\mathrm{i}^{\mathfrak{D}} B).$$

Proof: ($\Rightarrow$) Assume $(\exists \mathrm{Y} \subseteq \omega)(\mathrm{N} + \mathrm{Y} \not\models_\mathrm{i}^{\mathfrak{D}} B)$. Choose such a Y. Then $\mathrm{N + Y + H + Q} \not\models B$, where H and Q are the extensions assigned to H and Q by the least fixed point of $\delta_{\mathfrak{D},\mathrm{N+Y}}$. Now, $\mathrm{N} \models (AX \,\&\, PA^-)$ and $\mathrm{N + H} \models \forall xHx$, so $\mathrm{N + Y + H + Q} \models (AX \,\&\, PA^- \,\&\, \forall xHx)$. So $\mathrm{N + Y + H + Q} \not\models ((AX \,\&\, PA^- \,\&\, \forall xHx) \rightarrow B)$. So, $\mathrm{N + Y} \not\models_\mathrm{i}^{\mathfrak{D}} ((AX \,\&\, PA^- \,\&\, \forall xHx) \rightarrow B)$. So $\not\models_\mathrm{i}^{\mathfrak{D}} ((AX \,\&\, PA^- \,\&\, \forall xHx) \rightarrow B)$, as desired.

($\Leftarrow$) Assume $\not\models_\mathrm{i}^{\mathfrak{D}} ((AX \,\&\, PA^- \,\&\, \forall xHx) \rightarrow B)$. So we can fix some $\mathrm{M} = \langle \mathrm{D,I} \rangle$ and some $\mathrm{Y} \subseteq \mathrm{D}$ such that $\mathrm{M + Y} \not\models_\mathrm{i}^{\mathfrak{D}} ((AX \,\&\, PA^- \,\&\, \forall xHx) \rightarrow B)$. So $\mathrm{M + Y + H + Q} \not\models ((AX \,\&\, PA^- \,\&\, \forall xHx) \rightarrow B)$, where H and Q are the extensions assigned to H and Q by the least fixed point of $\delta_{\mathfrak{D},\mathrm{M+Y}}$. So $\mathrm{M + Y + H + Q} \models (AX \,\&\, PA^- \,\&\, \forall xHx)$ and $\mathrm{M + Y + H + Q} \not\models \mathrm{B}$. Since $\mathrm{H} = \mathbb{N}_\mathrm{M}$ and since $\mathrm{M + Y + H + Q} \models \forall xHx$, $\mathrm{D} = \mathbb{N}_\mathrm{M}$. So M is isomorphic to N (by Lemma 7, Section 5). Let $\phi : \mathrm{D} \rightarrow \omega$ be this isomorphism, and for $\mathrm{X} \subseteq \mathrm{D}$, define $\phi(\mathrm{X}) = \{\phi(\mathrm{d}) : \mathrm{d} \in \mathrm{D}\} \subseteq \omega$. Then $\phi(\mathrm{H})$ and $\phi(\mathrm{Q})$ are the extensions assigned to H and Q by the least fixed point of $\delta_{\mathfrak{D},\mathrm{N}+\phi(\mathrm{Y})}$. Furthermore, $\mathrm{N} + \phi(\mathrm{Y}) + \phi(\mathrm{H}) + \phi(\mathrm{Q})$ is isomorphic to $\mathrm{M + Y + H + Q}$. So $\mathrm{N} + \phi(\mathrm{Y}) + \phi(\mathrm{H}) + \phi(\mathrm{Q}) \not\models B$. So $\mathrm{N} + \phi(\mathrm{Y}) \not\models_\mathrm{i}^{\mathfrak{D}} B$.

Corollary 27 *For every* $\mathrm{n} \in \omega$, $\mathrm{n} \in \pi$ *iff* $\models_\mathrm{i}^{\mathfrak{D}} ((AX \,\&\, PA^- \,\&\, \forall xHx) \rightarrow {\sim}Q\underline{\mathrm{n}})$.

As promised, we have shown that, for every Π_2^1 set $\pi \subseteq \omega$, there is a finite positive set $\mathfrak{D}$ of definitions (over the language $L \cup \{Y\}$) such that π is recursively embeddable in $\{A : A$ is valid on $\mathfrak{D}$ in $\mathbf{S}_\mathrm{i}\}$. This suffices for Lemma 16.

10 Theories of Truth [2] bases its theories of truth, $\mathbf{T}^{\#}$ and $\mathbf{T}^*$, on $\mathbf{S}^{\#}$ and $\mathbf{S}^*$. In the theories of truth there is a single circular concept, *truth*, and it is defined via an infinite set of *partial* definitions (p. 197). Furthermore, suppose that L is a first-order language which has a "quotation name" (p. 75) 'A' for each sentence A of L^+ (which results by adding the truth predicate $\boldsymbol{T}$ to L). Then the set of partial definitions is completely determined: it is the set of definitions of the form

$$\boldsymbol{T}\text{'}A\text{'} =_{\mathrm{Df}} A,$$

where A is a sentence of L^+.

Given the special features of $\mathbf{T}^\#$ and $\mathbf{T}^*$, our work on $\mathbf{S}^\#$ and $\mathbf{S}^*$ does not deliver any verdict regarding the complexity of $\mathbf{T}^\#$ and $\mathbf{T}^*$. We leave this as an open problem which we now make precise (Problem 32).

Definition 28 If L is a first-order language then L^+ is the result of adding the truth predicate $\boldsymbol{T}$ to L.

Definition 29 Suppose (1) L is a first-order language with a quotation name for each sentence of L^+; (2) $\mathrm{M} = \langle \mathrm{D,I} \rangle$ is a model for L; (3) D contains each sentence of L^+; and (4) for each sentence A of L^+, $\mathrm{I}(\text{'}A\text{'}) = A$. Then M is a *ground model* for L. (This diverges slightly from [2]'s definition.)

Definition 30 (See [2], p. 210.) Given a ground model, M (for a language L),

$$\mathbf{V}^\#_{\mathrm{M}} = \{A\colon A \text{ is valid in M by } \mathbf{T}^\#\} \text{ and } \mathbf{V}^*_{\mathrm{M}} = \{A\colon A \text{ is valid in M by } \mathbf{T}^*\}.$$

Definition 31 Given a first-order language L (with quotation names for the sentences in L^+),

$$\mathbf{V}^\#_L = \bigcap\{\mathbf{V}^\#_{\mathrm{M}}\colon \mathrm{M} \text{ is a ground model for } L\} \text{ and}$$

$$\mathbf{V}^*_L = \bigcap\{\mathbf{V}^*_{\mathrm{M}}\colon \mathrm{M} \text{ is a ground model for } L\}.$$

Problem 32 What is the complexity of $\mathbf{V}^\#_L$ and of $\mathbf{V}^*_L$?

Remark 33 McGee has pointed out (in correspondence) that Burgess [1] has made some progress toward solving this problem. In particular, let L be the language of arithmetic (as in Section 2) and let N be the standard model of arithmetic (as in Section 6). Furthermore, identify each formula A of L^+ with it Gödel number, $\mathrm{Gn}(A)$, so that $\mathrm{I}(\text{'}A\text{'}) = \mathrm{Gn}(A)$. Burgess [1] shows that $\mathbf{V}^*_{\mathrm{N}}$ (which is called "$\square$T") is complete Π^1_2. (Theorem 12.3, p. 676; Burgess [1]'s statement of this theorem contains a typo.) McGee has noted that the same holds for $\mathbf{V}^\#_{\mathrm{N}}$.

Acknowledgements I thank Nuel Belnap for introducing me to this topic; Aldo Antonelli for helpful discussions, particularly regarding the proof in Section 9; Vann McGee, for carefully reading the paper, for helpful comments, and for directing me to Burgess [1]; Yiannis Moschovakis for helpful e-mail correspondence and for sketching the proof in Section 9; and especially Anil Gupta, who carefully read an early draft of this paper and made many central observations and suggestions.

REFERENCES

[1] Burgess, J., "The truth is never simple," *Journal of Symbolic Logic*, vol. 51 (1986), pp. 663–681.

[2] Gupta, A. and N. Belnap, *The Revision Theory of Truth*, MIT Press, 1993.

[3] Moschovakis, Y., *Elementary Induction on Abstract Structures*, North-Holland, 1974.

Dept. of Philosophy
University of Toronto
Toronto, Ontario M5S 1A1
Canada
e-mail: pkremer@epas.utoronto.ca

Dept. of Philosophy
University of Pittsburgh
Pittsburgh, Pennsylvania 15217