Revista Colombiana de Matemáticas
Vol. XIX (1985), pág. 177 - 200

THE CONSISTENCY OF A VARIANT OF CHURCH'S THESIS WITH AN AXIOMATIC THEORY OF AN EPISTEMIC NOTION

W.N. Reinhardt

ABSTRACT. In this paper we prove the consistency of a variant of Church's Thesis than can be formulated as a schema in a first order language with a modal operator for intuitive provability. We also conjeture the consistency of a stronger variant.

1. INTRODUCTION.

We consider the language of arithmetic augmented by a new symbol B and the formation rule: If σ is a sentence (or formula) so is $B\sigma$. The informal meaning intended for $B\sigma$ is that σ is intuitively provable, so that for example $\neg B\sigma \wedge \neg B\neg\sigma$ expresses the (absolute) undecidability of σ. This interpretation suggest notions of intuitive decidability, for example

$$\forall x(B\theta(x) \vee B\neg\theta(x))$$

express the intuitive decidability of $\theta(x)$, and these motivate the formulation of our variant of Church's Thesis. Since Turing advocated the view that any intelectual activity of humanscan be carried out by a properly programmed computer, and in particular that theorem proving by an idealized human mathematician is essentially mechanical, the thesis we formulate migth appropriately be called Turing's thesis. I believe that B expresses an important epistemic notion and that the axiomatic theory given here can be used to illuminate for example some controversies regarding the philosophical significance of Gödel's incompleteness theorems. In this paper we leave these issues aside, and simply formulate the theory and prove it consistent with one variant of Church's Thesis. In a later paper we shall discuss these issues and the relation of B to earlier authors. (Gödel 1933, 1951, Löb 1955, Kalmar 1959, Myhill 1960, Lucas 1961, Montague 1963, Benacerraf 1967, Tharp 1973, Wang 1974, Boolos 1979, Shapiro 1980). I would like to thank Andrej Ščedrov for pointing out an error in the first version of this paper. In the earlier version a proof was claimed for conjecture-3-1a. of this paper. The problem remains open.

2. ARITHMETIC WITH B.

We split the axioms into four groups: those which may be regarded as applicable to any subject matter(the logical axioms), those pecuiliar to arithmetic, those involving the truth (or satisfaction) predicate for arithmetic, and those stating Church's Thesis. In the first and third groups some of the axioms are essentially classical (e.g. instances of classical schemas which may however involve B) and others are new, peculiar to languages with B. The arithmetic axioms are essentially classical.

2.1. Logical axioms.

We suppose our languages has variables $v_0, v_1, \ldots$, a one place sentential connective $\neg$ (for negation), a two place sentential connective $\rightarrow$ (for truth functional implication), the universal quantifier $\forall$, a one place sentential connective B (for provability), and equality $\doteq$. We allow relation symbols and certain function symbols, but logic with function symbols in the general case requires restrictions not familiar from classical logic. (In effect, we may allow function symbols for recursive functions with no restrictions, or arbitrary function symbols with certain caveats which will be mentioned). We shall use $\neg$, $\rightarrow$, $\forall$ not as names for symbols but as names for operations. Thus if θ, ϕ are formulas, $\neg\theta$, $(\theta \rightarrow \phi)$ are to be formulas. We treat defined connectives such as $\vee$, $\wedge$, $\leftrightarrow$ similarly.

We have the usual formation rules for first order languages, plus the rule: if θ is a formula, $B\theta$ is a formula with the same free variables as θ. A sentence is a formula with no free variables.

In order to state the axioms for the truth predicate (in §2.3), it will be convenient to suppose that all syntactic objects have been identified with their Gödel numbers, in one of the usual ways. Thus the syntactic operations $\neg$, $\rightarrow$, etc. are all primitive recursive. It will not much matter how this is done, but for convenience in describing substitution operations, one may think of formulas as strings of symbols. What is important is that the various syntactic operations $\neg$, $\rightarrow$, substitution, etc. will be primitive recursive.

DEFINITION 2.1.

a) By a B-*closure* of a formula θ we mean a sentence obtained from θ by iterated applications of universal quantification and B. Thus $\forall x \forall y\, x \doteq y$, $B\forall x B \forall y B\, x \doteq y$, and $B\forall x \forall y\, x \doteq y$ are all B-closures of $x \doteq y$. If σ is a sentence it is a B-closure of itself.

b) $\theta(x/y)$ is the expression obtained from θ by replacing all free occurrences of x in θ by y.

c) By the *logical axioms* we understand the B-closures of the following schemas (where u,x,y,z are variables):

L1. truth functional tautologies

L2. $\forall x(\theta \to \phi) \to (\forall x\theta \to \forall x\phi)$

L3. $\forall y(\forall x\theta \to \theta(x/y))$, where x,y are variables and x is free for y in θ,

L4. $\theta \to \forall x\theta$, where x is not free in θ,

L5. $x \doteq x$

L6. $x \doteq y \to [\theta(u/x) \to \theta(u/y)]$, where u is free for x,y in θ,

L7. $B(\theta \to \phi) \to (B\theta \to B\phi)$

L8. $B\theta \to \theta$

L9. $B\theta \to BB\theta$

L10. $B\forall x\theta \to \forall xB(\exists y(x = y) \to \theta)$, where y is a variable distinct from x, and means as usual $\neg\forall\neg$,

L11. $B\exists zB(t = z) \to [\forall x\theta \to \theta(x/t)]$, where x,z are variables, t is a term and

i) z does not occur free in t

ii) x is free for t in θ,

L12. $\exists y(t = y) \to [\forall x\theta \to \theta(x/t)]$, provided that i), ii) above hold, and in addition x does not occur free within the scope fo B in θ.

d) A *theory* in the language with B is a set of sentences containing the logical axioms and closed under modus ponens. We write $A \vdash \sigma$ to mean σ is in every theory including A. We write as usual $\vdash\sigma$ for $\emptyset \vdash \sigma$.

We note:

PROPOSITION 2.2.

a) *Suppose that* A *is a set of sentences such that whenever* $\sigma \in A$, $B\sigma \in A$. *Then* $A \vdash \sigma$ *implies* $A \vdash B\sigma$.

b) *If the sentence* σ *is a classical validity, in a language with no function symbols, then* $\vdash\sigma$.

c) *If the sentence* σ *is a classical validity in a language with function symbols* f_i, *and* A *is the set of sentences* $\forall x\exists y\ f(\vec{x}) = \vec{y}$, *then* $A \vdash \sigma$.

Proof. a) The only rule is modus ponens, so apply L7, L2.

b) Since L1-L6 are the usual classical schemas, this is obvious.

c) By L12, it is sufficient to see that $A \vdash \exists y(t = y)$ for all terms t built from the f's. This is easily seen by induction on t; e.g. if $t = f(t_1,t_2)$, and $A \vdash \exists y_1(t_1 = y_1) \wedge \exists y_2(t_2 = y_2) \wedge \forall y_1y_2\exists z(f(y_1,y_2) = z)$, then by L12, $A \vdash \exists z(f(t_1,t_2) = z)$.

We note also the following:

L3' $\forall x\theta \to \theta(x/y)$,

where x is free for t in θ, provided that we restrict the introduction of function symbols to (for example) primitive recursive functions. This is because the antecedent of L11 says that t is a term which may be effectively evaluated (this is the caveat referred to earlier) and it will follow from our arithmetic axioms that this is so for primitive recursive terms. The antecedent in L12 may of course be omitted altogether if we follow the usual practice of classical logic of assuming all functions everywhere defined.

We remark without proof that the axioms L11-L12 are chosen so that the following is true.

THEOREM 2.3. *If* T *is a theory with* $T \vdash B\forall\vec{x}\exists!y$, *no* x_i *occurs in the scope of* B *in* θ, f *is new function symbol, and* T' *has as axioms* T *together with all* T *closures of*

$$\theta(\vec{x},y) \leftrightarrow f(\vec{x}) = y,$$

Then T' *is a conservative extension of* T.

In particular, if T has no function symbols, then we may add them with impunity, but only by observing the caveats regarding the principle of universal instantiation with complex terms (the case of functions fx = y corresponding to $\theta(x,y)$ with x in the scope of B would require a further restriction in L12).

2.2. Arithmetic axioms.

By the language of (first order) Peano arithmetic we understand the language with an individual constant for 0, a unary function symbol for the succesor operation S, and a function symbol d for each primitive recursive definition of a function f_d (we could of course do with only function symbols for plus and times, but it will be convenient to have terms for certain primitive recursive functions). Just as we used $\forall$, $\neg$, $\doteq$ etc. for syntactic operations on formulas and variables, we use $\bar{0}$, $\bar{S}$, $\bar{f}_d$ for syntactic operations on terms. Thus $\bar{0}$ is a term, and if t is a term, so are $\bar{S}(t)$, $\bar{f}_d(t)$. If f is primitive recursive, we shall often write $\bar{f}$, leaving the reader to find d.

REMARK 2.4. It does not much matter how one thinks of the definitions d, except that "d is such a definition" should be primitive recursive, and "x is the denotation of t" should be definable for terms t built up with the $\bar{f}_d$'s. We observe however that a nice way is the following simultaneous definition of function symbols and terms. We write V(e) for the free variable of the expression e.

1. $\bar{0}$ is a term with $V(p) = \emptyset$.

2. If v is a variable, v is a term with $V(v) = \{v\}$.
3. $\bar{S}$ is a function symbol with $V(\bar{S}) = \emptyset$.
4. If f is a function symbol and t is a term then f(t) is a term with $V(f(t)) = V(f) \cup V(t)$.
5. If t,s are terms and v is a variable, $v \notin V(t)$, then Pvst is a function symbol with $V(Pvst) = V(t) \cup V(s) - \{v\}$.

Modulo an assignment of functions f_d to the function symbols d, an assignment a to the variables now determines (in the usual way) a value $t[a]$ for each term. We explain the assignment f_d as follows. If we single out a variable, say u, to stand as argument, there is a (one place) function f_t for each term, namely $f_t(n)$ is just the value of t under the assignment which is like a except for assigning to u the value n. We intend that for a function symbol d, and term du (obtained by concatenating d with u), f_d will be the same as f_{du} (with u singled out). In particular, for the functional symbol d = Pvst, the value at n of the one-place function f_d is obtained by interating f_s (determined by selecting v as the argument) n times, starting with f_t (which, since $v \notin V(t)$, is constant when v is selected to mark the argument):

$$f_d(n) = f_s f_s \ldots f_s f_t = f_s^{(n)} f_t \quad ,$$

i.e.

$$f_d(0) = f_t$$
$$f_d(n+1) = f_s f_d(n).$$

Thus the axioms for d are (the closures of)

$$d(0) \doteq t$$
$$d(\bar{S}v) \doteq s(v/dv).$$

Note that the syntactic operation $\bar{+}$ takes variables x,y and produces the term $(x \mathbin{\bar{+}} y)$ with free variables x,y, whereas the function symbol for "adding x to" or "iterate the succesor operation starting with x" is $Py\bar{S}yx$, with one free variable x. Thus $x \mathbin{\bar{+}} y$ is $Py\bar{S}yxy$. Similarly, using Polish notation and dropping the bars for legibility, $\cdot xy = +x +x \ldots +x0 = (+x)^{(y)}0 = (Py + xy0)y$.

DEFINITION 2.5. Let L be any language which includes the language of Peano arithmetic. The *Peano axioms for* L are the B-closures (or the ordinary closures, if L is classical) of the following:

A1. The usual Peano axioms for $\bar{0}$ and $\bar{S}$, which assert that S is 1-1 and onto all but 0. We takes this to include $\exists x(\bar{0} \doteq x)$ and $\exists y(\bar{S}x \doteq y)$.

A2. The usual Peano axioms stating that +, $\cdot$ satisfy their recursive definitions, and in general similar axioms for each primitive recursive function symbol $\bar{f}$, including $\forall\vec{x}\exists y(f(\vec{x}) = y)$.

A3. The usual induction schema, allowing formulas from L:

$$\theta(x/\bar{0}) \wedge \forall x(\theta \to \theta(x/\bar{S}x)) \to \forall x\theta.$$

THEOREM 2.6.

a) $B\forall x B\exists y(x = y)$. *This yields the* B-*closures of* $B\forall x\theta \to \forall xB\theta$ *and* $\exists xB\theta \to B\exists x\theta$.

b) *From the* B-*closure of* $\theta \to B\theta$, *we obtain those of* $\exists x\theta \to B\exists x\theta$ *and* $\exists x\theta \to \exists xB\theta$.

c) *For each (primitive recursive) term* t *and variable* z, *we have the* B-*closure of* $t = z \to B(t = z)$. *In particular*

(i) $Sx = z \to B(Sx = z)$

(ii) $x+y = z \to B(x+y = z)$

(iii) $x \cdot y = z \to B(x \cdot y = z)$

(iv) $\exists zB(t = z)$

(v) $x < y \to B(x < y)$

(vi) $t \neq z \to B(t \neq z)$

(vii) $t_1 = t_2 \to B(t_1 = t_2)$.

d) *We have the* B-*closures of* $\forall x < y(B\theta) \to B(\forall x < y\,\theta)$.

REMARK. We have suppressed the bars and dots in the interests of readability.

Proof. a) We prove the B-clousures of $\exists y(x = y)$ by induction on x. $\exists y(0 = y)$ is a classical validity, so $B\exists y(0 = y)$ by Prop. 2.2. For the induction note that $\forall x\exists y(Sx = y)$ is a classical validity, so again by Prop. 2.2, $B\forall x\exists y(Sx = y)$. By L10 then

$$\forall xB(\exists y(x = y) \to \exists y(Sx = y)),$$

and by L7, L2

$$\forall x(B\exists y(x = y) \to B\exists y(Sx = y)),$$

which is the induction step.

We note that L10 now simplifies to $B\forall x\theta \to \forall xB\theta$,

a) con't. Furthermore by Prop. 2.2, $B\forall x(\theta \to \exists x\theta)$, so by L10 $\forall xB(\theta \to \exists x\theta)$, so by L7 $\forall x(B\theta \to B\exists x\theta)$, whence $\exists xB\theta \to B\exists x\theta$.

b) From $\forall x(\theta \to B\theta)$ we get $\forall x(\sim B\theta \to \sim\theta)$, $\forall x \sim B\theta \to \forall x \sim \theta$, $\sim\forall x\sim\theta \to \sim\forall x\sim B\theta$, i. e. $\exists x\theta \to \exists xB\theta$. Also, we have $\exists xB\theta \to B\exists x\theta$, so $\exists x\theta \to B\exists x\theta$.

c) We show that $\forall x(Sx = z \to B(Sx = z))$ by induction on z. The case $z = 0$ is vacuously true, as $\forall x \neg(Sx = 0)$. For the induction step we need the

LEMMA 2.7. *We have the* B-*closure of* $x = y \to B(x = y)$.

Proof. We give two proofs, first using the equality axioms, then using only induction. Now L6 gives

$$x = y \to [B(x = u)(u/x) \to B(x = y)(u/y)]$$

i.e.

$$x = y \to [B(x = x) \to B(x = y)].$$

But L5 gives $B(x = x)$, so a tautology yields

$$x = y \to B(x = y).$$

We now prove $\forall y(x = y \to B(x = y))$ by induction on x. First, $\forall y(0 = y \to B(0 = y))$, this is done by induction on y. If $y = 0$, it is $B(0 = 0)$. Since $0 = 0$ is clas-sicaly valid, Prop. 2.2 gives $B(0 = 0)$. The conclusion of the induction step is $\forall y(0 = Sy \to B(0 = Sy))$, which is vacuolusly satisfied. We now return to the induction on x. We assume $\forall y(x = y \to B(x = y))$, and are to prove $\forall y(Sx = y \to B(Sx = y))$. Again we proceed by induction on y. The case $y = 0$ is vacuously satisfied. So we must see $\forall y(Sx = y \to B(Sx = y)) \to \forall y(Sx = Sy \to B(Sx = Sy))$. Now $\forall y \forall x(x = y \to Sx = Sy)$ is a classical theorem of Peano arithemtic, so by Prop. 2.2 and A10, $\forall y \forall x B(x = y \to Sx = Sy)$, and thus

$$B(x = y) \to B(Sx = Sy).$$

Consequently

$$\begin{aligned} Sx = Sy &\to x = y \\ &\to B(x = y) \quad \text{induction on } x \\ &\to B(Sx = Sy). \end{aligned}$$

This completes the induction on y, hence that on x, and so the lemma.

We return to the induction on z for the theorem. We show directly that the conclusion of the induction step holds, namely

$$\forall x(Sx = Sz \to B(Sx = Sz)).$$

This is because $Sx = Sz \to x = z$

$\to B(x = z)$ by the Lemma.

As in the second proof of the Lemma,

$$B\forall x(x = y \to Sx = Sy)$$

$$\forall x B(x = y \to Sx = Sy)$$

$$\forall x(B(x = y) \to B(Sx = Sy)),$$

so

$$Sx = Sz \to B(Sx = Sz),$$

as desired.

For the iterations of S, note first that the above argument shows $0 = z \to B(0 = z)$. Now suppose that we have the B-closures of

$$f(x,0) = k(x)$$

$$f(x,n+1) = g(x,fn),$$

and those of $kx = z \to B(kx = z)$, $g(x,y) = z \to B(g(x,y) = z)$.

We show,

$$f(x,n) = z \to B(f(x,n) = z)$$

by induction on n. For n = 0, this is

$$\begin{aligned} f(x,0) = z &\to k(x) = z \\ &\to B(k(x) = z) \\ &\to B(f(x,0) = z). \end{aligned}$$

For Sn it is

$$\begin{aligned} f(x,Sn) = z &\to g(x,f(x,n)) = z \\ &\to \exists u(f(x,n) = u \wedge g(x,u) = z) \\ &\to B\exists u(f(x,n) = u \wedge g(x,u) = z) \quad \text{(by b)} \\ &\to B(f(x,Sn) = z). \end{aligned}$$

This takes care of c) and in particular (i)-(iii); (iv) follows from b). Writing $x < y$ as $\exists z(x+Sz = y)$, so does (v). We get (vi), (viii) the same way, viewing $t \neq z$ as $\exists u(t = u \wedge (t < u \text{ or } u < t))$, and $t_1 = t_2$ as $\exists z(t_1 = z = t_2)$.

d) This is a straightforward induction on y.

2.3. Axioms for the truth predicate.

The axioms are the usual ones for arithmetical truth, with the addition of a clause for sentences Bσ. Satisfaction is definable from truth, since for example if θ has one free variable x, then $\models \theta[n]$ iff $\models \theta(x/\bar{n})$, where $\bar{n}$ denotes n. We must, however, state more than the truth schema $Tr(\bar{\sigma}) \leftrightarrow \sigma$ to get the corresponding satisfaction schemas such as

$$Sat(\bar{\theta},y) \leftrightarrow \theta(y).$$

We shall need several syntactic operations, and formulas and terms arithmetizing syntactic notions, which we summarize in

NOTATION 1.8.

a) For each $n \in w$, $\bar{n}$ is the term $\bar{S}\ldots\bar{S}\bar{0}$ (n iterations).
The corresponding function is $b: n \mapsto \bar{n}$.

b) $Sb(\theta,x,n)$ is $\theta(x/\bar{n})$, the result of substituting $\bar{n}$ at the free occurrences of x in θ.

c) $Vb(x)$, $Tm(x)$, $Fm_L(x,y)$, $Sent_L(x)$, are formulas (of Peano arithmetic) expresing "x is a variable", "x is a primitive recursive term", "x is a formula of L with one free variable y", and "x is a sentence of L". L may be classical or allow B, but does not admit Tr_L; Tr_L is always a unary predicate not in L.

d) den(x,y) is a formula expressing "x is a primitive recursive term and y is the denotation of x".

e) Recall that if f is a primitive recursive function of n arguments, then $\bar{f}$ is a syntactic operation taking n terms to a term. If for example f has 2 arguments and u, v are variables $\bar{f}(u,v)$ will have two free variables u, v, and for all m, n, $\bar{f}(\bar{m},\bar{n})$ will denote f(m,n). For the axioms we are interested specifically in $\bar{\neg}$, $\bar{\rightarrow}$, $\bar{\forall}$, $\bar{=}$, $\bar{B}$, and $\overline{Sb}$. Note that for example $(\bar{\theta} \bar{\rightarrow} \bar{\phi})$ denotes $\theta \rightarrow \phi$, $\bar{B}\bar{\theta}$ denotes $B\theta$, and $\overline{Sb}(\bar{\theta},\bar{x},\bar{n})$ denotes $Sb(\theta,x,n)$.

DEFINITION 2.9. The *satisfaction axioms for* L are the B-closures of the following. Here u,v,w,x,y,z are variables.

S1. $Tm(u) \wedge Tm(v) \rightarrow$

$$Tr(u \bar{=} v) \leftrightarrow \exists z(den(u,z) \wedge den(v,z))$$

S2. $Sent(x) \rightarrow$

$$Tr(\bar{\neg}\, x) \leftrightarrow \neg\, Tr(x)$$

S3. $Sent(x) \wedge Sent_L(y) \rightarrow$

$$Tr(x \bar{\rightarrow} y) \leftrightarrow (Tr(x) \rightarrow Tr(y))$$

S4. $Fm(x,w) \rightarrow$

$$Tr(\bar{\forall} wx) \leftrightarrow \forall z Tr(\overline{Sb}(x,w,z))$$

S5. $Sent(x) \rightarrow$

$$Tr(\bar{B}x) \leftrightarrow BTr(x)$$

Note that for appropriate L, S1-S5 are the usual satisfaction axioms over Peano arithmetic; S5 is the obvious addition when the language includes B. For readability we omitted the subscript L from Sent, Fm, and Tr.

DEFINITION 2.10.

a) **P** is classical first order Peano arithmetic, i.e. the classical theory of axioms A1-A3. **BP** is Peano arithmetic in the language with B adjoined, i.e. the (Def. 2-1d) theory of A1-A3.

b) If A is a theory in a language L with finitely many function and relation symbols, then A^+ has in addition to the axioms of A the (B)-closures of the satisfaction axioms for Tr_L.

2.4. Church's Thesis.

Let U(e,n) be an r.e. formula universal for r.e. sets (provably in P). Now we may state the version of Church's Thesis which concern us. They are

DEFINITION 2.11.

CT. $\forall n(\theta n \rightarrow B\theta n) \rightarrow \exists n \forall n(\theta n \leftrightarrow Uen)$

BCT. $B[\forall n(\theta n \rightarrow B\theta n) \rightarrow \exists e \forall n(\theta n \leftrightarrow Uen)]$.

REMARK 2.12.

a) Note that CT implies that every intuitively decidable set (i.e. $\forall x[B\theta \vee B\neg\theta]$) is recursive, as then both $\forall x(\theta \rightarrow B\theta)$ and $\forall x(\neg\theta \rightarrow B\neg\theta)$, so both θ and $\neg\theta$ define r.e. sets, hence θ is recursive.

b) We remark without proof that the effectivized version of CT

ECT. $B\forall n(\theta n \to B\theta n) \to \exists eB\forall n(\theta n \leftrightarrow Uen)$

is refutable (this is essentially the content of Godel's first incompleteness theorem).

3. ON THE CONSISTENCY OF BCT.

We can now state the main theorem.

THEOREM 3.1. *The theory* $(BP)^+$ *is consistent with* CT. *That is, Peano arithmetic in the language with* B *and the truth (or satisfaction) predicate is consistent with the weak form of Church's Thesis. Specifically,* $(BP)^+$ *includes the axiom groups* L1-L12, A1-A4, *and* S1-S5.

The following corollary can be stated without the fuss of arithmetization required for 3·1.

COROLLARY 3.2. *The theory* BP *is consistent with* CT.

These results are much weaker than the corresponding conjectures for BCT:

Conjecture 3-1a. $(BP)^+$ *is consistent with* BCT.

Conjecture 3-2a. BP *is consistent with* BCT.

We begin by indicating the proof in outline. We shall first prove 3.2, then observe that the same method works to obtain 3.1. The proof proceeds by an interpretation **I**. **I** interprets $(BP)^+$ into the languages of P^+ so that an important instance of CT holds. Namely, that where θ is the formula Bw(x) expressing "x is an intuitively provable Tr-free sentence", i.e. the formula $Tr(\bar{B}(x))$. CT(Bw) is equivalent to the assertion that Bw is r.e.

Since Bw is r.e., all instances $CT(\theta)$ for θ not involving Tr follow from this. This proves 3.2. The proof of 3.1 proceeds the same way, beginning with an interpretation I of $(BP^+)^+$ into the language of P^{++} and using $CT(Bw^+)$ to complete the proof.

With each formula θ and assignment a to the free variables θ we can associate a sentence σ which asserts that θ holds of the assigned numbers. For example, to $\theta(u)$ (with one free variable u) and assignment a which sets u to n, we associate $\theta(\bar{n})$; to $\theta(u,v)$ and $a = (u/m, v/n)$ we associate $\theta(\bar{n},\bar{m})$, etc. In the sequel it will be convenient to associate with each θ a term $[\theta]$ with the same free variables as θ such that, under the assignment a, $[\theta]$ will denote the above mentioned sentence. For example, in the case $\theta(u)$, $\overline{Sb}(\bar{\theta},\bar{u},\bar{u})$

is such a term.

DEFINITION 3.3.

a) Let $u_1, \ldots, u_n$ be the free variables of θ. Then

$$[\theta] = \overline{Sb}(\ldots\overline{Sb}(\overline{Sb}(\bar{\theta},\bar{u}_1,u_1),\bar{u}_2,u_2)\ldots,\bar{u}_n,u_n)$$

(if $n = 0$, $[\theta] = \bar{\theta}$). Note that $\theta \mapsto [\theta]$ is primitive recursive.

b) $Bw(x)$ is the formula $Sent_L(x) \wedge Tr(\bar{B}(x))$ of $(BP)^+$ (which expresses "x is an intuitively provable sentence of L").

c) Let π be any formula of classical arithmetic (not involving Tr). We define the interpretation $\mathbf{I}$ ($\mathbf{I}_\pi$ if we need to make π explicit) in two stages. First, for formulas not involving Tr, then for those which do.

(i) $\alpha^{\mathbf{I}} = \alpha$ if α is atomic (without Tr).

$$(\neg\theta)^{\mathbf{I}} = \theta^{\mathbf{I}}$$

$$(\theta \to \phi)^{\mathbf{I}} = \theta^{\mathbf{I}} \to \phi^{\mathbf{I}}$$

$$(\exists u\theta)^{\mathbf{I}} = \exists u\theta^{\mathbf{I}}$$

$$(B\theta)^{\mathbf{I}} = \theta^{\mathbf{I}} \wedge \pi[\theta^{\mathbf{I}}]$$

(this interpretation of B was suggested by M.H. Löb). It will be convenient to introduce the notation

$$B_\pi\theta = \theta \wedge \pi[\theta].$$

(ii) Let $\mathbf{I}_0$ be the (primitive recursive) mapping (from formulas of arithmetic with B but without Tr to formulas of arithmetic with neither B nor Tr) defined by (i) (and the condition $\mathbf{I}_0(x) = 0$ if x is not such a formula). We put

$$(Tr(x))^{\mathbf{I}} = Tr(\bar{\mathbf{I}}_0(x)),$$

where $\bar{\mathbf{I}}_0(x)$ is a primitive recursive term representing $\mathbf{I}_0$.

PROPOSITION 3.4. CT(Bw), *the instance of* CT *obtained by taking* $\theta(x) = Bw(x)$, *is equivalent to*

$$\exists e\forall x(Bw(x) \leftrightarrow U(e,x)),$$

BCT(Bw) *to*

$$B\exists e\forall x(Bw(x) \leftrightarrow U(e,x)),$$

and ECT(Bw) *to*

$$\exists eB\forall x(Bw(x) \leftrightarrow U(e,x)).$$

Proof. It suffices to prove the antecedent

$$B\forall x(Bw(x) \to B\,Bw(x)).$$

Now $Bw(x)$ is $Tr(\bar{B}(x))$, and S5 is the B-closure of

$$Tr\bar{B}(x) \to B\,Tr(x).$$

Thus

$$\begin{aligned} Bw(x) &\to B\ Tr(x) && \text{by S5} \\ &\to B\ BTrx && \text{by L9} \\ &\to B\ Tr(\bar{B}(x)) && \text{by S5} \\ &\to B\ Bw(x). \end{aligned}$$

THEOREM 3.5. *There is a r.e. formula* π *such that under the interpretation* $\mathbf{I}_\pi$, $(BP)^+$ *is true, as well as* CT(Bw). *In particular*

(a) $\sigma \in BP$ *implies* $P \vdash \sigma^{\mathbf{I}}$

(b) $\sigma \in (BP)^+$ *implies* $P^+ \vdash \sigma^{\mathbf{I}}$

(c) $\sigma = CT(Bw)$ *implies* $P^{++} \vdash \sigma^{\mathbf{I}}$.

Proof. We break the proof into a series of lemmas and propositions. The first lemma asserts the existence of a π satisfyingthe hypotheses of the others.

LEMMA 3.6. *There is an r.e. formula* π *such that*

(i) $P^+ \vdash \sigma$ *implies* $P \vdash \pi[\sigma]$

(ii) $P \vdash \pi[\theta \to \phi] \to \pi[\theta] \to \pi[\phi]$

(iii) $P \vdash \pi[\theta] \to \pi[\pi[\theta]]$

(iv) $P \vdash \theta \to \pi[\theta]$, *for* θ *any primitive recursive equation.*

(v) $P \vdash Sent_L(y) \to \pi([Tr(y)] \leftrightarrow y)$

(vi) $P \vdash Sent_L(x) \wedge Sent_L(y) \wedge \pi(x \bar{\to} y) \wedge \pi(x) \to \pi(y)$

(vii) $P^+ \vdash Sent_L(x) \wedge \pi_0(x) \to Tr(x)$,

where $\pi_0(x)$ *expresses "x is provable in P".*

(viii) $P^{++} \vdash Sent_L(x) \wedge \pi(x) \to Tr(x)$

In (v)-(viii), L *is the language of* P.

Proof. Take π to be a standard formula expressing provability in P^+. Then (i)-(iii) are the schemas traditionaly used to prove Gödel's second incompleteness theorem. The proof may be found for example in the second volume of Hilbert and Bernays. (iv) is the main lemma used in the proof of (iii). (v) is the formalization of $Tr(\bar{\sigma}) \leftrightarrow \sigma$, and will be discussed at lemma 3.17. (vi) is the main lemma used in proving (ii); we list it separately for ease of reference. (vii) is a standard fact about P^+; it contains enough to prove the correctness of P. The specific theorem of P we will need to apply this to is

(vii') $\qquad Fm(x,w) \to \pi_0(\bar{I}\bar{S}\bar{b}(x,w,z) \bar{\leftrightarrow} \bar{S}b(\bar{I}x,w,z))$

which formalizes Claim 3.12.

PROPOSITION 3.7. *Let* A *be a theory such that*

(i) $A \vdash \sigma$ *implies* $A \vdash \pi[\sigma]$

Suppose furthermore that

(a) *The universal closures of the* **I** *images of the schema*

$$B\forall x\theta \to \forall xB\theta$$

are provable in A.

(b) *The universal closure of the* **I** *image of* ϕ *is provable in* A.

Then the **I** *image of any* B-*closure of* ϕ *is provable in* A.

(Thus it will suffice to show in P^+ the universal closures of the **I** images of the schemas we wish to interpret, including that of (a)).

Proof. Let σ be the universal closure of $\phi^{\mathbf{I}}$. From (i) we get A $\vdash\sigma$ implies A $\vdash$B . Using (a), the prefixed B may be moved in the quantifier prefix anywhere before $\phi^{\mathbf{I}}$. The process may be repeated with a new prefixed B to obtain all B-closures of $\phi^{\mathbf{I}}$.

We now examine the various groups of axioms.

PROPOSITION 3.8. *The* **I** *images of the classical schemas* L1, L2, L4, L5 *are logically valid, as well as* L6 *for* θ *atomic, and* L8.

Prop. 3.7 then shows that the **I** images of their B-closures are provable in A.

The rest of the classical axioms (L3, L6) will be taken care of in Lemma 3.10.

Proof. L1, L8. The schema $B_\pi\theta \to \theta$ is tautologous $(\theta \wedge \pi[\theta] \to \theta)$, and $(B\theta \to \theta)^{\mathbf{I}} = B_\pi\theta^{\mathbf{I}} \to \theta^{\mathbf{I}}$ is an instance of this. Evidently I takes tautologies to tautologies in general.

L2, L4, L5; L6 for θ atomic. The **I** images of these are all instances of the schema, since **I** preserves the classical logical operations. The only point to check arises in the case of L6, where we must check that $\theta(x/u)^{\mathbf{I}} = \theta^{\mathbf{I}}(x/u)$. For later reference we prove this where u is any term. For α atomic in classical arithmetic, $\alpha^{\mathbf{I}} = \alpha$; since $\alpha(x/u)$ is also atomic,

$$[\alpha(x/u)]^{\mathbf{I}} = \alpha(x/u) = \alpha^{\mathbf{I}}(x/u).$$

(Here the brackets are used only as parentheses). For α of the form Tr(t),

$$[\mathrm{Tr}(t)(x/u)]^{\mathbf{I}} = [\mathrm{Tr}(t(x/u))]^{\mathbf{I}} = \mathrm{Tr}\bar{\mathbf{I}}_0(t(x/u)),$$

whereas

$$[\mathrm{Tr}(t)]^{\mathbf{I}}(x/n) = [\mathrm{Tr}(\bar{\mathbf{I}}_0(t))](x/u) = \mathrm{Tr}(\bar{\mathbf{I}}_0 t(x/u)),$$

which establishes the desired equality.

PROPOSITION 3.9. a) *Suppose that in* A *we have the universal closures of*

(i) $\pi[\theta]$, *for* θ *logically valid*,

(ii) $\pi[\theta \to \phi] \to (\pi[\theta] \to \pi[\phi])$

Then the universal closures of the **I** *images of the schemas* L7, L8, L10 *(propositional and quantifiers axioms for* B) *are provable in* A.

b) *Moreover, the hypothesis* (a) *of prop. 3.7 is also satisfied, that is* $B_\pi \forall x\theta \to \forall x B_\pi \theta$.

c) *If in addition*

(iii) $\pi[\theta] \to \pi[\pi[\theta]]$

then L9 *is also provable in* A.

Proof. For L7, we show the schema

$$B_\pi(\theta \to \phi) \to (B_\pi\theta \to B_\pi\phi).$$

This is just $(\theta \to \phi) \wedge \pi[\theta \to \phi] \to (\theta \wedge \pi[\theta] \to \phi \wedge \pi[\phi])$, and follows from (i), $\pi[\theta \to \phi] \to (\pi[\theta] \to \pi[\phi])$. The **I** images of L7 are instances of the former.

L8 was proved already.

For L9, we show the schema $B_\pi\theta \to B_\pi B_\pi\theta$, of which the **I** images of L9 are instances. This is just $\theta \wedge \pi[\theta] \to \theta \wedge \pi[\theta] \wedge \pi[\theta \wedge \pi[\theta]]$, for which it will suffice to see $\pi[\theta] \to \pi[\theta \wedge \pi[\theta]]$. Now $\theta \to (\pi[\theta] \to \theta \wedge \pi[\theta])$ is a tautology, so by (i)

$$\pi[\theta \to (\pi[\theta] \to (\theta \wedge \pi[\theta]))].$$

Thus using (iii) $\pi[\theta] \to \pi[\pi[\theta]]$ and several instances of (ii), we obtain the desired result.

For L10 we show the schema

(a) $$B_\pi \forall x\theta \to \forall x B_\pi \theta.$$

(This will satisfy the hypotheses (a) of 3.7). The schema $B_\pi \forall x\theta \to \forall x B_\pi(\exists y (x = y) \to \theta)$ follows from this using a tautology and the distributivity of B_π over $\to$ (from the proof of L7). Now (a) is just

$$\forall x\theta \wedge \pi[\forall x\theta] \to \forall x(\theta \wedge \pi[\theta]),$$

for which it will suffice to see

$$\pi[\forall x\theta] \to \pi[\theta].$$

Since $\forall x\theta \to \theta$ is a logical validity, by (i) we have $\pi[\forall x\theta \to \theta]$, so (ii) gives the desired result.

LEMMA 3.10. a) *Let* A *include* P *and satisfy assumption* (ii) *of Prop. 3.9a, as well as*

(i) $A \vdash \sigma$ *implies* $A \vdash \pi[\sigma]$.

Then the **I**-*images of* L3 *and* L6 *are provable in* A.

Suppose that in addition, A *satisfies*

(iv) $A \vdash (\theta \to \pi[\theta])$, *for all primitive recursive equations* θ.

Then the universal closures of the I *images of* L11, L12 *(the universal instantiation schemas for terms), are provable in* A.

Before giving the proof, we make the following remark regarding the significance of the antecedents of L11, L12,

REMARK 3.11.

a) Because of assumption (i), the antecedent of L12 is always satisfied. This is because $\exists y(t = y)$ is logically valid (classically), and (i) assures that $A \vdash B_\pi \exists y(t = y)$.

b) Since the schema $B\forall x\theta \rightarrow \forall xB\theta$ yield $\exists xB\theta \rightarrow B\exists x\theta$ (see the proof of Theorem 2.6 a), the antecedent of L11 becomes $\exists xB(t = z)$. This is

$$\exists z(t = z \wedge \pi[t = z]).$$

In view of (iv), it is always satisfied for primitive recursive t.

Proof. We prove the **I**-images of the conclusion of L11, L12,

$$\forall x\theta \rightarrow \theta(x/t).$$

(In case t is a variable, (iv) will not be needed, which will prove part a)). Now $(\forall x\theta \rightarrow \theta(x/t))^{\mathbf{I}} = \forall x\theta^{\mathbf{I}} \rightarrow (\theta(x/t))^{\mathbf{I}}$. Since $\forall x\theta^{\mathbf{I}} \rightarrow \theta^{\mathbf{I}}(x/t)$ is an instance of the classical instantiation schema, will thus suffice to prove the

CLAIM 3.12. *In* A *we have the closure of* $(\theta(x/t))^{\mathbf{I}} \leftrightarrow \theta^{\mathbf{I}}(x/t)$.

Proof. The proof proceeds by induction on the formula θ.

1. $\theta = \alpha$ atomic. The proof is the same as in the case t = u, u a variable, which was given in prop. 3.8 for L6; we get $(\theta(x/t))^{\mathbf{I}} = \theta^{\mathbf{I}}(x/t)$.

2. $\theta \rightarrow \phi$, $\neg\theta$, $\exists y\theta$ are all easily checked since substitution is a homomorphism on these logical operations.

3. $B\theta$. Assume for θ, that $A \vdash \theta^{\mathbf{I}}(x/t) \leftrightarrow (\theta(x/t))^{\mathbf{I}}$. We are to see that $A \vdash (B\theta)^{\mathbf{I}}(x/t) \leftrightarrow (B\theta(x/t))^{\mathbf{I}}$. The l.h.s. of this is $(\theta^{\mathbf{I}} \wedge \pi[\theta^{\mathbf{I}}])(x/t)$, i.e.

$$\theta^{\mathbf{I}}(x/t) \wedge (\pi[\theta^{\mathbf{I}}])(x/t),$$

whereas the r.h.s. is

$$(\theta(x/t))^{\mathbf{I}} \wedge \pi[(\theta(x/t))^{\mathbf{I}}].$$

Thus it will suffice to see that

$$\pi[\theta(x/t)^{\mathbf{I}}] \leftrightarrow (\pi[\theta^{\mathbf{I}}])(x/t).$$

Now by induction, $A \vdash \theta^{\mathbf{I}}(x/t) \leftrightarrow (\theta(x/t))^{\mathbf{I}}$. Using (i)-(iv) we can apply Prop. 3.9 b), and Prop. 3.7 to obtain any B_π closure of this in A, in particular

$$A \vdash \pi[\theta^{\mathbf{I}}(x/t) \leftrightarrow (\theta(x/t))^{\mathbf{I}}].$$

Hence by (ii) we may distribute π to obtain

$$A \vdash \pi[\theta^{\mathbf{I}}(x/t)] \leftrightarrow \pi[\theta(x/t)^{\mathbf{I}}].$$

Thus it will suffice to see in A that

$$\pi[\theta^{\mathbf{I}}(x/t)] \to (\pi[\theta^{\mathbf{I}}])(x/t).$$

We shall prove now in A the general

SCHEMA 3.13: $\pi[\phi(x/t)] \leftrightarrow (\pi[\phi])(x/t)$.

To do this we need the syntactic

FACT 3.14. *If* x,w *are variables, then* $P \vdash [\theta](x/w) \doteq [\theta(x/w)]$.

We leave this to the reader to check, but note that in the case of θ with one free variable x, it follows from the formalization of

$$(\theta(x/w))(x/\bar{n}) = \theta(x/\bar{n}).$$

Letting

$$\mathrm{Sub}(\theta,x,w) = \theta(x/w),$$

this is

$$\mathrm{Sb}(\mathrm{Sub}(\theta,x,w,n) = \mathrm{Sb}(\theta,x,n),$$

so the formalization is

$$\overline{\mathrm{Sb}}(\overline{\mathrm{Sub}}(v_o,v_1,v_2),v_2,v_3) \doteq \overline{\mathrm{Sb}}(v_o,v_1,v_3)$$

which yields on taking v_o to $\bar{\theta}$, v_1 to $\bar{x}$, v_2 to $\bar{w}$, and v_3 to w

$$\overline{\mathrm{Sb}}(\overline{\mathrm{Sub}}(\bar{\theta},\bar{w},\bar{w}),\bar{w},\bar{w}) \doteq \overline{\mathrm{Sb}}(\bar{\theta},\bar{x},\bar{w})$$

i.e.

$$\overline{\mathrm{Sb}}(\overline{\theta(x/w)},w,w) \doteq \overline{\mathrm{Sb}}(\bar{\theta},x,x)(x/w)$$

i.e.

$$[\theta(x/w)] \doteq [\theta](x/w).$$

We proceed to check the schema 3.13. Now using the above fact,

$$(\pi[\phi])(x/w) = \pi([\phi](x/w))$$
$$\leftrightarrow \pi[\phi(x/w)].$$

In case t is a variable, this is schema 3.13, so the proof of a) is now complete. In general we have, for w a variable not in ϕ or t,

$$\phi(x/t) \leftrightarrow \forall w(t = w \to \phi(x/w)),$$

$$(\pi[\phi])(x/t) \leftrightarrow \forall w(t = w \to (\pi[\phi])(x/w))$$

$$\leftrightarrow \forall x(t = \dot{w} \to \pi[\phi(x/w)]), \tag{1}$$

whereas (using (i), (ii))

$$\pi[\phi(x/t)] \leftrightarrow \pi[\forall w(t = w \to \phi(x/w))];$$

again using (i) etc. (as in Prop. 3.9b)

$$\rightarrow \forall w \pi[t = w \rightarrow \phi(x/w)],$$

sp by (ii)

$$\rightarrow \forall w(\pi[t = w] \rightarrow \pi[\phi(x/w)]).$$

Thus by (iv)

$$t = w \rightarrow \pi[t = w],$$

and it follows that

$$\pi[\phi(x/t)] \rightarrow \forall w(t = w \rightarrow \pi[\phi(x/w)]).$$

Thus by (1)

$$\pi[\phi(x/t)] \rightarrow (\pi[\phi])(x/t).$$

This completes one half of the proof of the schema. Note that we have used the computability of t.

For the other half, we use

$$\chi(x/t) \leftrightarrow \exists w(t = w \wedge \chi(x/w)):$$

$$\pi[\phi](x/t) \leftrightarrow \exists w(t = w \wedge \pi[\phi](x/w)) \qquad (2)$$

$$\leftrightarrow \exists w(t = w \wedge \pi[\phi(x,w)])$$

so that by (iv)

$$\rightarrow \exists w(\pi[t = w] \wedge \pi[\phi(x/w)]);$$

using (i) with a suitable tautology, and (ii), gives

$$\rightarrow \exists w(\pi[t = w \wedge \phi(x/w)])$$

so that using tha validity $\chi \rightarrow \exists w\chi$ under π, we have by (ii)

$$\rightarrow \pi[\exists w(t = w \wedge \phi(x/w))].$$

Thus applying (i) to (2) with $\chi = \phi$

$$\rightarrow \pi[\phi(x/t)]$$

which completes the proof of schema 3.13, and Lemma 3.10, and hence the checking of the logical axioms under I.

We turn now the arithmetical axioms.

PROPOSITION 3.15. *Let* A *be a theory in a classical language* L *including arithmetic, and suppose that* A *satisfies the Peano axioms. Let* π *satisfy the conditions* (i)-(iii) *of Lemma 3.10a, as well as*

(iv) $$A \vdash \bar{S}(x) = w \rightarrow \pi[\bar{S}(x) = w],$$
$$A \vdash \bar{0} = w \rightarrow \pi[\bar{0} = w].$$

Then the I_π *images of the universal closures of the schemas* A1-A3 *are provable in* A.

Proof. I_π takes A1, A2 to themelves, so no assumption is needed on π. The claim 3.12 for $t = \bar{0}$, $\bar{S}(x)$ obtains and $\mathbf{I}_\pi$ preserves the classical logical symbols, so A3 goes to a formula equivalent to an instance of A3.

This completes the proof of Theorem 3.5a.

Next the satisfaction axioms.

PROPOSITION 3.16. a) *For any formula* π, *the universal closures of the* $\mathbf{I}$ *images of the satisfaction axioms* S1-S3 (*for the language* L^B *of arithmetic with* B) *are provable in* P^+.

b) *If* π *satisfies the conditions*

(vi) $\mathrm{Sent}_L(x) \wedge \mathrm{Sent}_L(y) \wedge \pi(x \to y) \wedge \pi(x) \to \pi(y)$,

(vii) $\mathrm{Sent}_L(x) \wedge \pi_o(x) \to \mathrm{Tr}(x)$,

then the $\mathbf{I}$ *image of* S4 *is provable in* P^+.

c) *If in addition to* (vi) π *satisfies*

(v) $\mathrm{Sent}_L(y) \to \pi([\mathrm{Tr}(y)] \leftrightarrow y)$,

then S5 *is also obtained.*

Proof. For atomic formulas, $I_o\alpha = \alpha$, so in P

$$\mathrm{Tm}(u) \wedge \mathrm{Tm}(v) \to \bar{\mathbf{I}}(u \doteq v) \doteq (u \doteq v).$$

Thus

$$\begin{aligned} \mathrm{Tr}^{\mathbf{I}}(u \doteq v) &\leftrightarrow \mathrm{Tr}\bar{\mathbf{I}}(u \doteq v) \\ &\leftrightarrow \mathrm{Tr}(u \doteq v), \end{aligned}$$

and S1 follows from S1 in P^+.

S2 becomes

$$\mathrm{Sent}_{(L^B)}(x) \to (\mathrm{Tr}(\bar{\mathbf{I}}_o \bar{\neg} x) \leftrightarrow \neg \mathrm{Tr}\bar{\mathbf{I}}_o x). \tag{1}$$

Since $\mathbf{I}_o \neg\sigma = \neg \mathbf{I}_o\sigma$, in P we have

$$(\bar{\mathbf{I}}_o \bar{\neg} x) \doteq (\bar{\neg}\bar{\mathbf{I}}_o x)$$

Using this and $P \vdash \mathrm{Sent}_{(L^B)}(x) \to \mathrm{Sent}_L(\mathbf{I}_o x)$, (1) follows from S2 in P^+. S3 is proved the same way, formalizing $\mathbf{I}_o(\theta \to \phi) = \mathbf{I}_o\theta \to \mathbf{I}_o\phi$.

b) For the proof of S4 we need the formalization of Claim 3.12. Taking $A = P$ in that Claim, and confining our attention to the case where θ has one free variable w, and t is a term $\bar{n}$, this yields

$$\left\{ \begin{array}{l} \text{For all n, and all formulas } \theta \text{ with one free variable w,} \\ (3) \quad P \vdash (\theta(w/\bar{n}))^{\mathbf{I}} \leftrightarrow \theta^{\mathbf{I}}(w/\bar{n}). \end{array} \right. \tag{2}$$

The proof of (2) may be carried out in P; since (3) may be written as

$$P \vdash \mathbf{I}\mathrm{Sb}(\theta,w,n) \leftrightarrow \mathrm{Sb}(\mathbf{I}\theta,w,n).$$

and π_o expresses provability in P, the formalization of (2) is

(vii') $$P \vdash \mathrm{Fm}(x,w) \to \pi_o(\bar{I}\overline{Sb}(x,w,z) \bar{\leftrightarrow} \overline{Sb}(\mathbf{\bar{I}}x,w,z))$$

as indicated after Lemma 3.6. Here Fm(x,w) means for L with B,Tr. At this point we only need it for L^B without Tr; for such we get $\mathrm{Fm}_{(L^B)}(x) \to \bar{I}(x) = \bar{I}_o(x)$, of course. Now S4 is

$$\mathrm{Fm}(x,w) \to (\mathrm{Tr}(\bar{\forall}wx) \leftrightarrow \forall z \mathrm{Tr}(\overline{Sb}(x,w,z))),$$

so we must see that $\mathrm{Fm}_{(L^B)}(x,w) \to$

$$\mathrm{Tr}(\mathbf{\bar{I}}_o\bar{\forall}wx) \leftrightarrow \forall z \mathrm{Tr}\mathbf{\bar{I}}_o\overline{Sb}(x,w,z).$$

Now, $\mathbf{I}_o\forall u\theta = \forall u\mathbf{I}_o\theta$, so in P, $\mathbf{\bar{I}}_o\bar{\forall}wx \doteq \bar{\forall}x\mathbf{\bar{I}}_ox$. Thus

$$\mathrm{Tr}(\mathbf{\bar{I}}_o\bar{\forall}wx) \leftrightarrow \mathrm{Tr}(\bar{\forall}w\mathbf{\bar{I}}_ox),$$

and using S4 in P^+,

$$\leftrightarrow \forall z \mathrm{Tr}\overline{Sb}(\mathbf{I}_ox,w,z).$$

Using (vii'), (vii), and S2,3 in P^+ gives the result.

c) S5. We shall need the following

LEMMA 3.17. a) *If* θ *is any formula of the language* L *of Peano arithmetic, then in* P^+ *we have the closure of* $\mathrm{Tr}[\theta] \leftrightarrow \theta$.

b) *In particular, restricting* θ *to sentences gives: for all sentences* σ *of* L, $P^+ \vdash (\mathrm{Tr}(\bar{\sigma}) \leftrightarrow \sigma)$.

c) *The proof of* (b) *can be carried out in* P, *so that if* π *expresses probability in* P^+ *(or something stronger)*

$$P \vdash \forall y(\mathrm{Sent}_L(y) \to \pi([\mathrm{Tr}(y)] \bar{\leftrightarrow} y))$$

Proof. These are standard results so we indicate the proof only briefly.

a) Proved by induction on θ. The notation $[\theta]$ extends to terms t. For atomic formulas use $[t \doteq s] \doteq ([t] \bar{\doteq} [s])$ and $\mathrm{den}([t],t)$. For the other cases use $[\neg\theta] = \bar{\neg}[\theta]$, etc., and the corresponding clause of the satisfaction axioms.

b) is immediate from a).

c) Note that

$$\mathrm{Tr}(\bar{\sigma}) = (\mathrm{Tr}(y))(y/\bar{\sigma}) = \mathrm{Sb}(\mathrm{Tr}(y),y,\sigma),$$

so its formalization is $\overline{Sb}(\overline{\mathrm{Tr}(y)},\bar{y},y) = [\mathrm{Tr}(y)]$.

Proof of 3.16. The new satisfaction axiom is

$$\mathrm{Sent}(x) \to (\mathrm{Tr}(\bar{B}x) \leftrightarrow B\mathrm{Tr}(x)),$$

so under $\mathbf{I}$ it becomes

$$\mathrm{Sent}(x) \to (\mathrm{Tr}(\bar{I}_o\bar{B}x) \leftrightarrow B_\pi\mathrm{Tr}(\mathbf{I}_ox)). \qquad (1)$$

Here Sent(x) expresses "is a sentence in the language L^B of arithmetic with

B", but of course without Tr. Now

$$\mathbf{I}_o B\sigma = \mathbf{I}_o\sigma \wedge \pi[\mathbf{I}_o\sigma] = \mathbf{I}_o\sigma \wedge \pi(\overline{\mathbf{I}_o\sigma})$$
$$= \mathbf{I}_o\sigma \wedge \mathrm{Sb}(\pi, y, \mathbf{I}_o\sigma).$$

Where y is the free variables of π. The formalization of this is

$$\mathrm{Sent}(x) \to \bar{\mathbf{I}}_o\bar{B}x \doteq \bar{\mathbf{I}}_o x \bar{\wedge}\ \overline{\mathrm{Sb}}(\bar{\pi}, \bar{y}, \bar{\mathbf{I}}_o x)\ ,$$

which is provable in P. Thus we have, suppressing the antecedent Sent(x),

$$\mathrm{Tr}(\bar{\mathbf{I}}_o\bar{B}x) \leftrightarrow \mathrm{Tr}(\bar{\mathbf{I}}_o x \bar{\wedge}\ \overline{\mathrm{Sb}}(\bar{\pi}, \bar{y}, \bar{\mathbf{I}}_o x)) \tag{2}$$

so that by S2, S3 in P^+,

$$\leftrightarrow \mathrm{Tr}(\bar{\mathbf{I}}_o x) \wedge \mathrm{Tr}(\overline{\mathrm{Sb}}(\bar{\pi}, \bar{y}, \bar{\mathbf{I}}_o x)).$$

Now

$$B_\pi \mathrm{Tr}(\bar{\mathbf{I}}_o x) \leftrightarrow \mathrm{Tr}(\bar{\mathbf{I}}_o x) \wedge \pi[\mathrm{Tr}\bar{\mathbf{I}}_o x], \tag{3}$$

so to check (1) it will by (2), (3) be enough to check

$$\mathrm{Tr}(\overline{\mathrm{Sb}}(\bar{\pi}, \bar{y}, \bar{\mathbf{I}}_o x)) \leftrightarrow \pi[\mathrm{Tr}\bar{\mathbf{I}}_o x].$$

Now in P

$$\mathrm{Sent}_{(L^B)}(x) \to \mathrm{Set}_L(\bar{\mathbf{I}}x),$$

so it will suffice (by instantiating $\bar{\mathbf{I}}_o x$ for y) to see

$$\mathrm{Sent}_L(y) \to (\mathrm{Tr}(\overline{\mathrm{Sb}}(\bar{\pi}, \bar{y}, y)) \leftrightarrow \pi[\mathrm{Tr}(y)]),$$

i.e., suppressing the antecedent, and noting $[\pi] = \overline{\mathrm{Sb}}(\bar{\pi}, \bar{y}, y)$

$$\mathrm{Tr}[\pi(y)] \leftrightarrow \pi[\mathrm{Tr}(y)]. \tag{4}$$

The result now follows easily from Lemma 3.17. By 3.17a,

$$\mathrm{Tr}[\pi(y)] \leftrightarrow \pi(y).$$

While by 3.17c,

$$\pi([\mathrm{Tr}(y)] \mathrel{\bar{\leftrightarrow}} y). \tag{5}$$

Thus by (vi) (modus ponens for π),

$$\pi[\mathrm{Tr}(y)] \leftrightarrow \pi(y). \tag{6}$$

Combining (5), (6) gives (4) and hence completes the proof of 3.16.

This completes the proof of Theorem 3.5b.

We now turn to CT(Bw).

PROPOSITION 3.18. *If* π *satisfies the condition* (vi) (vii) *of Prop. 3.16 b, as well as*

(viii) $P^{++} \vdash \mathrm{Sent}_L(x) \wedge \pi(x) \to \mathrm{Tr}(x),$

then the **I** *image of* CT(Bw) *is provable in* P^+.

Proof. The **I** image of CT(Bw) is

$$\exists e \forall x(Bw^{\mathbf{I}}(x) \leftrightarrow Uex),$$

i.e.

$$\exists e \forall x(Sent(x) \wedge Tr(\bar{I}_0 \bar{B} x) \leftrightarrow Uex).$$

As before (in 3.16c), this is

$$Bw^{I}(x) \leftrightarrow Tr(\bar{I}_0 x) \wedge Tr(\overline{Sb}(\bar{\pi}, \bar{y}, \bar{I}_0 x))$$
$$\leftrightarrow Tr(\bar{\mathbf{I}}_0 x) \wedge \pi(\bar{\mathbf{I}}_0 x).$$

But since π is a correct notion of proof (viii),

$$\leftrightarrow \pi(\bar{I}_0 x).$$

Finally, since π is an r.e. formula, there is a Turing machine e which enumerates the arithmetical sentences σ such that $P^+ \vdash \sigma$; since I_0 is recursive, it is evident that $\{\sigma | P^+ \vdash \sigma^{I_0}\}$ is also. I.e. that

$$\exists e \forall x(\pi(\bar{I}_0 x) \wedge Sent_{(LB)}(x) \leftrightarrow Uex),$$

so that

$$\exists e \forall x(Bw^{\mathbf{I}}(x) \leftrightarrow Uex)$$

as desired.

This completes the proof of Theorem 3.5.

It is now easy to prove Corollary 3.2. We must see the schema BCT for formulas θ of the language of BP (i.e. without Tr).

Proof of 3.2. Suppose that $\forall x(\theta \rightarrow B\theta)$. Then

$$\theta \leftrightarrow B\theta$$
$$\leftrightarrow BTr[\theta]$$
$$\leftrightarrow Bw[\theta]$$
$$\leftrightarrow U(e, [\theta]).$$

Now this just says that the inverse image of a certain r.e. set (Bw) under the recursive function $f(n) = \theta(\bar{n})$ is also r.e. Indeed, there is a primitive recursive function g such that if e is the Turing machine for a set E, g(e) will be the Turing machine for $f^{-1}E$ (and moreover, one can get g effectively from f). Thus we have in P,

$$\forall e(\forall x(U(e, \bar{f}x) \leftrightarrow U(\bar{g}e, x)), \qquad (1)$$

so that

$$\forall e[\forall x(Bw(x) \leftrightarrow Uex) \rightarrow \forall x(Bw(\bar{f}x) \leftrightarrow U(\bar{g}e, x))].$$

Since $\bar{f}(x) = [\theta(x)]$, this completes the proof of Corollary 3.2.

To complete the proof of Theorema 3.1, we extend the definition of **I** to the language of $(BP)^{++}$ using for $\pi(x)$ a formula expressing "$P^{++} \vdash x$": Let $\mathbf{I}_1$ be the previous definition and put

$$Tr_+^I(x) = Tr_+(\bar{\mathbf{I}}_1(x)).$$

We show that **I** interpretes $(BP)^{++} + CT(Bw_+)$.

Prop. 3.8 requires the addition clauses (obtained by replacing Tr by Tr_+ and $\mathbf{I}_o$ by $\mathbf{I}_1$):

$$[Tr_+(\tau)(x/u)]^{\mathbf{I}} = [Tr_+(t(x/u))]^{\mathbf{I}} = Tr_+\bar{\mathbf{I}}_1(t(x/u)),$$

$$[Tr_+(\tau)]^{\mathbf{I}}(x/u) = [Tr_+(\bar{\mathbf{I}}_1(\tau)](x/u) = Tr_+(\bar{\mathbf{I}}_1 t(x/u)).$$

In Prop. 3.16, the first new satisfaction axiom is $S1^+$:

$$Tm(u) \to (Tr_+\overline{Tr}u \leftrightarrow Tr(u)),$$

which goes to $Tm(u) \to$

$$Tr_+(\bar{\mathbf{I}}_1\overline{Tr}u) \leftrightarrow Tr(\bar{\mathbf{I}}_o(u)).$$

Now

$$Tr^{\mathbf{I}}(x) = Tr\bar{\mathbf{I}}_o(x),$$

so in P

$$(\bar{\mathbf{I}}_1\overline{Tr}\,x) \doteq (\overline{Tr}\,\bar{\bar{\mathbf{I}}}_o x).$$

Thus

$$\begin{aligned} Tr_+(\mathbf{I}_1\overline{Tr}u) &\leftrightarrow Tr_+(\overline{Tr}\bar{\bar{\mathbf{I}}}_o u) \\ &\leftrightarrow Tr_+|Tr\bar{\mathbf{I}}_o u] \\ &\leftrightarrow Tr(\bar{\mathbf{I}}_o u). \end{aligned}$$

The new axioms $S2^+$, $S3^+$, $S4^+$ are the same as the old with Tr replaced by Tr_+, and the antecedent syntactic conditions (Fm,Sent) chaged accordingly. Thus the proofs are the same with I_o replaced by $\mathbf{I}_1$. The proof of $S5^+$ requires use of

$$P^{++} \vdash Tr_+[\theta] \leftrightarrow \theta\ ,$$

and the various formalizations of Lemma 3.17. The proof is again the same, replacing Tr by Tr_+, $\mathbf{I}_o$ by $\mathbf{I}_1$. The proof of $CT(Bw_+)$ requires the correctness of π, which is now provable in P^{+++}, but is otherwise the same. Theorem 3.1 now follows by the same argument as in the proof of 3.2.

We would like to make a few remarks to indicate the nature of the difficulty in extending these results to obtain conjectures 3.1a, 3.2a.

The situation now is that we have T consistent (where T is $(BP)^{\dagger}$), $T \vdash \alpha$, and we would like to show the consistency of $T \cup \{B\alpha\}$. However, it is certainly not true in general that $T \cup \{B\alpha\}$ is consistent. For example, if α is $(\gamma \wedge \neg B\gamma)$, then $B\alpha$ (and hence $T \cup \{B\alpha\}$) is inconsistent:

$$B(\gamma \wedge \neg B\gamma) \rightarrow B\gamma \wedge B \neg B\gamma$$
$$\rightarrow B\gamma \wedge \neg B\gamma.$$

Now CT(Bw) certainly implies that there are sentences such as γ; in fact BCT implies for some arithmetical θ

$$B\exists e(\theta(e) \wedge \neg B\theta(e))$$

so that

$$\exists e(\theta(e) \wedge \neg B\theta(e))$$

is provable in T.

We remark that one obvious way to try to interpret BCT(Bw) is to use the same method used for CT(Bw) but to replace $\pi[\theta]$ by $\pi[\alpha \rightarrow \theta] = \pi_{\alpha}[\theta]$. It turns out that no such interpretation will have BCT(Bw). In fact, if π is any reasonable proof predicate (i.e. one satisfying Gödel's second theorem), then de-Jongh and the author have observed that BCT(Bw) fails in the Löb interpretation using π.

To see this, consider the following schema, which is a consequence of BCT:

1) $B\exists e\forall x(B\theta(x) \leftrightarrow U(e,[\theta(e)]))$,

where $\theta(x)$ is a formula with one free variable x. This is equivalent to BCT($B\theta$), since the correspondence $x \mapsto [\theta(x)]$ may be assumed to be 1-1 recursive. Apply the fixed point theorem to obtain $\theta(x)$ an arithmetical formula (no B's, no Tr's, etc.) so that

2) $B\forall x(\theta(x) \leftrightarrow \neg U(x,[\theta(x)]))$.

Now we also have in general

3) $B\forall x(B\theta(x) \rightarrow \theta(x))$.

Thus taking x to be e, we have $B\exists e\phi(e)$, where $\phi(e)$ is the conjunction of

1') $B\ (e) \leftrightarrow U(e,[\theta(e)])$

2') $\theta(e) \leftrightarrow \neg U(e,[\theta(e)])$

3') $B\theta(e) \rightarrow \theta(e)$.

Since $\theta(e) \wedge \neg B\theta(e)$ is a tautological consequence of $\theta(e)$, we thus have

4) $B\exists e(\theta(e) \wedge \neg B\theta(e))$.

Consequently, any interpretation which makes (1) come out true will also have (4). In particular, if $B\theta$ is interpreted as $\theta \wedge \pi[\theta]$ (which, since θ is arithmetical, will be the case for any Löb interpretation), the truth of 4) gives

$$\pi[\exists e(\theta(e) \wedge \neg(\theta(e) \wedge \pi[\theta(e)]))]$$

$$\vdash_{\pi} \exists e(\theta(e) \wedge \neg\pi[\theta(e)])$$

$$\vdash_{\pi} \exists e(\neg\pi[\theta(e)])$$

$$\vdash_\pi \exists x \neg\pi(x)$$
$$\vdash_\pi Con_\pi,$$

in other words, Gödel's second theorem fails for π.

Agradecimiento. Muchas gracias a los organizadores de este simposio, especialmente a María Mercedes Sánchez A. y a los Profesores Xavier Caicedo, Clara Helena Sánchez y Rolando Chuaqui.

REFERENCES.

1. Benacerraf, Paul, *God, the devil, and Gödel*, The Monist 51 (1967), 9-32.
2. Boolos, George, *The Unprobability of Consistency*, Cambridge, Cambridge University Press, 1979.
3. Gödel, K., *Eine Interpretation des intuitionistischen Aussagenkalküls*, Ergebnisse eines mathematischen Kolloquiums 4 (1933), 39-40.
4. Gödel, K., *25th* Joshia Willard Gibbs Lecture at Providence, December 26, 1951 (unpublished).
5. Kalmar, Lászlő. *Constructivity in Mathematics*, Proceedings of a Colloquium held at Amsterdam 1957, Ed. A. Heyting, North Holland, 1959, pp.72-80.
6. Lob, M.H., *Solution of a problem of Henkin*, Journal of Symbolic Logic 20 (1955), 115-118.
7. Lucas, J.R., *Minds, machines, and Godel*, Philosophy, vol. 36 (1961), 112-127.
8. Montague, R., *Syntactical Treatments of Modality, with Corollaries of Reflection Principles and Finite Axiomatizability*, Acta Philosophica Fennica 16 (1963), 153-167. Reprinted in R. Thomason (ed), *Formal Philosophy*, Yale (1974).
9. Myhill, John, *Some Remarks on the Notion of Proof*, Journal of Philosophy, Vol. LVII (1960).
10. Shapiro, Stewart, *Epistemic and Intuitionistic Arithmetic* (preprint, Dept. of Philosophy, Ohio State University, Newark, Ohio 43055), 1981.
11. Tharp, Leslie, H., *All Truths Are Absolutely Probable*, 1973.
12. Reinhardt, W.N., *Necessity predicates and operators*, Journal of Philosophical Logic 9 (1980), 437-450.
13. Wang, Hao, *From Mathematics to Philosophy*, Humanities Press, N.Y., 1974.

***

Department of Mathematics
University of Colorado
Boulder, Colorado, U.S.A.