Bollettino U. M. I.
(7) 3-A (1989), 177-184

Σ_n^0-Interpretations of Modal Logic.

ALESSANDRO BERARDUCCI

Sunto. – *Consideriamo la logica modale predicativa della dimostrabilità nella Aritmetica di Peano (PA), in cui però la interpretazione di ogni formula modale atomica viene ristretta ad appartenere all'insieme Σ_n^0. Viene dimostrato che per distinti n le corrispondenti logiche modali formano una gerarchia stretta e che ognuna di esse è Π_2^0-completa.*

1. – Introduction.

The language L_0 of (predicate) modal logic consists of a countable set of n-ary relation symbols (for every $n > 0$), propositional constants (i.e. 0-ary relation symbols), variables, boolean connectives, quantifiers, and the unary operator $\Box$ (box). In other words we expand the language of first order logic (without identity and function symbols) by adjoining the operator $\Box$. An (arithmetical) interpretation of the modal language L_0 is a map f which assigns to every formula θ of L_0, a formula θ' of the (first order) language of Peano Arithmetic (PA) having the same free variables and such that f commutes with boolean connectives and quantifiers and interprets the $\Box$-operator as the provability predicate of PA. So for example $(\forall x \Box\, \theta(x))' = \forall x\, \mathrm{Thm}_{PA}(\mathrm{sub}(\ulcorner\theta(v_0)'\urcorner, x))$. The usual conventions about renaming of bound variables are assumed. We also require that f commutes with substitutions of variables, hence if two atomic formulas differ only by a change of variables so do their interpretations. Clearly f is determined by its restriction to the atomic formulas. The (predicate modal) logic of provability is the set of all the modal formulas θ such that for every arithmetical interpretation f, $PA \vdash \theta'$. So for example $\Box \forall x A(x) \to \forall x \Box A(x)$ belongs to the logic of provability. In [Var85] Vardayan showed that the logic of provability is Π_2^0-complete (thus answering a question by F. Montagna in [Mon84]). For Γ a set of formulas of PA and f an (arithmetical) interpretation, we say that f is a Γ-intepretation if $A' \in \Gamma$ whenever A is atomic. The Γ-logic of provability is the set of all the (predicate) modal formulas θ such that for all Γ-interpretations f, $PA \vdash \theta^f$. So for example $A \to \Box A$ belongs to the Σ^0_1-logic of provability but not to the Π^0_1-logic of provability. Let S_n be the Σ^0_n-logic of provability and P_n be the Π^0_n-logic of provability. It is clear that $S_n \supseteq S_{n+1}$ and $P_n \supseteq P_{n+1}$ $(n \geqslant 1)$. We prove that all these inclusions are strict and that for each $n \geqslant 1$ the modal theories S_n and P_n are Π^0_2-complete (so in particular they are not recursively axiomatizable). Note that in this terminology the logic of provability is just the intersection of all the theories S_n (and is Π^0_2-complete by Vardayan's result). To extablish our results we will use a technique employed in [Var85], [Art85] and [BM87], namely a formalized version of Tennenbaum's theorem (every recursive model of PA is standard), which we will present in the next section. To place our results in their context we want to mention some related investigations about the «true» logics of provability. In [Sol76] R. Solovay proved that the true propositional modal logic of provability, namely the theory $\{\theta$ propositional modal formula $|\forall f, \omega \models \theta^f\}$, is decidable. In [Art85] Artemov showed that the true predicate modal logic of provability is not arithmetical and in [BM87] Boolos and McGee proved that this theory is actually Π^0_1-complete in the theory of ω. From the proof of our results it will follow that the true Σ^0_n-logics of provability, namely the predicate modal theories $\{\theta |$ for every Σ^0_n-interpretation f $\omega \models \theta^f\}$, are all distinct for different natural numbers $n \geqslant 1$.

2. – Tennenbaum's theorem.

Tennenbaum's theorem says that every recursive model of PA is standard (it is known on the other hand that there are Δ^0_2 non-standard models of PA). It turns out that the full strenght of PA is not actually needed, a big enough finitely axiomatizable subtheory, for example $I\Sigma_1$ will suffice. So if M is a recursive model of $I\Sigma_1$, then $M \cong \omega$. Our next goal is to state a formalized version of Tennenbaum's theorem which is provable in PA. In the following we will always assume that the language L_0 of modal logic contains $L(PA)$ (formulated as a relational language). So L_0 contains the relation symbols $0(x)$ (zero), $s(x, y)$ (successor), $+(x, y, z)$, $\cdot(x, y, z)$, $x = y$. An arithmetical interpretation f of L_0 induces a model of the language $L(PA)$ (with underlying set ω) simply by interpreting $0, s, +, \cdot, =$ as the relation defined by $0^f, s^f, +^f, \cdot^f, =^f$ in ω. We do not assume that $=^f$ is the identity relation. Let $\psi_{I\Sigma_1}$ be a sentence which axiomatizes $I\Sigma_1$. Since $L_0 \supset L(PA)$, $\psi_{I\Sigma_1}$ is a sentence of L_0, so it makes sense to speak of the relativized sentence $\psi^f_{I\Sigma_1}$ (which is a sentence of PA). Now let $A(x)$ be a formula of PA. We express, in PA, the fact that $\{x|A(x)\}$ is recursive by:

$$\exists e\, \forall x\, [A(x) \to \{e\}(x) = 1 \wedge \neg A(x) \to \{e\}(x) = 0]\,.$$

In the following definitions f is a fixed arithmetical interpretation of the modal language L_0.

DEFINITION. – «*f induces a recursive model of $I\Sigma_1$*» is defined as the sentence of PA obtained by taking the conjunction of $\psi^f_{I\Sigma_1}$ and the sentences expressing that $0^f, s^f, +^f, \cdot^f, =^f$ are recursive.

Notice that if the sentence just defined holds (in the standard model), then the structure of the language of PA induced by f is indeed a recursive model of $I\Sigma_1$. As already remarked we do not assume that $=^f$ is the identity relation, hovewer if $I\Sigma_1$ holds, then in particular $=^f$ is a congruence, hence we can consider the quotient structure modulo $=^f$ (which will still be a model of $I\Sigma_1$). The fact that this quotient model is standard can be expressed in PA as follows:

DEFINITION. – «*The quotient model induced by f is standard*» is the sentence $\forall y\, \exists x\, R_f(x, y)$ where $R_f(x, y)$ is the formula of PA which naturally expresses the following: $\exists u$: u codes a finite sequence of length $x + 1$ and

1) $u(x) = y$;

2) $\forall z < x$, $s^f(u(z), u(z+1))$;

3) $u(0) = 0^f$ (that is $\exists z, w[u(z) = w \wedge 0(z) \wedge 0^f(w)]$).

REMARK. – Note that $\psi^f_{I\Sigma_1}$ implies, in PA, that $\forall x\, \exists y\, R_f(x, y)$. We can think at R_f as the graph of an embedding of the standard model into the model induced by f (except that $=^f$ migh not be the identity). So $\forall y\, \exists x\, R_f(x, y)$ says that this embedding is onto. Notice that if f is a Σ^0_n-interpretation, then $R_f(x, y)$ is a Σ^0_n-formula. This will be important in the sequel.

The promised formalized version of Tennenbaum's theorem, which we state without proof, is the following:

THEOREM 1. – *For every interpretation f PA proves: «f induces a recursive model of $I\Sigma_1$» $\to$ «the quotient model induced by f is standard».*

This version of Tennenbaum's theorem is essentially Lemma 4 of [BM87]. The results of [Art85], [Var85] and [BM87] are all based on the above theorem and on the existence of a modal sentence I (of L_0) such that for every interpretation f, the relativised sentence I^f implies, in PA, that f induces a recursive model of $I\Sigma_1$.

DEFINITION. – The modal formula «I» is the conjunction of the following:

1) $\psi_{I\Sigma_1}$
2) $\neg\Box\bot$
3) $\forall x_1, x_2, x_3(A(\boldsymbol{x}) \to \Box A(\boldsymbol{x}))$
where A ranges over the atomic and negated atomic formulas of PA, that is $A(x)$ is one of the following: $0(x_1)$, $s(x_1, x_2)$, $+(x_1, x_2, x_3)$, $\cdot(x_1, x_2, x_3)$, $x_1 = x_2$, or the negation of one of these formulas.

THEOREM 2. – *For every interpretation f PA proves: $I^f \to$ «f induces a recursive model of $I\Sigma_1$».*

PROOF. – I^f says that PA is consistent and that the relations 0^f, s^f, $+^f$, $\cdot^f$, $=^f$ are decided by PA. This clearly implies that these relations are recursive.

COROLLARY. – For any sentence φ of PA (and every interpretation f) PA proves $I^f \to [\varphi \leftrightarrow \varphi^f]$. More generally for any formula $\varphi(\boldsymbol{x})$ of PA, PA proves: I^f implies $\varphi(\boldsymbol{x}) \leftrightarrow \varphi(\boldsymbol{y})^f$ whenever the x's and the y's satisfy the relation $R_f(x, y)$.

PROOF. – By meta-induction on φ using the fact that I^f implies that the quotient model induced by f is standard.

This result can be extended as follows (with a similar proof): Let $A(y)$ be a new predicate symbol (not in $L(PA)$), let $\varphi[A]$ be a sentence in the enriched language, and let E be the sentence expressing the fact that «$=$» is a congruence (even with respect to the new predicate A). Then we have:

THEOREM 3. – *PA proves $(I\wedge E)^f \to (\varphi[\hat{A}] \leftrightarrow \varphi[A]^f)$ where $\hat{A}(x) = \exists y\,[A^f(y)\wedge R_f(x, y)]$. More generally PA proves that $(I\wedge E)^f$ implies $\varphi[\hat{A}](\boldsymbol{x}) \leftrightarrow \varphi[A](\boldsymbol{y})^f$ whenever the x's and the y's satisfy the relation $R_f(x, y)$*

We can think at $\hat{A}$ as the isomorphic image in the standard model of the set defined by A^f in the quotient model induced by f (this makes sense since $=^f$ is a congruence with respect to A^f). For the details of the proof the reader can see [BM87] (Lemma 7).

3. – Σ_2^0-completeness.

We are finally in a position to prove that the theory S_n ($n\geq 1$) defined in the introduction is Π_2^0-complete. It is clear that S_n is a Π_2^0 set of formulas. To prove that it is complete we will show that for every $e \in \omega$ we can effectively find a modal sentence θ_e of L_0 such that

$$\{e\} \text{ is total} \leftrightarrow \text{for each } \Sigma_n^0\text{-interpretation } f,\ PA \vdash \theta_e^f$$

This will suffice since $\{e|e \text{ is total}\}$ is Π_2^0-complete.

Our modal language L_0 will include $L(PA)$ and a new unary relation symbol $A(x)$. For $n\geq 1$ we know that there is a Σ_n^0-formula of PA, $\mathrm{True}_{\Sigma_n^0}(x, y)$, which gives a truth definition for Σ_n^0-formulas of PA with one free variable. So if $\theta(v_0) \in \Sigma_n^0$ and $x = \ulcorner\theta(v_0)\urcorner$, then $\mathrm{True}_{\Sigma_n^0}(x, y) \leftrightarrow \theta(y)$ (provably in PA). Since $L_0 \supset L(PA)$, $\mathrm{True}_{\Sigma_n^0}(x, y)$ is a formula of L_0.

DEFINITION. – Let $B_e[A]$ be the formula (of L_0) which says that $\forall v$ if v is the least number such that $\forall y\,[\mathrm{True}_{\Sigma_n^0}(v, y)\leftrightarrow A(y)]$, then $(\forall u<v)\{e\}(u)\downarrow$.

DEFINITION. – Let E be the formula (of L_0) which says that «$=$» is a congruence (even with respect to the new predicate $A(x)$).

DEFINITION. – θ_e is the modal formula $I\wedge E \to B_e[A]$. Note that the $\Box$-operator appears only in the subformula I.

THEOREM 4. – *For $n\geq 1$, $\{e\}$ is total $\leftrightarrow$ for each Σ_n^0-interpretation f, $PA \vdash \theta_e^f$.*

PROOF. – ($\to$): Suppose $\{e\}$ is total. Let f be a Σ_n^0-interpretation. We have to show that $PA \vdash \theta_e^f$, that is $PA \vdash (I\wedge E \to B_e[A])^f$. Let $M \models PA$. Assume $(I\wedge E)^f$ holds in M. Let $\hat{A}(x) = \exists y\,(A^f(y)\wedge R_f(x, y))$. Then, in M, $B_e[A]^f \leftrightarrow B_e[\hat{A}]$ (by the results about Tennenbaum's theorem). Therefore it is enough to show that $B_e[\hat{A}]$ holds (in M). Now $B_e[\hat{A}]$ says that if v is the least element such that $\forall y\,[\mathrm{True}_{\Sigma_n^0}(v, y) \leftrightarrow \hat{A}(y)]$ then $(\forall u<v)\{e\}(u)\downarrow$. So let $v \in M$ be such a least element. Since f is a Σ_n^0-interpretation, $A^f \in \Sigma_n^0$ and (by a previous remark) $R_f \in \Sigma_n^0$. It follows from the definition of A that $\hat{A} \in \Sigma_n^0$. Thus clearly $v<\ulcorner\hat{A}\urcorner$. So v is standard. Now $\{e\}$ is total and «$\{e\}(x)\downarrow$» is a Σ_1^0-assertion, so for every standard n, $M \models \{e\}(n)\downarrow$, whence $M \models (\forall u<v)\{e\}(u)\downarrow$ and we are done.

($\leftarrow$): We will prove the following stronger assertion: if for every Σ_1^0-interpretation f, $\omega \models \theta_e^f$, then $\{e\}$ is total. So assume the antecedent. Let $k \in \omega$ and let f_k be the Σ_1^0-interpretation defined as follows:

1) f_k is standard on $L(PA)$
(that is 0^f, s^f, $+^f$, $\cdot^f$, $=^f$ coincide with 0, s, $+$, $\cdot$, $=$)

2) $A^{f_k}(y) \leftrightarrow (y = k)$
(more precisely $A^{f_k}(y)$ is the Σ_1^0-formula of our relational language of PA which naturally expresses « $y = k$ »)

By hypothesis $\omega \models \theta_e^{f_k}$, that is $\omega \models (I \wedge E \to B_e[A])^{f_k}$. Since f_k is standard on $L(PA)$, $\omega \models (I \wedge E)^{f_k}$. Therefore, $\omega \models B_e[A]^{f_k}$. So, using again the fact that f_k is standard on $L(PA)$, $\omega \models B_e[A^{f_k}]$. Hence, by definition of B_e, if $v_k \in \omega$ is the least number such that $\forall y\, [\mathrm{True}_{\Sigma_n^0}(v_k, y) \leftrightarrow A^{f_k}(y)]$, then $(\forall u < v_k)\{e\}(u)\downarrow$ (in ω). So to prove that $\{e\}$ is total it is enough to show that if $k \to \infty$, then $v_k \to \infty$. But this is clear since the map which sends k to v_k is injective.

4. – Hierarchy theorem.

We will prove that all the inclusions $S_n \supset S_{n+1} (n \geq 1)$ are strict; this will follow immediately from the following stronger result:

THEOREM 5. – *For every $n \geq 1$ there is a modal sentence θ_n such that*:

1) *for each Σ_n^0-interpretation f, $PA \vdash \theta_n^f$*

2) *there is a Π_n^0-interpretation f such that $\omega \not\models \theta_n^f$*

Like in the previous proofs our modal language will include $L(PA)$ and a new unary predicate $A(x)$. The formulas I and E are defined above. We recall in particular that E says that « $=$ » is a congruence even with respect to A. The desired modal sentence $\theta_n = \theta_n[A]$ is defined as follows:

DEFINITION. – $\theta_n[A] = (I \wedge E) \to \exists e\, \forall x\, [A(x) \leftrightarrow \mathrm{True}_{\Sigma_n^0}(e, x)]$

PROOF OF THEOREM 5. – 1) Let f be a Σ_n^0-interpretation. So $A^f \in \Sigma_n^0$ and (by a previous remark) $R_f \in \Sigma_n^0$. Let $\hat{A}(x) = \exists y\, [A^f(y) \wedge R_f(x, y)]$. Then $\hat{A}$ is also Σ_n^0, so there is a natural number $e \in \omega$ such that $PA \vdash \forall x [\hat{A}(x) \leftrightarrow \mathrm{True}_{\Sigma_n^0}(e, x)]$. Let M be a model of PA. We have to show that $M \models \theta_n[A]^f$. So assume that $(I \wedge E)^f$ holds in M. Let $e' \in M$ be such that $M \models R_f(e, e')$. Since $M \models PA$, $M \models \forall x [\hat{A}(x) \leftrightarrow \mathrm{True}_{\Sigma_n^0}(e, x)]$. So by the results about Tennenbaum's theorem, $M \models \forall x [A(x) \leftrightarrow \mathrm{True}_{\Sigma_n^0}(e', x)]^f$, whence $M \models \theta_n[A]^f$ and we are done.

2) For the second part choose f so that f is standard on $L(PA)$ and A^f is a Π_n^0-formula which is not equivalent, in ω, to any Σ_n^0-formula (for example $A^f(x) = \neg \mathrm{True}_{\Sigma_n^0}(x, x)$). Since f is standard on $L(PA)$, $(I \wedge E)^f$ holds (in ω), but on the other hand $\exists e \forall x [A^f(x) \leftrightarrow \mathrm{True}_{\Sigma_n^0}(e, x)]$ is false. Thus $\omega \not\models \theta_n[A]^f$ (note: since f is standard on $L(PA)$, $\theta_n[A]^f = \theta_n[A^f]$). This completes the proof of the hierarchy theorem.

REMARK. – From the hierarchy theorem it also follows that the predicate modal theories $\{\theta | \text{for every } \Sigma_n^0\text{-interpretation } f\ \omega \models \theta^f\}$, are all distinct for different natural numbers $n \geq 1$.

5. – Π_n^0-interpretations.

In this section we extende our analysis to Π_n^0-interpretations. With a proof completely similar to the one that the Σ_n^0-logics of provability form an hierarchy, we can prove the analogous result for the Π_n^0-logics of provability, namely P_{n+1} strictly includes P_n (P_n and S_n are defined in the introduction). It is also easy to see that P_n and S_n are many-one reducible to each other (whence P_n is also Π_2^0-complete). To see this let $\tilde{\theta}$ be the result of replacing each atomic subformula of θ by its negation, then the map which sends θ to $\tilde{\theta}$ gives both reductions. Notice that for Π_n^0-interpretations f the relation R_f is not necessarily Π_n^0, so we cannot give a direct proof of the fact that P_n is Π_2^0-complete analogous to the one for S_n.

I want to thank Prof. Robert Solovay for many helpful discussions concerning the modal logic of provability.

REFERENCES

[Art85] S. N. ARTEMOV, *Nonarithmeticity of truth predicate logic of provability*, Soviet Math. Dokl., **32**, 2 (1985), 403-405 (Translated from: Dokl. Akad. Nauk, SSSR (1985)).

[BM87] G. BOOLOS - V. MCGEE, *The degree of the set of sentences of predicate provability logic that are true under every interpretation*, J. Symbolic Logic, **52**, 1 (1987), 165-171.

[Mon84] F. MONTAGNA, *The predicate modal logic of provability*, Notre Dame J. Formal Logic, **25**, 2 (1984), 179-189.

[Sol76] R. M. SOLOVAY, *Provability interpretations of modal logic*, Israel J. Math., **25** (1976), 287-304.

[Var85] V. A. VARDAYAN, *O predicatnoĭ logike dokazuemosti*, Nauchnyĭ Sovet po Komplexnoĭ Probleme Kibernetiki, Akademiia Nauk SSSR, Moscow, 1985 (in Russian).

Via Stibbert 33/A - 50134 Firenze

Pervenuta in Redazione
il 2 maggio 1988