Friedman's A-Translation

Peter Selinger

October 27, 1992

Theorem 1 *Peano arithmetic is a conservative extension of Heyting arithmetic for Π^0_2 sentences.*

1 Heyting and Peano arithmetic

Definition *Heyting arithmetic* (**HA**) and *Peano arithmetic* (**PA**) are formal systems based on the following language $\mathcal{L}$:

$\mathcal{L}$ is a first-order-language, with logical constants $\bot$, $\wedge$, $\vee$, $\to$, $\forall$, $\exists$, numerical variables x, y, $z\ldots$, a constant $\mathbf{0}$, a unary function constant $\mathbf{S}$, constant function symbols for all primitive recursive functions (indicated by F, G, H…) and a single binary predicate constant $=$. Terms and formulas are defined as usual. Formulas are indicated by Φ, $\Psi\ldots$ and $\neg\Phi$ abbreviates $\Phi \to \bot$.

The *axioms and rules* of **HA** (**PA**, respectively) are the axioms and rules of intuitionistic (respectively classical) first-order predicate logic (e.g. in a standard Hilbert-style formalization or one of several natural or sequent calculi) together with the following non-logical axioms:

$$x = x \qquad (\textit{refl})$$

$$x = y \wedge z = y \to x = z \qquad (\textit{trans})$$

$$x_i = x_i' \to \mathrm{F}(x_1, \ldots, x_i, \ldots, x_n) = \mathrm{F}(x_1, \ldots, x_i', \ldots, x_n) \qquad (\textit{cong}_\mathrm{F})$$

for any n-ary function constant F, $1 \le i \le n$,

$$\mathbf{S}x \neq \mathbf{0} \text{ (as abbreviation for } \mathbf{S}x = \mathbf{0} \to \bot) \qquad (\textit{succ1})$$

$$\mathbf{S}x = \mathbf{S}y \to x = y \qquad (\textit{succ2})$$

furthermore all instances of the axiom schema

$$\Phi\mathbf{0} \wedge \forall x(\Phi x \to \Phi(\mathbf{S}x)) \to \forall x \Phi x \qquad (\textit{ind})$$

as well as defining axioms for all primitive recursive functions. Every primitive recursive function F except the 0-ary $\mathbf{0}$ and the 1-ary $\mathbf{S}$ is defined by exactly one axiom of one of the following forms:

$$\mathrm{F}(x_1, \ldots, x_i, \ldots, x_n) = x_i \qquad (\textit{proj}_\mathrm{F})$$

$$\mathrm{F}(x_1,\ldots,x_n) = \mathrm{G}(\mathrm{H}_1(x_1,\ldots,x_n),\ldots,\mathrm{H}_m(x_1,\ldots,x_n)) \qquad (comp_{\mathrm{F}})$$

$$\begin{aligned} \mathrm{F}(0,x_1,\ldots,x_n) &= \mathrm{G}(x_1,\ldots,x_n) \\ \wedge\ \mathrm{F}(\mathbf{S}y,x_1,\ldots,x_n) &= \mathrm{H}(\mathrm{F}(y,x_1,\ldots,x_n),y,x_1,\ldots,x_n) \qquad (rec_{\mathrm{F}}) \end{aligned}$$

where G, H, H_1, ..., H_m have been defined before. [Tro73]

Note that **HA** and **PA** differ only in that **HA** uses intuitionistic, **PA** classical logic. We therefore have the immediate

Lemma 2 $\vdash_{\mathrm{HA}}\ \Phi \Rightarrow \vdash_{\mathrm{PA}}\ \Phi$ *for any formula* $\Phi \in \mathcal{L}$. □

The next lemma states that every quantifier-free formula is essentially of the form $\mathrm{F}(x_1,\ldots,x_n) = \mathbf{0}$, where F is a primitive recursive function symbol and $x_1,\ldots,x_n$ the (free) variables of the formula.

Lemma 3 *Let* Ψ *be any formula without quantifiers and with (free) variables* $x_1,\ldots,x_n$. *Then there is an* n*-ary primitive recursive function symbol* F *of* $\mathcal{L}$ *with* $\vdash_{\mathrm{HA}}\ \Psi \leftrightarrow \mathrm{F}(x_1,\ldots,x_n) = \mathbf{0}$. *(* $\leftrightarrow$ *is the usual abbreviation).*
Proof: Note first that the 2-ary functions + (addition), $\cdot$ (multiplication) and $\dot{-}$ (cut-off subtraction) are primitive recursive ([Tro88], p. 116), and that the following are provable: $\vdash_{\mathrm{HA}}\ x = \mathbf{0} \wedge y = \mathbf{0} \leftrightarrow x + y = \mathbf{0}$, $\vdash_{\mathrm{HA}}\ x = \mathbf{0} \vee y = \mathbf{0} \leftrightarrow x \cdot y = \mathbf{0}$, $\vdash_{\mathrm{HA}}\ x = \mathbf{0} \to y = \mathbf{0} \leftrightarrow (1 \dot{-} x) \dot{-} (1 \dot{-} y) = \mathbf{0}$, $\vdash_{\mathrm{HA}}\ \perp \leftrightarrow \mathbf{S}x = \mathbf{0}$. From this it should be clear how the proof goes by induction on the structure of Ψ. □

Definition $\Pi^0_2 \subset \mathcal{L}$ is the class of all formulas of the form

$$(\forall x_1)(\forall x_2)\ldots(\forall x_i)(\exists y_1)(\exists y_2)\ldots(\exists y_j)\Psi$$

where $i, j \geq 0$ and Ψ quantifier-free.

Lemma 4 *Every closed formula* $\Phi \in \Pi^0_2$ *is* **HA***-provably equivalent to*

$$(\forall x)(\exists y)\mathrm{F}(x, y) = \mathbf{0}$$

for some primitive recursive function symbol F.
Proof: Successive quantifiers of the same kind can be contracted by pairing; if additional quantifiers are necessary, "dummy" variables can be introduced. The existence of F follows from Lemma 3. □

2 Motivation of Theorem 1

Looking back at Theorem 1, it tells us that **PA** and **HA** have the same provable Π^0_2 formulas. Before we start proving this, let me try to motivate it a bit.

Observe that in **HA** (with a natural style system for the underlying logic) the following easy induction yields a proof for $(\forall x) x = \mathbf{0} \vee x \neq \mathbf{0}$:

$$
\dfrac{
\dfrac{\begin{array}{c}(\mathit{refl})\\ \mathbf{0}=\mathbf{0}\end{array}}{\mathbf{0}=\mathbf{0} \vee \mathbf{0}\neq\mathbf{0}}
\qquad
\dfrac{\dfrac{\dfrac{\begin{array}{c}(\mathit{succ1})\\ \mathbf{S}x\neq\mathbf{0}\end{array}}{\mathbf{S}x=\mathbf{0}\vee\mathbf{S}x\neq\mathbf{0}}}{x=\mathbf{0}\vee x\neq\mathbf{0}\to\mathbf{S}x=\mathbf{0}\vee\mathbf{S}x\neq\mathbf{0}}}{(\forall x)(x=\mathbf{0}\vee x\neq\mathbf{0}\to\mathbf{S}x=\mathbf{0}\vee\mathbf{S}x\neq\mathbf{0})}
}{(\forall x)x=\mathbf{0}\vee x\neq\mathbf{0}}\ (\mathit{ind})
$$

By application of the $\forall$-*elim* rule we can therefore get

$$\vdash_{\mathrm{HA}}\ \mathrm{F}(x_1,\ldots,x_n)=\mathbf{0}\vee\mathrm{F}(x_1,\ldots,x_n)\neq\mathbf{0}$$

for every n-ary primitive recursive function symbol F, and with lemma 3 for every quantifier-free formula Ψ:

$$\vdash_{\mathrm{HA}}\ \Psi\vee\neg\Psi.$$

We can therefore say that Heyting arithmetic has a certain amount of classical logic already "built in."

But don't be pleased too early. Our above discovery says nothing about formulas with quantifiers. For example, if $\Phi = (\forall x)\Psi$ with quantifier-free Ψ, then by the above method we can easily get $\vdash_{\mathrm{HA}}\ (\forall x)(\Psi\vee\neg\Psi)$, but this allows us *not* to conclude $\vdash_{\mathrm{HA}}\ \Phi\vee\neg\Phi$. Perhaps we can now more appreciate Theorem 1, that assures us that **HA** has classical logic "built in" even for (quantified) Π^0_2 formulas. This result is particularly nice because many statements of arithmetic can be expressed in Π^0_2 form. As an example we take the formula

$$(\forall x)(\exists y)(y\geq x\wedge\texttt{prime}(y)=\mathbf{0}),$$

where `prime` is the characteristic function of the prime numbers (`prime` and $\geq$ are primitive recursive, cf. [Tro88], p. 117). The formula states that there are infinitely many prime numbers. The best known proof for this fact is typically non-constructive, starting with the words "suppose not." However, once we have established Theorem 1 we get a constructive proof for the existence of infinitely many prime numbers for free.

The argument that we give here to prove Theorem 1 is due to H. Friedman [Fri78]. Other proofs were known earlier, but they were much more painful and required a delicate proof theoretic or semantic analysis, which we will not need. In the following we introduce two translations of formulas and some basic facts about them. The proofs are straightforward.

3 Double negation translation

Definition The *double negation translation* Φ° of some first-order formula Φ is defined by adding "$\neg\neg$" before every atomic, disjunctive or existential subformula:

$$\begin{array}{rcll}
\bot^\circ & = & \bot & \\
\Phi^\circ & = & \neg\neg\Phi & \text{where } \Phi \neq \bot \text{ atomic} \\
(\Phi \wedge \Psi)^\circ & = & \Phi^\circ \wedge \Psi^\circ & \\
(\Phi \vee \Psi)^\circ & = & \neg\neg(\Phi^\circ \vee \Psi^\circ) & \\
(\Phi \to \Psi)^\circ & = & \Phi^\circ \to \Psi^\circ & \\
(\forall x \Phi)^\circ & = & \forall x(\Phi^\circ) & \\
(\exists x \Phi)^\circ & = & \neg\neg\exists x(\Phi^\circ) &
\end{array}$$

Lemma 5 *Let* $\vdash_{\mathrm{C}}$ *stand for classical,* $\vdash_{\mathrm{I}}$ *for intuitionistic deducibility. The double negation translation has the following properties (*Φ *a formula,* Γ *a set of formulas, where* $\Gamma^\circ = \{\Psi^\circ | \Psi \in \Gamma\}$*):*

1. $\vdash_{\mathrm{C}} \Phi^\circ \leftrightarrow \Phi$
2. $\neg\neg\Phi^\circ \vdash_{\mathrm{I}} \Phi^\circ$
3. $\Gamma \vdash_{\mathrm{C}} \Phi \Rightarrow \Gamma^\circ \vdash_{\mathrm{I}} \Phi^\circ$
4. In general *not* $\Phi \vdash_{\mathrm{I}} \Phi^\circ$ □

In particular property 3 is interesting; it says that classical logic is embedded into (or reduced to) intuitionistic logic; therefore the term double negation "translation." Note that the converse of 3 trivially holds with 1. A counterexample for 4 is $\Phi = \neg\forall x \Psi x$.

4 A-translation

Definition Let A and Φ be formulas such that no bound variable of Φ is free in A. The *A-translation* Φ_A of some first-order formula Φ is defined by simultaneously replacing every atomic subformula Ψ by $\Psi \vee A$:

$$\begin{array}{rcll}
\bot_A & = & A & \\
\Phi_A & = & \Phi \vee A & \text{where } \Phi \neq \bot \text{ atomic} \\
(\Phi \wedge \Psi)_A & = & \Phi_A \wedge \Psi_A & \\
(\Phi \vee \Psi)_A & = & \Phi_A \vee \Psi_A & \\
(\Phi \to \Psi)_A & = & \Phi_A \to \Psi_A & \\
(\forall x \Phi)_A & = & \forall x(\Phi_A) & \\
(\exists x \Phi)_A & = & \exists x(\Phi_A) &
\end{array}$$

Here it is important that $\neg\Phi$ is only an abbreviation for $\Phi \to \bot$; note that the A-translation of $\neg\Phi$ is *not* $\neg\Phi_A$.

Lemma 6 *The A-translation has the following properties (Φ a formula and Γ a set of formulas, such that Φ_A and Γ_A are defined, where $\Gamma_A = \{\Psi_A | \Psi \in \Gamma\}$):*

1. $\vdash_{\mathrm{C}} \ \Phi_A \leftrightarrow \Phi \vee A$
2. $A \ \vdash_{\mathrm{I}} \ \Phi_A$
3. $\Gamma \ \vdash_{\mathrm{I}} \ \Phi \Rightarrow \Gamma_A \ \vdash_{\mathrm{I}} \ \Phi_A$
4. In general *not* $\Phi \ \vdash_{\mathrm{I}} \ \Phi_A$ □

The proof of property 3 is a bit tricky where eigenvariable conditions are involved. The rest is straightforward. Note that $\Phi \equiv \neg\neg A$ is a counterexample for 4.

5 Proof of Theorem 1

The proof goes in two steps. Given $\vdash_{\mathrm{PA}} \ (\exists y)\mathrm{F}(x,y) = \mathbf{0}$ we first conclude $\vdash_{\mathrm{HA}} \ \neg\neg(\exists y)\mathrm{F}(x,y) = \mathbf{0}$, using double negation translation, then $\vdash_{\mathrm{HA}} \ (\exists y)\mathrm{F}(x,y) = \mathbf{0}$, using A-translation. The proof will last on the following crucial properties of the axioms that we stated in Section 1:

Lemma 7 *For every non-logical axiom Ψ of Heyting/Peano arithmetic (including instances of axiom schemata) both translations Ψ° and Ψ_A are provable in* **HA***.*
Proof: Note that from property 4 in Lemmas 5 and 6 this is not true for a general formula Ψ. However, if Ψ is of the form Φ, $\Phi_1 \wedge \Phi_2$, $\Phi_1 \to \Phi_2$ or $\Phi_1 \wedge \Phi_2 \to \Phi_3$ (where Φ, Φ_1, Φ_2, Φ_3 atomic), then we can easily show $\Psi \ \vdash_{\mathrm{I}} \ \Psi^\circ$ and $\Psi \ \vdash_{\mathrm{I}} \ \Psi_A$. All axioms except *(ind)* are of this form. Suppose now Ψ is an instance of *(ind)*:

$$\Psi \equiv \Phi\mathbf{0} \wedge \forall x(\Phi x \to \Phi(\mathbf{S}x)) \to \forall x \Phi x$$

for some formula Φx. Then

$$\Psi^\circ \equiv \Phi^\circ\mathbf{0} \wedge \forall x(\Phi^\circ x \to \Phi^\circ(\mathbf{S}x)) \to \forall x \Phi^\circ x,$$

$$\Psi_A \equiv \Phi_A\mathbf{0} \wedge \forall x(\Phi_A x \to \Phi_A(\mathbf{S}x)) \to \forall x \Phi_A x,$$

which are themselves axioms of **HA**. □

Corollary 8 *The following hold for all formulas $\Phi \in \mathcal{L}$:*

1. $\vdash_{\mathrm{PA}} \ \Phi \Rightarrow \vdash_{\mathrm{HA}} \ \Phi^\circ$
2. $\vdash_{\mathrm{HA}} \ \Phi \Rightarrow \vdash_{\mathrm{HA}} \ \Phi_A$, if Φ_A defined.

Proof: 1. If Γ are the non-logical axioms of **PA** used in the proof of $\vdash_{\mathrm{PA}} \Phi$, then

$$\Gamma \vdash_{\mathrm{C}} \Phi \overset{\textit{Lemma } 5.3}{\Longrightarrow} \Gamma^{\circ} \vdash_{\mathrm{I}} \Phi^{\circ} \overset{\textit{Lemma } 7}{\Longrightarrow} \vdash_{\mathrm{HA}} \Phi^{\circ}.$$

2. If Γ are the non-logical axioms of **HA** used in the proof of $\vdash_{\mathrm{HA}} \Phi$, then

$$\Gamma \vdash_{\mathrm{I}} \Phi \overset{\textit{Lemma } 6.3}{\Longrightarrow} \Gamma_A \vdash_{\mathrm{I}} \Phi_A \overset{\textit{Lemma } 7}{\Longrightarrow} \vdash_{\mathrm{HA}} \Phi_A.$$

□

Proof of Theorem 1: If $\vdash_{\mathrm{PA}} (\exists y)\mathrm{F}(x,y) = \mathbf{0}$ then $\vdash_{\mathrm{HA}} \neg\neg(\exists y)\mathrm{F}(x,y) = \mathbf{0}$ by the Corollary. Having $\vdash_{\mathrm{HA}} (((\exists y)\mathrm{F}(x,y) = \mathbf{0}) \to \bot) \to \bot$, using $A \equiv (\exists y)\mathrm{F}(x,y) = \mathbf{0}$, we have

$$\vdash_{\mathrm{HA}} \Big(((\exists y)\mathrm{F}(x,y) = \mathbf{0}) \vee ((\exists y)\mathrm{F}(x,y) = \mathbf{0}) \to ((\exists y)\mathrm{F}(x,y) = \mathbf{0}) \Big) \\ \to ((\exists y)\mathrm{F}(x,y) = \mathbf{0}),$$

hence $\vdash_{\mathrm{HA}} (\exists y)\mathrm{F}(x,y) = \mathbf{0}$. □

Note that this argument can easily be applied to theories other than **HA**, as long as their axioms satisfy Lemma 7. Friedman does this in his paper [Fri78] for the theory of finite types and for Zermelo-Fraenkel set theory. A further development of Friedman's methods is found in [Lev85], where in particular large classes of axioms satisfying Lemma 7 are described syntactically.

References

[Fri78] H.M. Friedman: Classically and Intuitionistically Provably Recursive Functions, in: *Mueller and Scott (eds.): Higher Set Theory (1978) 21-27*, Springer, Berlin

[Lev85] D. Leivant: Syntactic Translations and Provably Recursive Functions, in: *The Journal of Symbolic Logic, vol. 50 (1985) 682-688*

[Tro73] A.S. Troelstra: Metamathematical Investigation of Intuitionistic Arithmetic and Analysis, in: *Lecture Notes in Mathematics, vol. 344 (1973)*, Springer, Berlin

[Tro88] A.S. Troelstra, D. van Dalen: *Constructivism in Mathematics, vol. I (1988)*, North Holland