Calibrating the Negative Interpretation

Joan Rand Moschovakis

Occidental College, Emerita

12th Panhellenic Logic Symposium, Anogeia, Crete
June 26-30, 2019

This talk is in memory of Anne S. Troelstra, 1939-2019, who contributed significantly to logic in Greece. He lectured for $\mu\Pi\lambda$A on the intuitionism of Brouwer and Heyting, and on proofs as types. After retiring from the University of Amsterdam he donated much of his personal logical library to the University of Athens and continued visiting Crete for the wildflowers.

Elementary Facts: In a language with all the usual logical connectives and quantifiers, Hilbert-style classical predicate logic can be formulated so that intuitionistic predicate logic has the same rules, and the same axioms except that $\neg A \to (A \to B)$ replaces the stronger, classical $\neg\neg A \to A$.

Gödel and Gentzen independently proved that classical predicate logic can be faithfully interpreted in the *negative fragment* of its intuitionistic subsystem (involving only $\&, \neg, \to$ and $\forall$), e.g. by

1. replacing predicate letters with their double negations, and
2. hereditarily replacing $A \vee B$ by $\neg(\neg A \;\&\; \neg B)$, and $\exists x A(x)$ by $\neg\forall x \neg A(x)$.

Hence: To prove that a classical theory **T** is equiconsistent with its intuitionistic subtheory **S**, it is enough to show that **S** *proves the negative interpretations of the mathematical axioms of* **T**.

Classical arithmetic **PA** and intuitionistic arithmetic **HA**, with $=$, $0, ', +, \cdot$ and full mathematical induction, satisfy this condition.

The **Gödel-Gentzen negative interpretation** E^g of a formula E of the language of arithmetic is defined inductively:

- Prime formulas are unchanged: $(\mathrm{s} = \mathrm{t})^g \equiv (\mathrm{s} = \mathrm{t})$. (This is possible because $\vdash_{\mathbf{HA}} \neg\neg(\mathrm{s} = \mathrm{t}) \leftrightarrow (\mathrm{s} = \mathrm{t})$.)
- Negative connectives pass through: $(\forall \mathrm{x} \mathrm{A}(\mathrm{x}))^g = \forall \mathrm{x} \mathrm{A}^g(\mathrm{x})$, $(\mathrm{A} \ \& \ \mathrm{B})^g \equiv (\mathrm{A}^g \ \& \ \mathrm{B}^g)$ and $(\mathrm{A} \to \mathrm{B})^g \equiv (\mathrm{A}^g \to \mathrm{B}^g)$.
- Disjunction $\vee$ and existence $\exists$ are interpreted classically: $(\mathrm{A} \vee \mathrm{B})^g \equiv \neg(\neg \mathrm{A}^g \& \neg \mathrm{B}^g)$ and $(\exists \mathrm{x} \mathrm{A}(\mathrm{x}))^{\mathrm{g}} \equiv \neg \forall \mathrm{x} \neg \mathrm{A}^{\mathrm{g}}(\mathrm{x})$.

Theorem 1. (Gödel) **PA** and **HA** are equiconsistent.

Proof: For every arithmetical formula E:

- $\vdash_{\mathbf{PA}} (\mathrm{E} \leftrightarrow \mathrm{E}^g)$.
- $\vdash_{\mathbf{PA}} \mathrm{E}$ if and only if $\vdash_{\mathbf{HA}} \mathrm{E}^g$.

Remarks:

- The negative interpretation is easily extended to a language for analysis, with variables $\alpha, \beta, \ldots$ over infinite sequences of natural numbers. Set $(\exists \alpha \mathrm{B}(\alpha))^g \equiv \neg\forall\alpha\neg\mathrm{B}^g(\alpha)$.
- The *neutral* (classically and intuitionistically correct) basic subsystem **B** of Kleene's formal system **I** for intuitionistic analysis has mathematical axioms (countable choice and bar induction) whose negative interpretations are unprovable in **B**.
- The negative interpretation of Brouwer's continuity principle (the axiom separating **I** from **B**) is *refutable* in **B** *and* in **I**.

Question: What must be added to a subsystem **S** of Kleene's formal system **I** of intuitionistic analysis, in order to prove the negative interpretations of the classically correct axioms of **S**?

Let $\mathbf{S}^{+g}$ be the *minimum classical extension of* **S** in this sense, and let $\mathbf{S}^g$ be the *negative fragment* of $\mathbf{S}^{+g}$.

Theorem 2. If **S** $\subseteq$ **B**, then

(a) $\mathbf{S}^{+g}$ and $\mathbf{S}^g$ and $\mathbf{S} + (\neg\neg \mathrm{A} \rightarrow \mathrm{A})$ are equiconsistent, and have exactly the same classical ω-models as **S**.

(b) $\mathbf{S}^{+g}$ is consistent with Kleene's intuitionistic analysis **I**.

Proofs: (a): If $\mathbf{S} \subseteq \mathbf{B}$ then $\mathbf{S}^g \subseteq \mathbf{S}^{+g} \subseteq \mathbf{S} + (\neg\neg \mathrm{A} \rightarrow \mathrm{A})$, and for every formula E of the language of analysis:

- $\mathbf{S} + (\neg\neg \mathrm{A} \rightarrow \mathrm{A}) \vdash \mathrm{E} \leftrightarrow \mathrm{E}^g$.
- $\mathbf{S} + (\neg\neg \mathrm{A} \rightarrow \mathrm{A}) \vdash \mathrm{E}$ if and only if $\mathbf{S}^{+g} \vdash \mathrm{E}^g$, which happens if and only if $\mathbf{S}^g \vdash \mathrm{E}^g$ *using only negative rules and axioms*.

(b) holds because all the axioms of **I**, and all classically correct negative formulas, are Kleene function-realizable; therefore so is every theorem of $\mathbf{I} + \mathbf{S}^{+g}$, but $0 = 1$ is not.

Challenge: Clarify the classical vs. the intuitionistic mathematical content of a given subsystem **S** of Kleene's neutral analysis **B**, by finding a *nice* characterization of $\mathbf{S}^{+g}$ (which is consistent with **I**).

Mathematical axioms of B:

- $=$ is an equivalence relation.
- 0 is not a successor, and $'$ is one-to-one.
- $x = y \rightarrow \alpha(x) = \alpha(y)$.
- Primitive recursive defining equations for finitely many function constants, including the characteristic function of Kleene's T-predicate and the result-extracting function U.
- Mathematical induction: $A(0) \,\&\, \forall x(A(x) \rightarrow A(x')) \rightarrow A(x)$.
- λ-reduction: $(\lambda x.r(x))(t) = r(t)$ for terms $r(x), t$.
- Countable choice (^{x}2.1 in Kleene-Vesley 1965):

  $$\mathsf{AC}_{01} : \forall x \exists \alpha A(x, \alpha) \rightarrow \exists \beta \forall x A(x, \lambda y.\beta(\langle x, y\rangle)).$$
- The "bar theorem" (^{x}26.3b in Kleene-Vesley 1965):

  $$\mathsf{BI}_1 : \forall \alpha \exists x \rho(\overline{\alpha}(x)) = 0 \ \&\ \forall w(\mathrm{Seq}(w) \ \&\ \rho(w) = 0 \rightarrow A(w)) \ \&\ \forall w(\mathrm{Seq}(w) \ \&\ \forall s A(w * \langle s+1 \rangle) \rightarrow A(w)) \rightarrow A(\langle\,\rangle).$$

The logic of **B** *is intuitionistic.* [Let **C** $\equiv$ **B** $+$ $(\neg\neg A \rightarrow A)$.]

Two weak but useful subsystems of B:

Two-sorted intuitionistic arithmetic $\mathbf{IA_1}$ is the fragment of Kleene's basic system **B** obtained by omitting the axioms of countable choice and bar induction. There is full mathematical induction, but no comprehension or choice. The primitive recursive functions form a classical ω-model of $\mathbf{IA_1}$. It is easy to show $\mathbf{(IA_1)}^{+g} = \mathbf{IA_1}$.

Intuitionistic recursive analysis **IRA** adds to $\mathbf{IA_1}$ the axiom

$$\text{qf-AC}_{00} : \forall \mathrm{x} \exists \mathrm{y} \rho(\langle \mathrm{x}, \mathrm{y} \rangle) = 0 \to \exists \alpha \forall \mathrm{x} \rho(\langle \mathrm{x}, \alpha(\mathrm{x}) \rangle) = 0$$

of quantifier-free countable choice, which guarantees that the class of functions is closed under "recursive in." The general recursive functions form the smallest classical ω-model, but $\mathbf{IRA}^{+g} \neq \mathbf{IRA}$.

Note: Troelstra's **EL** (Troelstra 1973 and Troelstra and van Dalen 1988) has a constant and axioms for every primitive recursive function but otherwise is like **IRA**, with full induction and qf-AC_{00}. Vafeiadou 2012 gives the precise comparison, and many others.

Axioms stronger than qf-AC_{00} but weaker than AC_{01}:

Countable comprehension ("unique choice") is

$$\mathsf{AC}_{00}! : \forall x \exists ! y A(x, y) \to \exists \alpha \forall x A(x, \alpha(x)),$$

where $\exists ! y B(y)$ abbreviates $\exists y B(y)$ & $\forall y \forall z (B(y) \& B(z) \to y = z)$.

Countable choice for numbers is

$$\mathsf{AC}_{00} : \forall x \exists y A(x, y) \to \exists \alpha \forall x A(x, \alpha(x)).$$

Arithmetical countable choice AC_{00}^{Ar} restricts this to arithmetical A.

Lemma. **IRA** $\subsetneq$ **IA**$_1$ + $\mathsf{AC}_{00}!$ = **IA**$_1$ + $\mathsf{AC}_{01}!$ $\subsetneq$ **IA**$_1$ + AC_{00} where $\mathsf{AC}_{01}!$ is a "unique" version of AC_{01}. **IRA** $\subsetneq$ **IA**$_1$ + AC_{00}^{Ar}.

Weaker than the bar theorem is the binary fan theorem:

FT_1. $\forall \alpha_{B(\alpha)} \exists x\, \rho(\overline{\alpha}(x)) = 0 \to \exists n \forall \alpha_{B(\alpha)} \exists x \leq n\, \rho(\overline{\alpha}(x)) = 0$ (where $B(\alpha) \equiv \forall y\, \alpha(y) \leq 1$). The arithmetical functions form a classical ω-model of **IRA** + FT_1 but not of **B**. **IRA** + FT_1 proves "pointwise continuous functions on [0,1] are uniformly continuous".

The form MP_1: $\forall\alpha(\neg\neg\exists \mathrm{x}\,\alpha(\mathrm{x}) = 0 \to \exists \mathrm{x}\,\alpha(\mathrm{x}) = 0)$ of *Markov's Principle* was rejected by Brouwer but is consistent with **I** (Kleene). Consequences of MP_1 consistent with **I** + ¬ MP_1 include the **double negation shift principles**

DNS_1. $\forall\rho[\forall\alpha\neg\neg\exists\mathrm{x}\rho(\overline{\alpha}(\mathrm{x})) = 0 \to \neg\neg\forall\alpha\exists\mathrm{x}\rho(\overline{\alpha}(\mathrm{x})) = 0]$,

Σ^0_1-DNS_0. $\forall\alpha[\forall\mathrm{x}\neg\neg\exists\mathrm{y}\alpha(\langle\mathrm{x},\mathrm{y}\rangle) = 0 \to \neg\neg\forall\mathrm{x}\exists\mathrm{y}\alpha(\langle\mathrm{x},\mathrm{y}\rangle) = 0]$,

and the *Gödel-Dyson-Kreisel Principle*, which is equivalent over **IRA** to the weak completeness of intuitionistic predicate logic:

GDK. $\forall\rho[\forall\alpha_{\mathrm{B}(\alpha)}\neg\neg\exists\mathrm{x}\rho(\overline{\alpha}(\mathrm{x})) = 0 \to \neg\neg\forall\alpha_{\mathrm{B}(\alpha)}\exists\mathrm{x}\rho(\overline{\alpha}(\mathrm{x})) = 0]$.

Lemma (Scedrov-Vesley) **IRA** + $\mathsf{DNS}_1 \vdash \Sigma^0_1$-$\mathsf{DNS}_0$ & GDK.

Theorem 3. (**S** with at most qf-AC_{00}, but perhaps FT_1 or BI_1)

(a) $\mathbf{IRA}^{+g}$ = **IRA** + Σ^0_1-DNS_0.

(b) $(\mathbf{IA}_1 + \mathsf{FT}_1)^{+g} = \mathbf{IA}_1 + \mathsf{FT}_1$ + GDK.

(c) $(\mathbf{IRA} + \mathsf{FT}_1)^{+g}$ = **IRA** + FT_1 + Σ^0_1-DNS_0 + GDK.

(d) $(\mathbf{IRA} + \mathsf{BI}_1)^{+g} \subseteq$ **IRA** + BI_1 + DNS_1.

Theorem 4. (Vafeiadou) $AC_{00}!$ is equivalent over **IRA** to the *characteristic function principle for decidable* $A(x)$:

$$\mathsf{CF}_d.\ \forall x(A(x) \vee \neg A(x)) \rightarrow \exists \chi_{B(\chi)} \forall x(\chi(x) = 0 \leftrightarrow A(x)).$$

Weak characteristic function principles, of the form

$$\mathsf{WCF}_0.\ \neg\neg\exists\chi\forall x(\chi(x) = 0 \leftrightarrow A(x)),$$

assert only that it is *consistent* to assume that $A(x)$ has a characteristic function. Three useful special cases are

- Π^0_1-WCF_0. $\forall\alpha[\neg\neg\exists\chi\forall x(\chi(x) = 0 \leftrightarrow \forall y\alpha(\langle x, y\rangle) = 0)]$.
- WCF_0^{Ar} (the restriction of WCF_0 to *arithmetical* $A(x)$).
- WCF_0^{-} (the restriction of WCF_0 to *negative* $A(x)$).

Theorem 5. (**S** satisfying **IRA** $\subsetneq$ **S** $\subsetneq$ $\mathbf{IA}_1 + AC_{01}$)

(a) $(\mathbf{IA}_1 + AC^{Ar}_{00})^{+g} = \mathbf{IA}_1 + AC^{Ar}_{00} + \Sigma^0_1\text{-}DNS_0 + \Pi^0_1\text{-}WCF_0$.

(b) $(\mathbf{IA}_1 + AC_{00}!)^{+g} = \mathbf{IA}_1 + AC_{00}! + \Sigma^0_1\text{-}DNS_0 + WCF^{-}_0$.

(c) $(\mathbf{IA}_1 + AC_{00})^{+g} = \mathbf{IA}_1 + AC_{00} + \Sigma^0_1\text{-}DNS_0 + WCF^{-}_0$.

Theorem 6. (Solovay)

(a) $\mathbf{IRA} + \mathsf{BI}_1 + \mathsf{MP}_1 \vdash \Sigma^0_1\text{-}\mathsf{WCF}_0$, hence

(b) $\mathbf{IRA} + \mathsf{BI}_1 + \mathsf{MP}_1 \vdash \mathsf{WCF}_0^{Ar}$, hence

(c) $\mathbf{IA}_1 + \mathsf{AC}_{00}^{Ar} + \mathsf{BI}_1 + (\neg\neg \mathrm{A} \to \mathrm{A})$ is negatively interpretable in $\mathbf{IRA} + \mathsf{BI}_1 + \mathsf{MP}_1$.

Analysis of Solovay's clever proof of (a) shows that MP_1 can be replaced in (c) by the intuitionistically more acceptable DNS_1.

Theorem 7.

(a) $\mathbf{IRA} + (\mathsf{BI}_1)^g \vdash \Pi^0_1\text{-}\mathsf{WCF}_0$, hence

(b) $\mathbf{IRA} + \mathsf{BI}_1 + \mathsf{DNS}_1$ proves the weak characteristic function principle for all *negative* arithmetical formulas, hence

(c) $(\mathbf{IA}_1 + \mathsf{AC}_{00}^{Ar} + \mathsf{BI}_1)^g \subseteq \mathbf{IRA} + \mathsf{BI}_1 + \mathsf{DNS}_1$.

Open question: Precisely characterize $(\mathbf{IRA} + \mathsf{BI}_1)^{+g}$, $\mathbf{B}^{+g}$ and intermediate systems with BI_1 such as $(\mathbf{IA}_1 + \mathsf{AC}_{00} + \mathsf{BI}_1)^{+g}$.