Gödel's Dialectica Interpretation

Klaus Frovin Jørgensen

Section for Philosophy and Science Studies, RUC

May 6, 2010

Question to be Answered

Kurt Gödel devised (around 1941) an interpretation of intuitionistic arithmetic into a calculus of functionals.

Together with Gödel's $\neg\neg-$translation (1933) this was also a consistency proof classical arithmetic.

What are the philosophical conclusions?

The Context of the Interpretation

The 1920s and 1930s:

- The *Grundlagestreit*
- Hilbert's programme (1920s)
- The development of intuitionistic logic (Heyting, Kolmogorov $\sim$ 1930)
- The incompleteness theorems (1931) and the Gödel/Gentzen $\neg\neg$-translation (1933)
- Church's and Turing's negative answer to the *Entscheidungsproblem* (1935-6).
- Gentzen's proof of consistency of formal number theory (1936)

Heyting's Proof Interpretation
$p : A$

Some of the clauses of Heyting's interpretation are:

($\rightarrow$) $p : A \rightarrow B$ iff p is a construction taking any q such that $q : A$ into $p(q)$ such that $p(q) : B$.

($\vee$) $p : A \vee B$ iff p is a pair (p_0, p_1), $p_0 \in \{0, 1\}$ and $p_1 : A$ if $p_0 = 0$ and $p_1 : B$ if $p_0 = 1$. $q : A$ into $p(q)$ such that $p(q) : \bot$.

($\forall$) $p : \forall x A(x)$ iff p is a construction taking any t from the intended domain into $p(t)$ such that $p(t) : A(t)$.

($\exists$) $p : \exists x A(x)$ iff p is a pair (p_0, p_1), where p_0 is an object of the domain and $p_1 : A(p_0)$.

Intuitionistic Propositional Logic

The following is Spector's (1962) formulation:

$$A \to A, \qquad \bot \to A,$$

$$A \to A \lor B, \qquad B \to A \lor B,$$

$$A \land B \to A, \qquad A \land B \to B,$$

$$\frac{A \quad A \to B}{B} \qquad \frac{A \to B \quad B \to C}{A \to C} \qquad \frac{A \to B \quad A \to C}{A \to B \land C}$$

$$\frac{A \land B \to C}{A \to (B \to C)} \qquad \frac{A \to (B \to C)}{A \land B \to C} \qquad \frac{A \to C \quad B \to C}{A \lor B \to C}$$

Tertium Non Datur

Tertium non datur:

$$A \vee \neg A$$

is not sound under the proof interpretation.

Intuitionistic Quantifier Rules

$$\frac{B \to A(b)}{B \to \forall x A(x)} \qquad \forall x A(x) \to A(t),$$

$$A(t) \to \exists x A(x), \qquad \frac{A(b) \to B}{\exists x A(x) \to B}$$

Here b is eigenvariable meaning that b is not allowed to occur free in B.

Disjunction Property and Existence Property

- *Existence property:* If $\mathrm{S} \vdash \exists x A(x)$ then $\mathrm{S} \vdash A(t)$ for a certain term t.
- *Disjunction property*: If $\mathrm{S} \vdash A \vee B$, for $A,\ B$ closed then $\mathrm{S} \vdash A$ or $\mathrm{S} \vdash B$.

Impredicative Methods (1/2)

A definition is impredicative if it refers to a collection which contains the object to be defined. If one sees definitions as somehow creating or constructing, then circularity is involved. For example the 'least upper bound' of a set is defined to be the 'smallest among the upper bounds'. This is seen, for instance, in the completeness axiom:

Any non-empty subset of real numbers bounded above has a least upper bound.

Example: There exists a real number x such that $x^2 = 2$.

Another example is the Russell set R.

Impredicative Methods (2/2)

Yet an example is the intuitionistic meaning of implication: The proof interpretation). We know $A \longrightarrow B$ precisely when we know what may count as a proof of $A \longrightarrow B$. A proof of $A \longrightarrow B$ transforms any proof of A into a proof of B.

The System Σ (1/3)

The ground type consits of the natural numbers. There are symbols for zero and successor and variables of all types.
If F is an operation of type $\sigma \to \tau$ this is written as $F^{\sigma \to \tau}$
Operations in T are defined from combinators (which introduce λ-abstraction)

$$\mathrm{K}(x, y) = x \qquad \mathrm{S}(x, y, z) = x(z)(yz)$$

The combinators give us λ-abstraction $\lambda x.t$ for terms t, with the following equality:

$$(\lambda x.t[x])s = t[s].$$

Moreover, we have primitive recursion

$$\begin{aligned} \mathrm{R}(x, y, 0) &= x \\ \mathrm{R}(x, y, (z+1)) &= y(\mathrm{R}(x, y, z),\, z) \end{aligned}$$

The System Σ (2/3)

Quantifier free induction

$$\frac{A(0) \quad A(x^0) \to A(Sx^0)}{A(x^0)}$$

Substitution:

$$\frac{A(x^\sigma)}{A(t^\sigma)}$$

The System Σ (3/3)

We can do elementary arithmetic i Σ:

$$\begin{aligned}
x + y &:\equiv \mathsf{R}_0 x(\lambda w, u.Sw)y \\
\mathsf{prd} &:\equiv \lambda x.\mathsf{R}_0 0(\lambda w, u.u)x \\
x \dot{-} y &:\equiv \mathsf{R}_0 x(\lambda w, u.\mathsf{prd}(w))y \\
|x - y| &:\equiv (x \dot{-} y) + (y \dot{-} x) \\
\mathrm{Cond} &:\equiv \lambda x^\sigma, y^\sigma, z^0.\mathrm{R}_\sigma x(\lambda v^\sigma, w^0.y)z \\
\mathsf{max} &:\equiv \lambda x^0, y^0.\mathrm{Cond}(y, x, x \dot{-} y) \\
\mathsf{min} &:\equiv \lambda x^0, y^0.\mathrm{Cond}(x, y, x \dot{-} y)
\end{aligned}$$

The prime formulas of Σ are decidable (proved by induction on the complexity of the terms). Therefore, Σ actually has classical logic which can, moreover, be represented in the system.

Definition of D-translation (1/2)

To each formula A of $\mathcal{L}(\mathrm{HA})$ is now associated its Dialectica translation A^D which is a formula of Σ.

$$A^D \equiv \exists \boldsymbol{x} \forall \boldsymbol{y} A_D(\boldsymbol{x}, \boldsymbol{y}),$$

where A_D is quantifier free. Intuitively: If A is provable then according to the translation A^D there are $\boldsymbol{x}$ making A_D 'true' for any $\boldsymbol{y}$.

The lengths and types of the fresh variables $\boldsymbol{x}$ and $\boldsymbol{y}$ depend only on the logical structure of A.

Definition of D-translation (2/2)

Let A^D have the form $\exists \boldsymbol{x} \forall \boldsymbol{y} A_D(\boldsymbol{x}, \boldsymbol{y})$ and B^D the form $\exists \boldsymbol{u} \forall \boldsymbol{v} B_D(\boldsymbol{u}, \boldsymbol{v})$.

Definition.

$$
\begin{array}{rcl}
A^D & :\equiv & A_D :\equiv A, \text{ if } A \text{ is prime,} \\
(A \wedge B)^D & :\equiv & \exists \boldsymbol{x}, \boldsymbol{u} \forall \boldsymbol{y}, \boldsymbol{v} (A_D(\boldsymbol{x}, \boldsymbol{y}) \wedge B_D(\boldsymbol{u}, \boldsymbol{v})), \\
(A \vee B)^D & :\equiv & \exists z^0, \boldsymbol{x}, \boldsymbol{u} \forall \boldsymbol{y}, \boldsymbol{v} \big((z = 0 \to A_D(\boldsymbol{x}, \boldsymbol{y})) \wedge \\
 & & \qquad (z \neq 0 \to B_D(\boldsymbol{u}, \boldsymbol{v})) \big), \\
(\exists z A(z))^D & :\equiv & \exists z, \boldsymbol{x} \forall \boldsymbol{y} A_D(\boldsymbol{x}, \boldsymbol{y}, z), \\
(\forall z A(z))^D & :\equiv & \exists \boldsymbol{X} \forall \boldsymbol{y}, z A_D(\boldsymbol{X} z, \boldsymbol{y}, z), \\
(A \to B)^D & :\equiv & \exists \boldsymbol{U}, \boldsymbol{Y} \forall \boldsymbol{x}, \boldsymbol{v} \big(A_D(\boldsymbol{x}, \boldsymbol{Y} \boldsymbol{x} \boldsymbol{v}) \to B_D(\boldsymbol{U} \boldsymbol{x}, \boldsymbol{v}) \big).
\end{array}
$$

The Translation of '$\longrightarrow$'

Assume $\exists x \forall y A_D(x, y) \to \exists u \forall v B_D(u, v)$. A reasonable reading is:

$$\exists U \forall x (\forall y A_D(x, y) \to \forall v B_D(Ux, v)). \tag{1}$$

What could a possible interpretation of $\forall x C(x) \to \forall y D(y)$ be? Given any counter-example to D we can construct a counter-example to C, i.e. $\exists X \forall y (\neg D(y) \to \neg C(Xy))$. This implies:

$$\exists U \forall x \exists Y' \forall v (\neg B_D(Ux, v) \to \neg A_D(x, Y'v)). \tag{2}$$

The quantifier free formulas are stable. Therefore, 'by' the proof interpretation we get:

$$\exists U, Y \forall x, v (A_D(x, Yxv) \to B_D(Ux, v)).$$

On 'Constructive Meaning'

The following principles are not constructively valid *according* to the proof interpretation by Heyting.

- Different forms of independence-of-premise:

$$(A \to \exists y B(y)) \to \exists y (A \to B(y)),$$

  where $y \notin \mathrm{FV}(A)$ and different restrictions on A.
- Markov's principle: $\neg\neg\exists x A_{\mathrm{qf}}(x) \to \exists x A_{\mathrm{qf}}(x)$.

Gödel's translation changes the intuitionistic meaning:

$$\mathrm{HA}^\omega + \mathrm{IP}^\omega_\forall + \mathrm{MP}^\omega + \mathrm{AC} \vdash A \leftrightarrow A^D.$$

This is not necessarily a bad thing.

Soundness

Theorem [Gödel 1941].

$$\text{If } \mathrm{HA} \vdash A, \text{ then } \Sigma \vdash A_D(\boldsymbol{T}, \boldsymbol{y}),$$

where $\boldsymbol{T}$ is a sequence of terms which can be extracted from a proof of A in HA.

Examples from the Proof of Soundness (1/2)

The proof is by induction on the length of the proof of A in HA.

Case 1. Axiom $A \to A$. This translates to

$$\exists \boldsymbol{X}, \boldsymbol{Y} \forall \boldsymbol{x}, \boldsymbol{y}(A_D(\boldsymbol{x}, \boldsymbol{Y}\boldsymbol{x}\boldsymbol{y}) \to A_D(\boldsymbol{X}\boldsymbol{x}, \boldsymbol{y})).$$

From this we see that with $\boldsymbol{T}_1 :\equiv \lambda \boldsymbol{x}.\boldsymbol{x}$ and $\boldsymbol{T}_2 :\equiv \lambda \boldsymbol{x}, \boldsymbol{y}.\boldsymbol{y}$ we have

$$\Sigma \vdash A_D(\boldsymbol{x}, \boldsymbol{T}_2\boldsymbol{x}\boldsymbol{y}) \to A_D(\boldsymbol{T}_1\boldsymbol{x}, \boldsymbol{y}).$$

Examples from the Proof of Soundness (2/2)

Case 2. Modus Ponens. Assume as induction hypothesis

$$\begin{array}{ll} \text{(i)} & \Sigma \vdash A_D(\boldsymbol{T}_1, \boldsymbol{y}), \\ \text{(ii)} & \Sigma \vdash A_D(\boldsymbol{x}, \boldsymbol{T}_2 \boldsymbol{x} \boldsymbol{v}) \to B_D(\boldsymbol{T}_3 \boldsymbol{x}, \boldsymbol{v}), \end{array}$$

for given $\boldsymbol{T}_1$, $\boldsymbol{T}_2$, and $\boldsymbol{T}_3$. Find $\boldsymbol{T}_4$ such that $\Sigma \vdash B_D(\boldsymbol{T}_4, \boldsymbol{v})$. Set $\boldsymbol{x}$ in (ii) to $\boldsymbol{T}_1$ and let $\boldsymbol{y}$ in (i) be $\boldsymbol{T}_2 \boldsymbol{T}_1 \boldsymbol{v}$. Then use MP (in Σ) to obtain

$$\Sigma \vdash B_D(\boldsymbol{T}_3 \boldsymbol{T}_1, \boldsymbol{v}).$$

Let $\boldsymbol{T}_4$ be equal to $\boldsymbol{T}_3 \boldsymbol{T}_1$.

Note the similarities to the proof interpretation.

Gödel's Results

In 1941 the following results are mentioned:

1. For a certain quantifier free formula $A(x)$ let $C \equiv \neg\forall x(A(x) \vee \neg A(x))$. Then $\mathrm{HA} + C$ is consistent.
2. If HA proves $\exists x A(x)$ then Σ proves the *translated* formula $A_D(t)$, for a term t.
3. $\neg\neg$-translation (1933) together with the new interpretation proves consistency of classical arithmetic relative to T.

The interpretation was ultimately published in *Dialectica* in 1958: "Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes".

Philosophy (1/3)

Weak versus Strong Counterexamples (1/3)

π is transcendental, but we can approximate π arbitrarily well:

$$\pi = \frac{4}{1} - \frac{4}{3} + \frac{4}{5} - \frac{4}{7} + \frac{4}{9} - \frac{4}{11} + \cdots$$

Let A be the statement: "There exists one hundred 9s in a row in the decimal expansion of π."

Weak versus Strong Counterexamples (2/3)

1. Start the *computation* of π.
2. We start writing a real number a. The first digit is 0 followed by a point:

$$0.$$

3. Construction of the n-th digit of a:
   - If the decimal expansion of π up to digit number $n-1$ has not verified A, then the n-th digit of a is 0,
   - Otherwise 1

Is $a = 0$ or $a \neq 0$?

Weak versus Strong Counterexamples (3/3)

Gödel produces with his interpretation a *strong* counterexample. $\neg\forall x(A(x) \vee \neg A(x))$ is demonstrably incompatible with classical logic.

Constructivity is Understood Locally (1/3)

The following principles are not constructively valid *according* to the proof interpretation by Heyting.

- Different forms of independence-of-premise:

$$(A \to \exists y B(y)) \to \exists y (A \to B(y)),$$

  where $y \notin \mathrm{FV}(A)$ and different restrictions on A.
- Markov's principle: $\neg\neg\exists x A_{\mathrm{qf}}(x) \to \exists x A_{\mathrm{qf}}(x)$.

Constructivity is Understood Locally (2/3)

Different interpretations validate different principles:

- Modified realisability validates full extensionality, $\mathrm{IP}^{\omega}_{\mathrm{ef}}$ and AC.
- Functional interpretation validates MP^{ω}, $\mathrm{IP}^{\omega}_{\forall}$ and AC.

Can there be a better interpretation; one which is more optimal with respect to these principles?

Constructivity is Understood Locally (3/3)

Two incompatible constructive theories:

- The proof interpretation is a global interpretation (a rule of thumb); locally one can accept more if the goal is computable existence.
- The combination of extensionality, Markov's principle and restricted forms of independence-of-premise is a subtle issue.

$$\text{WE-HA}^\omega + \text{IP}^\omega_\forall + \text{MP}^\omega + \text{AC} + \Gamma,$$

Γ is any set of universal true sentences; has existence property, disjunction property and is closed under various rules.

$$\text{E-HA}^\omega + \text{IP}^\omega_{\text{ef}} + \text{AC} + \Gamma,$$

Γ is any set of true $\exists$-free sentences; has existence property, disjunction property and is closed under different rules, except Markov's rule.

Philosophy (2/3)

On the Existence and Disjunction Properties

Gödel mentions the following result:

- If HA proves $\exists x A(x)$ then Σ proves the *translated* formula $A_D(t)$, for a term t.

But this does not suffice for proving the existence and disjunction properties for HA, as the original formula is not in general intuitionistically provable from the translated. In 1945 Kleene (and Nelson) proved that realisability by numbers can be used for showing that.

Philosophy (3/3):

What should we think of the consistency proof?

Coherence

Gödel's interpretation is paradigmatic with respect to coherence (Σ is now called T):

— PA is interpreted in T
— T is consistent, we can prove strong normalisation, by:
  · Howard's strong computability predicates
    (uses König's lemma)
  · Tait's method of ascribing ordinals $< \varepsilon_0$ to terms of T
— Fits with Gentzen's partial cut-elmination (which again fits with Schütte's full cut-elmination)
— Tait's proof of termination fits with Gentzen's characterisation of PA as ε_0
— The no-counter-example interpretation can be derived from both the Dialectica interpretation and Gentzen's cut-elimination

Neurath's Ship
Welcome Aboard

We cannot eliminate impredicativity. There is no secure Archimedian point—there is no *cogito*.

We give up foundationalism: Instead we can account (I think) for actual mathematics—contemporary as well as historical.