Semina ☐ Nr 3
Scientiarum 2004

Leszek Wroński

Twierdzenia Gödla — dowody. Czy arytmetyka jest w stanie dowieść własną niesprzeczność?

W tym krótkim opracowaniu chciałbym przedstawić dowody obu twierdzeń Gödla wykorzystujące warunki dowodliwości Löba, a także problem pewnej, być może zbyt pochopnie wysuwanej filozoficznej implikacji drugiego twierdzenia Gödla dotyczącego niedowiedlności niesprzeczności arytmetyki środkami samej arytmetyki.

1. Wstęp

Niech $\mathfrak{M} =< \mathbb{N}, +, \cdot, S, 0 >$ będzie tzw. modelem Peano, czyli standardowym modelem arytmetyki. Niech PA będzie systemem standardowo sformalizowanym opartym o następujące aksjomaty:

(1) $\forall x, y(Sx = Sy \rightarrow x = y)$

(2) $\forall x(\neg(0 = Sx))$

(3) $\forall x(x + 0 = x)$

(4) $\forall x, y(x + Sy = S(x + y))$

(5) $\forall x(x \cdot 0 = 0)$

(6) $\forall x, y(x \cdot Sy = x \cdot y + x)$

(7) $[\alpha(0) \wedge \forall x[\alpha(x) \rightarrow \alpha(S(x))]] \rightarrow \forall x\alpha(x)$

(oczywiście, (7) powyżej to schemat nieskończenie wielu aksjomatów).

Łatwo dowieść, iż $\mathfrak{M} \models PA$, tak więc możemy stwierdzić, co następuje:

Twierdzenie 1.1 *PA jest teorią niesprzeczną.*

Jasne jest, że zbiór aksjomatów PA jest rozstrzygalny, czyli istnieje algorytm rozpoznawania aksjomatów PA w zbiorze wszystkich formuł języka arytmetyki. Pierwsze twierdzenie Gödla pokazuje, iż arytmetyka PA jest istotnie niezupełna, czyli niezupełne jest każde jej rozszerzenie o rozstrzygalnym zbiorze aksjomatów.

Definicja 1.2 *Mówimy, że wyrażona w języku teorii PA formuła $\alpha(x_1, \ldots x_k)$ mocno reprezentuje relację $R \subseteq \mathbb{N}^k$ jeśli dla dowolnych $n_1, \ldots n_k$ jest tak, iż*

(1) $< n_1, \ldots n_k > \in R \Rightarrow PA \vdash \alpha(\overline{n_1}, \ldots \overline{n_k})$

(2) $< n_1, \ldots n_k > \notin R \Rightarrow PA \vdash \neg\alpha(\overline{n_1}, \ldots \overline{n_k})$,

gdzie $\overline{n_1}, \ldots \overline{n_k}$ są tzw. termami kanonicznymi, które w języku arytmetyki desygnują liczby naturalne — term oznaczający liczbę n ma postać $\underbrace{SS \ldots S}_{n-razy}0$.

W dalszym ciągu — dla uproszczenia notacji — liczbę oraz desygnujący ją term będziemy często oznaczać tym samym symbolem w sytuacjach, w których kontekst nie pozostawia wątpliwości co do tego, co mamy na myśli. Zakładam, iż Czytelnikowi znana jest jakaś metoda arytmetyzacji, czyli jednoznacznego przyporządkowania liczb naturalnych wyrażeniom języka arytmetyki. Może być to metoda samego Gödla, albo np. Smullyana lub Boolosa. Nie będzie dla nas ważny sposób dokonania arytmetyzacji, lecz tylko fakt, iż jest ona możliwa. Numer Gödla formuły α będziemy oznaczać często używanym symbolem $\ulcorner\alpha\urcorner$. Wprowadźmy następujące relacje:

Definicja 1.3 *Prf(a,b)* wtw *a jest numerem Gödla dowodu formuły o numerze Gödla b.*

Definicja 1.4 *Pr(a)* wtw $\exists x : Prf(x,a)$
(w obu definicjach $a, b, x \in \mathbb{N}$).

Dla dowodu twierdzeń Gödla potrzebny nam będzie następujący fakt:

Fakt 1.5 *Relacja Prf jest mocno reprezentowalna w teorii PA.*

Dla uproszczenia notacji używać będziemy tego samego symbolu na określenie relacji Prf oraz formuły w języku PA, która ową relację mocno reprezentuje. Łatwo udowodnić następujący fakt:

Fakt 1.6 *Dla dowolnej formuły α, jeśli numer Gödla $\ulcorner\alpha\urcorner$ ma własność Pr, to $PA \vdash \alpha$.*

Przypomnijmy, że w zależności od kontekstu symbol Prf będzie oznaczał relację zawartą w $\mathbb{N} \times \mathbb{N}$ lub formułę arytmetyczną z dwiema zmiennymi wolnymi, która tę relację mocno reprezentuje w PA.

Definicja 1.7 *Mówimy, że formuła $\varphi(x, y, z)$ mocno reprezentuje funkcję f $(f : \mathbb{N}^2 \mapsto \mathbb{N})$, jeżeli dla dowolnych $n, m \in \mathbb{N}$ jest tak, że*

$$PA \vdash \forall z[\varphi(\overline{n}, \overline{m}, z) \leftrightarrow [z = \overline{f(n,m)}]].$$

Tak więc reprezentacją k-argumentowej funkcji jest formuła z $k + 1$ zmiennymi wolnymi.

Rozważmy funkcję $Sub : \mathbb{N}^2 \to \mathbb{N}$. Niech $Sub(\ulcorner\alpha\urcorner, n) = \ulcorner\alpha(x/\overline{n})\urcorner$, jeśli pierwszy argument jest numerem formuły, zaś w innym przypadku wartość funkcji nie jest określona. Wiemy, iż prawdziwy jest następujący fakt:

Fakt 1.8 *Istnieje formuła arytmetyczna $\mathfrak{S}(x, y, z)$, która mocno reprezentuje w PA funkcję Sub.*

Udowodnijmy następujące twierdzenie, zwane „lematem przekątniowym”:

Twierdzenie 1.9 *Dla dowolnej formuły arytmetycznej $\theta(x)$ zawierającej dokładnie jedną zmienną wolną x istnieje zdanie Δ takie, że*

$$PA \vdash \Delta \leftrightarrow \theta(\overline{\ulcorner \Delta \urcorner}).$$

Dowód: Nasze intuicje co do zdania Δ można wyrazić stwierdzeniem, iż mówi ono „Mój numer ma własność θ”. Zdefiniujmy formułę $\Gamma(x)$ z jedną zmienną wolną:

$$\Gamma(x) \stackrel{\text{def}}{=} \forall y[\mathfrak{S}(x, x, y) \rightarrow \theta(y)].$$

Niech $n = \ulcorner \Gamma(x) \urcorner$. Zdefiniujmy tak oto zdanie Δ:

$$\Delta \stackrel{\text{def}}{=} \Gamma(x/\overline{n}) \tag{1}$$

czyli

$$\Delta \stackrel{\text{def}}{=} \forall y[\mathfrak{S}(\overline{n}, \overline{n}, y) \rightarrow \theta(y)].$$

Wiemy, że $PA \vdash \Delta \leftrightarrow \Delta$, a zatem

$$PA \vdash \Delta \leftrightarrow \forall y[\mathfrak{S}(\overline{n}, \overline{n}, y) \rightarrow \theta(y)]. \tag{2}$$

Fakty, iż formuła $\mathfrak{S}$ mocno reprezentuje w PA funkcję Sub, a także, iż $n = \ulcorner \Gamma(x) \urcorner$ implikują[1]:

$$PA \vdash \forall y[\mathfrak{S}(\overline{n}, \overline{n}, y) \leftrightarrow y = \overline{\ulcorner \Gamma(\overline{n}) \urcorner}]. \tag{3}$$

Z (2) i (3) otrzymujemy

$$PA \vdash \Delta \leftrightarrow \forall y[(y = \overline{\ulcorner \Gamma(\overline{n}) \urcorner}) \rightarrow \theta(y)]. \tag{4}$$

Na mocy definicji (1):

$$\overline{\ulcorner \Gamma(\overline{n}) \urcorner} = \overline{\ulcorner \Delta \urcorner}.$$

[1] Stosujemy def. 1.7 do funkcji Sub i reprezentującej ją formuły $\mathfrak{S}$.

Po podstawieniu do (4) otrzymujemy:

$$PA \vdash \Delta \leftrightarrow \forall y[(y = \overline{\ulcorner \Delta \urcorner}) \rightarrow \theta(y)],$$

czyli

$$PA \vdash \Delta \leftrightarrow \theta(\overline{\ulcorner \Delta \urcorner}).$$

Q.E.D.

Dowody twierdzeń Gödla sprowadzają się do zwykłego rachunku zdań przy założeniu tzw. warunków dowodliwości Löba:

(D1) jeśli $PA \vdash \varphi$, to PA $\vdash Pr(\ulcorner \varphi \urcorner)$

(D2) $PA \vdash Pr(\ulcorner \varphi \urcorner) \rightarrow Pr(\ulcorner Pr(\ulcorner \varphi \urcorner) \urcorner)$

(D3) $PA \vdash [Pr(\ulcorner \varphi \urcorner) \wedge Pr(\ulcorner \varphi \rightarrow \psi \urcorner)] \rightarrow Pr(\ulcorner \psi \urcorner)$

2. Twierdzenia Gödla

Poniżej wykorzystywać będziemy zdanie φ_g, które wyraża własną niedowodliwość środkami PA. Fakt, iż owo zdanie da się napisać w języku arytmetyki uzasadniamy powołując się na (1.5) oraz (1.9).

Twierdzenie 2.1 (I twierdzenie Gödla) *Niech φ_g będzie takim zdaniem języka PA, że $PA \vdash \varphi_g \leftrightarrow \neg Pr(\ulcorner \varphi_g \urcorner)$. Wtedy:*

(i) $PA \nvdash \varphi_g$

(ii) $PA \nvdash \neg\varphi_g$.

Dowód: Najpierw dowiedziemy warunku (i). Załóżmy, że $PA \vdash \varphi_g$. Wtedy, z warunku D1 mamy $PA \vdash Pr(\ulcorner \varphi_g \urcorner)$. Ponieważ $PA \vdash \neg\varphi_g \leftrightarrow Pr(\ulcorner \varphi_g \urcorner)$, otrzymujemy wniosek, iż $PA \vdash \neg\varphi_g$, dochodzimy więc do sprzeczności z twierdzeniem (1.1). Zatem $PA \nvdash \varphi_g$.

Następnie dowiedziemy warunku (ii). Załóżmy, że $PA \vdash \neg\varphi_g$. Zatem $PA \vdash Pr(\ulcorner \varphi_g \urcorner)$, czyli, z faktu (1.6), $PA \vdash \varphi_g$, co prowadzi do sprzeczności. Tak więc $PA \nvdash \neg\varphi_g$. Q.E.D.

Twierdzenie 2.2 (II twierdzenie Gödla) *Niech* Con_{PA} *oznacza formułę* $\neg Pr(\ulcorner 0 = 1 \urcorner)$. *Wtedy* $PA \nvdash Con_{PA}$.

Dowód: Niech φ_g będzie zdaniem takim, że $PA \vdash \varphi_g \leftrightarrow \neg Pr(\ulcorner \varphi_g \urcorner)$. Chcemy pokazać, że $PA \vdash \varphi_g \leftrightarrow Con_{PA}$.

Udowodnimy najpierw implikację w prawo, $PA \vdash \varphi_g \rightarrow Con_{PA}$. Zauważmy, że $PA \vdash (0 = 1) \rightarrow \varphi_g$. Z warunku D1 otrzymujemy: $PA \vdash Pr(\ulcorner (0 = 1) \rightarrow \varphi_g \urcorner)$. Po wykorzystaniu warunku D3 mamy: $PA \vdash Pr(\ulcorner 0 = 1 \urcorner) \rightarrow Pr(\ulcorner \varphi_g \urcorner)$. Po zastosowaniu prawa kontrapozycji: $PA \vdash \neg Pr(\ulcorner \varphi_g \urcorner) \rightarrow \neg Pr(\ulcorner 0 = 1 \urcorner)$. Z definicji φ_g: $PA \vdash \varphi_g \rightarrow \neg Pr(\ulcorner \varphi_g \urcorner)$. Zatem, z przechodniości implikacji, $PA \vdash \varphi_g \rightarrow \neg Pr(\ulcorner 0 = 1 \urcorner)$, czyli $PA \vdash \varphi_g \rightarrow Con_{PA}$.

Następnie udowodnijmy implikację w lewo,
$PA \vdash Con_{PA} \rightarrow \varphi_g$.

Z definicji φ_g mamy

$$PA \vdash \varphi_g \rightarrow \neg Pr(\ulcorner \varphi_g \urcorner)$$

z czego po zastosowaniu transpozycji wynika

$$PA \vdash Pr(\ulcorner \varphi_g \urcorner) \rightarrow \neg \varphi_g.$$

Z warunku D1 otrzymujemy

$$PA \vdash Pr(\ulcorner Pr(\ulcorner \varphi_g \urcorner) \rightarrow \neg \varphi_g \urcorner)$$

zaś po zastosowaniu warunku D3

$$PA \vdash Pr(\ulcorner Pr(\ulcorner \varphi_g \urcorner) \urcorner) \rightarrow Pr(\ulcorner \neg \varphi_g \urcorner). \tag{5}$$

Z D2, (5) oraz przechodniości implikacji otrzymujemy

$$PA \vdash Pr(\ulcorner \varphi_g \urcorner) \rightarrow Pr(\ulcorner \neg \varphi_g \urcorner). \tag{6}$$

Wiemy, że $PA \vdash \varphi_g \rightarrow (\neg \varphi_g \rightarrow (\varphi_g \wedge \neg \varphi_g))$. Zatem, na mocy D1 i D3, otrzymujemy

$$PA \vdash Pr(\ulcorner \varphi_g \urcorner) \rightarrow (Pr(\ulcorner \neg \varphi_g \urcorner) \rightarrow Pr(\ulcorner \varphi_g \wedge \neg \varphi_g \urcorner))$$

a zatem, po zastosowaniu prawa komutacji,

$$PA \vdash Pr(\ulcorner \neg\varphi_g \urcorner) \rightarrow (Pr(\ulcorner \varphi_g \urcorner) \rightarrow Pr(\ulcorner \varphi_g \wedge \neg\varphi_g \urcorner)). \qquad (7)$$

Z samorozdzielności implikacji i prawa komutacji otrzymujemy tautologię

$$[p \rightarrow (q \rightarrow r)] \rightarrow [(q \rightarrow p) \rightarrow (q \rightarrow r)].$$

Za zmienną p podstawiamy $Pr(\ulcorner \neg\varphi_g \urcorner)$, za q $Pr(\ulcorner \varphi_g \urcorner)$ oraz za r $Pr(\ulcorner \varphi_g \wedge \neg\varphi_g \urcorner)$, by otrzymać

$$PA \vdash [Pr(\ulcorner \neg\varphi_g \urcorner) \rightarrow (Pr(\ulcorner \varphi_g \urcorner) \rightarrow Pr(\ulcorner \varphi_g \wedge \neg\varphi_g \urcorner))] \rightarrow$$
$$\rightarrow [(Pr(\ulcorner \varphi_g \urcorner) \rightarrow Pr(\ulcorner \neg\varphi_g \urcorner)) \rightarrow (Pr(\ulcorner \varphi_g \urcorner) \rightarrow Pr(\ulcorner \varphi_g \wedge \neg\varphi_g \urcorner))].$$

Odrywamy (7), (6) i otrzymujemy

$$PA \vdash Pr(\ulcorner \varphi_g \urcorner) \rightarrow Pr(\ulcorner \varphi_g \wedge \neg\varphi_g \urcorner). \qquad (8)$$

Wiemy, że $PA \vdash (\varphi_g \wedge \neg\varphi_g) \leftrightarrow (0 = 1)$. Zatem w szczególności $PA \vdash (\varphi_g \wedge \neg\varphi_g) \rightarrow (0 = 1)$. Z warunków D1 i D3 otrzymujemy:

$$PA \vdash Pr(\ulcorner \varphi_g \wedge \neg\varphi_g \urcorner) \rightarrow Pr(\ulcorner 0 = 1 \urcorner). \qquad (9)$$

Z (8) i przechodniości implikacji mamy

$$PA \vdash Pr(\ulcorner \varphi_g \urcorner) \rightarrow Pr(\ulcorner 0 = 1 \urcorner),$$

a zatem (z prawa kontrapozycji)

$$PA \vdash \neg Pr(\ulcorner 0 = 1 \urcorner) \rightarrow \neg Pr(\ulcorner \varphi_g \urcorner).$$

Z definicji Con_{PA} i φ_g mamy

$$PA \vdash Con_{PA} \rightarrow \varphi_g,$$

co chcieliśmy udowodnić.

Wnioskujemy więc, że $PA \vdash \varphi_g \leftrightarrow Con_{PA}$. Z udowodnionego wyżej pierwszego twierdzenia Gödla wiemy, że $PA \nvdash \varphi_g$. Zatem $PA \nvdash Con_{PA}$. Q.E.D.

3. Niesprzeczność arytmetyki

Często słyszy sie pogląd, iż drugie twierdzenie Gödla implikuje niedowiedlność niesprzeczności arytmetyki środkami samej arytmetyki. Można by sądzić, iż dobrą interpretacją powyższego poglądu jest następujące zdanie (występujący w nim symbol „$\bot$" oznacza wybraną „formułę absurdalną" (np. „$\neg(0 = 0)$"): nie może ona być twierdzeniem żadnej niesprzecznej teorii zawierającej arytmetykę):

Nie istnieje formuła arytmetyczna $\gamma(x)$ taka, że

(1) dla dowolnej formuły α, $PA \vdash \alpha$ wtw $PA \vdash \gamma(\ulcorner\alpha\urcorner)$ (czyli formuła γ wyraża dowiedlność w PA)

(2) $PA \vdash \neg\gamma(\ulcorner\bot\urcorner)$.

Jednakże zdanie powyższe jest fałszywe. Zdefiniujmy $\gamma(x)$ tak oto:

$$\gamma(x) \leftrightarrow \exists y[Prf(y, x) \wedge (x \neq \ulcorner\bot\urcorner)]$$

Jeśli $PA \vdash \alpha$ to, wobec niesprzeczności arytmetyki, $\ulcorner\alpha\urcorner \neq \ulcorner\bot\urcorner$; tak więc (ponieważ relacja różności między liczbami naturalnymi jest mocno reprezentowalna w PA) $PA \vdash (\ulcorner\alpha\urcorner \neq \ulcorner\bot\urcorner)$. Zatem

$$PA \vdash \exists y[Prf(y, \ulcorner\alpha\urcorner) \wedge (\ulcorner\alpha\urcorner \neq \ulcorner\bot\urcorner)]$$

z czego wnioskujemy, iż $PA \vdash \gamma(\ulcorner\alpha\urcorner)$. Mamy więc implikację w lewo z powyższego warunku (1). Implikacja w prawo jest oczywista z definicji własności Prf.

Aby dowieść warunku (2) zauważamy, że

$$\neg\gamma(\ulcorner\bot\urcorner) \leftrightarrow \neg\exists y : (Prf(y, \ulcorner\bot\urcorner) \wedge (\ulcorner\bot\urcorner \neq \ulcorner\bot\urcorner)) \leftrightarrow$$
$$\leftrightarrow \forall y : (Prf(y, \ulcorner\bot\urcorner) \to (\ulcorner\bot\urcorner = \ulcorner\bot\urcorner))$$

zaś to ostatnie jest prawem logiki.

Stajemy więc w obliczu takiej sytuacji: wiemy, iż

$$PA \vdash \alpha \text{ wtw } PA \vdash Pr(\ulcorner\alpha\urcorner) \text{ wtw } PA \vdash \gamma(\ulcorner\alpha\urcorner)$$

a także

$$PA \vdash \neg\gamma(\ulcorner\bot\urcorner)$$

oraz (z poprzedniego twierdzenia)

$$PA \nvdash \neg Pr(\ulcorner\bot\urcorner).$$

Tak więc istnieje formuła arytmetyczna $\gamma(x)$ wyrażająca dowiedlność w PA i taka, że $PA \vdash \neg\gamma(\ulcorner\bot\urcorner)$. Widać stąd, iż drugie twierdzenie Gödla wymaga, by formuła Pr wybrana została w sposób „właściwy”, tak, by miała pożądany sens arytmetyczny. Może się bowiem zdarzyć, iż dla odpowiednio skonstruowanej formuły Pr_1 formuła

$$Pr(\ulcorner\alpha\urcorner) \leftrightarrow Pr_1(\ulcorner\alpha\urcorner)$$

będzie prawdziwa w modelu $\mathfrak{M}$ (czyli $Th(\mathfrak{M}) \models Pr(\ulcorner\alpha\urcorner) \leftrightarrow Pr_1(\ulcorner\alpha\urcorner)$), ale nie będzie twierdzeniem PA (z pierwszego twierdzenia Gödla wiemy, że jest to możliwe)[2].

[2]Podane tu informacje można w dużej części znaleźć w książce Romana Murawskiego „Funkcje rekurencyjne i elementy metamatematyki” (Wydawnictwo Naukowe UAM, Poznań 2000) oraz drugim wykładzie z „Foundational Studies” Andrzeja Mostowskiego (PWN 1979).