Metalogika (8)

Jerzy Pogonowski

Zakład Logiki Stosowanej UAM
www.logic.amu.edu.pl
pogon@amu.edu.pl

Uniwersytet Opolski

Plan wykładu

W tym wykładzie podamy kilka ważnych twierdzeń metalogicznych, wraz z dowodami.

- Twierdzenie Gödla o niezupełności PA.
- Twierdzenie Rossera o niezupełności PA.
- Twierdzenie Gödla o niedowodliwości niesprzeczności PA w PA.
- Twierdzenie Löba.
- Twierdzenie Tarskiego o niedefiniowalności pojęcia prawdy arytmetycznej.

Dowody tych twierdzeń w istotny sposób wykorzystują procedurę arytmetyzacji składni opisaną w poprzednim wykładzie.

Teorie rekurencyjnie aksjomatyzowalne

- Procedurę arytmetyzacji składni można przeprowadzić dla dowolnej teorii pierwszego rzędu, której zbiór numerów gödlowskich symboli pozalogicznych jest rekurencyjny.
- Jeśli jednak zbiór numerów gödlowskich aksjomatów pozalogicznych teorii T nie jest rekurencyjny, to relacja $Dow_T(a, b)$ (czytaj: a jest numerem gödlowskim dowodu w teorii T formuły o numerze gödlowskim b) nie musi być rekurencyjna. W konsekwencji, w takim przypadku zbiór numerów gödlowskich twierdzeń teorii T nie musi być rekurencyjnie przeliczalny.

- Mówimy, że teoria T jest ***(rekurencyjnie) aksjomatyzowalna***, gdy zbiór numerów gödlowskich aksjomatów teorii T jest rekurencyjny.
- Arytmetyka PA jest rekurencyjnie aksjomatyzowalna.

Zupełność i rozstrzygalność

Niech T będzie teorią pierwszego rzędu, której zbiór numerów gödlowskich symboli pozalogicznych jest rekurencyjny. Mówimy, że T jest:

- ***zupełna***, gdy dla dowolnego zdania ψ jej języka: albo $T \vdash \psi$, albo $T \vdash \neg\psi$; w przeciwnym przypadku T nazywamy ***niezupełną***.
- ***rozstrzygalna***, gdy zbiór numerów gödlowskich jej twierdzeń jest rekurencyjny; w przeciwnym przypadku T nazywamy ***nierozstrzygalną***.

- Teoria T jest zatem zupełna, gdy dla dowolnego pytania rozstrzygnięcia sformułowanego w jej języku potrafimy w T udowodnić: albo odpowiedź TAK, albo odpowiedź NIE na to pytanie.
- Teoria T jest rozstrzygalna, gdy istnieje obliczalna metoda pozwalająca rozstrzygać o dowolnej formule jej języka czy jest ona twierdzeniem T czy nie jest [zakładamy tu Tezę Churcha: obliczalne=rekurencyjne].

ω-niesprzeczność

Niech T będzie teorią (pierwszego rzędu), w której języku mamy liczebniki (nazwy liczb naturalnych). Jak zwykle, $T \vdash \psi$ oznacza, że istnieje dowód formuły ψ w teorii T. Piszemy T *non* $\vdash \psi$, gdy nie zachodzi $T \vdash \psi$.

- Mówimy, że teoria T jest ***ω-niesprzeczna***, gdy dla każdej formuły $\psi(x)$: jeśli $T \vdash \psi(\underline{0})$, $T \vdash \psi(\overline{1})$, $T \vdash \psi(\overline{2})$, ..., $T \vdash \psi(\overline{n})$,..., to T *non* $\vdash \exists x\, \neg\psi(x)$.

- **Twierdzenie**. Jeśli PA jest ω-niesprzeczna, to jest niesprzeczna.
- **Zarys dowodu**. Wystarczy znaleźć choć jedną formułę, która nie jest twierdzeniem PA. Mamy: $PA \vdash x \doteq x \rightarrow x \doteq x$, a zatem $PA \vdash \overline{n} \doteq \overline{n} \rightarrow \overline{n} \doteq \overline{n}$ dla wszystkich n. Z założenia o ω-niesprzeczności mamy: PA *non* $\vdash \exists x\, \neg(x \doteq x \rightarrow x \doteq x)$.

Konstrukcja zdania Gödla

Funkcję num określamy przez schemat rekursji prostej:

- $\mathrm{num}(0) = \langle sn(\underline{0}) \rangle$
- $\mathrm{num}(a+1) = \langle sn(\underline{s}), \mathrm{num}(a) \rangle$.

Wtedy $\mathrm{num}(n)$ jest numerem gödlowskim liczebnika $\overline{n}$. Funkcja num jest rekurencyjna. Przypominamy, że $\langle\ \rangle$ jest tu funkcją kodowania ciągów zdefiniowaną w poprzednim wykładzie.

Nie zagub się! Należy odróżniać:

- liczbę naturalną n
- liczebnik $\overline{n}$
- numer gödlowski $\mathrm{num}(n)$ liczebnika $\overline{n}$.

Konstrukcja zdania Gödla

Niech sam będzie dwuargumentową relacją zdefiniowaną następująco:
$\mathrm{sam}(a, b) \equiv \mathrm{Form}(a) \wedge \mathrm{Fr}(a, 2) \wedge \mathrm{Dow}(b, \mathrm{sub}(a, 2, \mathrm{num}(a)))$.

- Jeśli a jest numerem gödlowskim formuły, powiedzmy, $\psi(x_1)$, to $\mathrm{sub}(a, 2, \mathrm{num}(a))$ jest numerem gödlowskim formuły, która powstaje z formuły $\psi(x_1)$ poprzez wstawienie za zmienną x_1 liczebnika nazywającego liczbę a, czyli nazywającego właśnie numer gödlowski samej formuły ψ.
- Tak więc, relacja sam zachodzi między liczbami a oraz b dokładnie wtedy, gdy:
  - a jest numerem gödlowskim formuły o zmiennej wolnej x_1,
  - b jest numerem gödlowskim dowodu formuły o numerze gödlowskim $\mathrm{sub}(a, 2, \mathrm{num}(a))$, czyli formuły otrzymanej z formuły o numerze gödlowskim a w wyżej podany sposób.

Konstrukcja zdania Gödla

Komentarz dydaktyczny. Mamy formułę, powiedzmy, $\psi(x_1)$ o jednej zmiennej wolnej x_1 (wybór tej właśnie zmiennej jest nieistotny).

- Formuła ta ma swój numer gödlowski, powiedzmy, a, czyli $\ulcorner\psi(x_1)\urcorner = a$.
- Liczba $\mathrm{num}(a)$ jest numerem gödlowskim liczebnika $\overline{a}$.
- Do formuły $\psi(x_1)$ chcemy wstawić, w miejsce zmiennej wolnej x_1 term $\overline{a}$, czyli chcemy otrzymać formułę $\psi(\overline{a})$, która (na mocy definicji liczby a) jest formułą $\psi(\overline{\ulcorner\psi(x_1)\urcorner})$.
- Liczba $\mathrm{sub}(a, 2, \mathrm{num}(a))$ jest właśnie numerem gödlowskim otrzymanej w ten sposób formuły: $\mathrm{sub}(a, 2, \mathrm{num}(a)) = \ulcorner\psi(\overline{\ulcorner\psi(x_1)\urcorner})\urcorner$.

- Pamiętaj: do formuły podstawiamy (w miejsce zmiennej wolnej) term. W szczególności, term ten może być liczebnikiem.

Konstrukcja zdania Gödla

Ponieważ sam jest relacją rekurencyjną, więc (na mocy twierdzenia o reprezentowalności) istnieje co najmniej jedna formuła języka PA, która mocno reprezentuje tę relację. Niech $\underline{\mathrm{sam}}(x, y)$ będzie taką formułą.

- Rozważmy formułę o postaci: $\forall y \neg \underline{\mathrm{sam}}(x, y)$.
- Niech $m = \ulcorner \forall y \neg \underline{\mathrm{sam}}(x, y) \urcorner$, czyli niech m będzie numerem gödlowskim formuły $\forall y \neg \underline{\mathrm{sam}}(x, y)$.
- Niech *god* będzie zdaniem: $\forall y \neg \underline{\mathrm{sam}}(\overline{m}, y)$.
- Zdanie *god* nazywamy ***zdaniem Gödla***.
- Zdanie *god* stwierdza zatem, że formuła o numerze gödlowskim m nie ma dowodu w PA.
- Ponieważ m jest numerem gödlowskim formuły $\forall y \neg \underline{\mathrm{sam}}(x, y)$, więc zdanie Gödla *god* stwierdza, że zdanie *god* nie jest twierdzeniem PA, czyli głosi ono samo o sobie: „nie jestem twierdzeniem PA.”

I Twierdzenie Gödla (o niezupełności PA)

I Twierdzenie Gödla (o niezupełności PA)

Jeśli PA jest ω-niesprzeczna, to ani zdanie god, ani zdanie $\neg\, god$ nie ma dowodu w PA:

1. $PA\ non \vdash god$
2. $PA\ non \vdash \neg\, god$.

Tak więc, PA jest niezupełna.

- Zauważmy, że jedno ze zdań: god, $\neg\, god$ musi być prawdziwe w modelu standardowym $\mathfrak{N}_0$. Zobaczymy, że $\mathfrak{N}_0 \models god$.
- Dla dowodu $PA\ non \vdash god$ wystarczy założenie niesprzeczności PA; dowód $PA\ non \vdash \neg\, god$ wymaga silniejszego założenia ω-niesprzeczności.

Dowód I Twierdzenia Gödla

PA non $\vdash$ *god*.

- Dla dowodu nie wprost przypuśćmy, że $PA \vdash god$, czyli że *god* ma dowód w PA.
- Niech k będzie numerem gödlowskim jakiegoś dowodu zdania *god* (pamiętamy, że dowody, jako ciągi formuł, też mają numery gödlowskie).
- Zachodzi zatem $\mathrm{sam}(m, k)$. Ponieważ $\underline{\mathrm{sam}}$ mocno reprezentuje relację sam, więc $PA \vdash \underline{\mathrm{sam}}(\overline{m}, \overline{k})$.
- Skoro $PA \vdash god$, czyli $PA \vdash \forall y \neg \underline{\mathrm{sam}}(\overline{m}, y)$, to $PA \vdash \neg \underline{\mathrm{sam}}(\overline{m}, \overline{k})$.
- Skoro $PA \vdash \underline{\mathrm{sam}}(\overline{m}, \overline{k})$ oraz $PA \vdash \neg \underline{\mathrm{sam}}(\overline{m}, \overline{k})$, to PA jest sprzeczna, wbrew założeniu (bo zakładamy, że PA jest nawet ω-niesprzeczna).
- Przypuszczenie dowodu nie wprost trzeba zatem odrzucić. Ostatecznie, *PA non* $\vdash$ *god*.

Dowód I Twierdzenia Gödla

PA *non* $\vdash \neg$ *god*.

- Pokazaliśmy, że PA *non* $\vdash$ *god*, a więc nie istnieje liczba naturalna n, która byłaby numerem gödlowskim dowodu *god* w PA.
- Dla każdej n: **nie** zachodzi zatem sam(m, n).
- Ponieważ $\underline{\text{sam}}$ mocno reprezentuje relację sam, więc dla wszystkich n mamy: $PA \vdash \neg\, \underline{\text{sam}}(\overline{m}, \overline{n})$.
- Na mocy ω-niesprzeczności PA mamy: PA *non* $\vdash \exists y \neg\neg\, \underline{\text{sam}}(\overline{m}, y)$, co jest równoważne temu, iż PA *non* $\vdash \neg\forall y \neg\, \underline{\text{sam}}(\overline{m}, y)$.
- Ponieważ $\neg\forall y \neg\, \underline{\text{sam}}(\overline{m}, y)$ jest formułą $\neg$ *god*, więc PA *non* $\vdash \neg$ *god*.

Dowód całego twierdzenia został tym samym zakończony.

I Twierdzenie Gödla: komentarz

- Zdanie Gödla *god* jest formułą generalnie skwantyfikowaną: $\forall y \neg \underline{\mathrm{sam}}(\overline{m}, y)$.
- Ponieważ:
  - PA $non \vdash god$ oraz
  - $\underline{\mathrm{sam}}$ mocno reprezentuje w PA relację sam,

  więc dla każdej liczby naturalnej n mamy: $PA \vdash \neg \underline{\mathrm{sam}}(\overline{m}, \overline{n})$.
- Tak więc, choć samo (generalnie skwantyfikowane) zdanie Gödla jest nierozstrzygalne w PA, to wszystkie jego szczególne przypadki (gdy pomijamy kwantyfikator generalny i wstawiamy liczebnik za zmienną) są twierdzeniami PA.

- Założenie ω-niesprzeczności można osłabić do zwykłej niesprzeczności, jak za chwilę zobaczymy.

Konstrukcja zdania Rossera

Zdefiniujmy dwuargumentową relacje rekurencyjną samneg:

$\mathrm{samneg}(a, b) \equiv \mathrm{Form}(a) \wedge \mathrm{Fr}(a, 2) \wedge \mathrm{Dow}(b, \mathrm{sub}(\langle 3, a\rangle, 2, \mathrm{num}(a)))$.

- Relacja samneg zachodzi zatem między liczbami a oraz b dokładnie wtedy, gdy a jest numerem gödlowskim pewnej formuły, powiedzmy, $\psi(x_1)$ o zmiennej wolnej x_1, natomiast b jest numerem gödlowskim dowodu formuły otrzymanej przez podstawienie w formule $\neg\psi(x_1)$ za zmienną x_1 liczebnika nazywającego liczbę a, czyli numer gödlowski samej formuły $\psi(x_1)$.

- Relacja samneg jest rekurencyjna, a zatem istnieje co najmniej jedna formuła języka PA, która ją mocno reprezentuje. Niech $\underline{\mathrm{samneg}}$ będzie taką formułą.
- Jak poprzednio, niech formuła $\underline{\mathrm{sam}}$ mocno reprezentuje relację sam.

Konstrukcja zdania Rossera

- Rozważmy formułę: $\forall y\ (\underline{\mathrm{sam}}(x,y) \to \exists z\ (z \underline{\leqslant} y \wedge \underline{\mathrm{samneg}}(x,z)))$. Tu $\underline{\leqslant}$ jest predykatem o denotacji $\leqslant$.
- Niech n będzie numerem gödlowskim tej formuły, czyli: $n = \ulcorner \forall y\ (\underline{\mathrm{sam}}(x,y) \to \exists z\ (z \underline{\leqslant} y \wedge \underline{\mathrm{samneg}}(x,z))) \urcorner$.
- Niech *ros* będzie zdaniem: $\forall y\ (\underline{\mathrm{sam}}(\overline{n},y) \to \exists z\ (z \underline{\leqslant} y \wedge \underline{\mathrm{samneg}}(\overline{n},z)))$.
- Zdanie *ros* nazwiemy ***zdaniem Rossera***.

Dla każdej liczby naturalnej y mamy:

- (†) $\mathrm{sam}(n,y)$ dokładnie wtedy, gdy y jest numerem gödlowskim dowodu w PA zdania *ros*
- (‡) $\mathrm{samneg}(n,y)$ dokładnie wtedy, gdy y jest numerem gödlowskim dowodu w PA zdania $\neg$ *ros*.

Twierdzenie Rossera

- Zdanie Rossera *ros* stwierdza zatem, że jeśli istnieje w PA dowód zdania *ros*, to istnieje w PA również dowód (o niewiększym numerze gödlowskim) zdania $\neg$ *ros*.
- Zdanie *ros* stwierdza więc, że jeśli ono samo jest twierdzeniem PA, to twierdzeniem PA jest także jego negacja.

Twierdzenie Rossera.

Jeśli PA jest niesprzeczna, to ani zdanie ros, ani zdanie $\neg$ ros nie ma dowodu w PA:

1. *PA non $\vdash$ ros*
2. *PA non $\vdash \neg$ ros.*

Tak więc, PA jest niezupełna.

Dowód Twierdzenia Rossera

1. PA *non* $\vdash$ *ros*

- Przypuśćmy, dla dowodu nie wprost, że $PA \vdash ros$ i niech k będzie numerem gödlowskim jakiegoś dowodu *ros* w PA.
- Wtedy (na mocy (†)) sam(n, k), a więc $PA \vdash \underline{\mathrm{sam}}(\overline{n}, \overline{k})$.
- Na mocy przypuszczenia dowodu nie wprost mamy: $PA \vdash \underline{\mathrm{sam}}(\overline{n}, \overline{k}) \rightarrow \exists z\ (z \underline{\leqslant} \overline{k} \wedge \underline{\mathrm{samneg}}(\overline{n}, z))$.
- Na mocy reguły odrywania mamy: $(*)$ $PA \vdash \exists z\ (z \underline{\leqslant} \overline{k} \wedge \underline{\mathrm{samneg}}(\overline{n}, z))$.
- Na mocy założenia, że PA niesprzeczna: nie istnieje w PA dowód zdania $\neg$ *ros*.
- Na mocy (‡), dla każdej y: ***nie*** zachodzi samneg(n, y).
- Ponieważ $\underline{\mathrm{samneg}}$ mocno reprezentuje samneg w PA, więc dla wszystkich i mamy: $PA \vdash \neg\, \underline{\mathrm{samneg}}(\overline{n}, \overline{i})$.

Dowód Twierdzenia Rossera

- W szczególności:
  $PA \vdash \neg\, \underline{\mathrm{samneg}}(\overline{n}, \underline{0}) \wedge \neg\, \underline{\mathrm{samneg}}(\overline{n}, \overline{1}) \wedge \ldots \wedge \neg\, \underline{\mathrm{samneg}}(\overline{n}, \overline{k})$.
- Na mocy faktu podanego w poprzednim wykładzie mamy:
  $PA \vdash (\neg\, \underline{\mathrm{samneg}}(\overline{n}, \underline{0}) \wedge \neg\, \underline{\mathrm{samneg}}(\overline{n}, \overline{1}) \wedge \ldots \wedge \neg\, \underline{\mathrm{samneg}}(\overline{n}, \overline{k})) \rightarrow \forall z\, (z \underline{\leqslant} \overline{k} \rightarrow \neg\, \underline{\mathrm{samneg}}(\overline{n}, z))$.
- Na mocy reguły odrywania mamy:
  $(**)\ \ PA \vdash \forall z\, (z \underline{\leqslant} \overline{k} \rightarrow \neg\, \underline{\mathrm{samneg}}(\overline{n}, z))$.
- Skoro zachodzą $(*)$ oraz $(**)$, to PA jest sprzeczna, wbrew założeniu.
- Tak więc, przypuszczenie dowodu nie wprost $PA \vdash ros$ trzeba odrzucić jako fałszywe.
- Ostatecznie, $PA\ non \vdash ros$.

Dowód Twierdzenia Rossera

2. PA *non* $\vdash \neg$ *ros*

- Przypuśćmy, dla dowodu nie wprost, że $PA \vdash \neg\, ros$ i niech r będzie numerem jakiegoś dowodu $\neg\, ros$ w PA.
- Na mocy (‡) mamy: samneg(n, r), a na mocy mocnej reprezentowalności samneg przez $\underline{\text{samneg}}$ mamy: $PA \vdash\ \underline{\text{samneg}}(\overline{n}, \overline{r})$.
- Z założenia niesprzeczności PA oraz przypuszczenia dowodu nie wprost mamy: PA *non* $\vdash$ *ros*.
- Tak więc, żadna liczba y nie jest numerem gödlowskim dowodu zdania *ros*, co oznacza, że dla każdej y: ***nie*** zachodzi sam(n, y).
- W konsekwencji, $PA \vdash \neg\, \underline{\text{sam}}(\overline{n}, \overline{i})$, dla wszystkich i.
- Mamy więc: $PA \vdash \neg\, \underline{\text{sam}}(\overline{n}, \underline{0}) \wedge \neg\, \underline{\text{sam}}(\overline{n}, \overline{1}) \wedge \ldots \wedge \neg\, \underline{\text{sam}}(\overline{n}, \overline{r})$.

Dowód Twierdzenia Rossera

- Tak samo jak w dowodzie punktu 1 otrzymujemy stąd: $PA \vdash y \underline{\leqslant} \overline{r} \rightarrow \neg\, \underline{\mathsf{sam}}(\overline{n}, y)$.
- Skoro $PA \vdash \underline{\mathsf{samneg}}(\overline{n}, \overline{r})$, to $PA \vdash \overline{r} \underline{\leqslant} y \rightarrow \exists z\, (z \underline{\leqslant} y \wedge \underline{\mathsf{samneg}}(\overline{n}, z))$.
- Z faktu podanego w poprzednim wykładzie mamy: $PA \vdash y \underline{\leqslant} \overline{r} \vee \overline{r} \underline{\leqslant} y$.
- Z trzech powyższych faktów otrzymujemy: $PA \vdash \neg\, \underline{\mathsf{sam}}(\overline{n}, y) \vee \exists z\, (z \underline{\leqslant} y \wedge \underline{\mathsf{samneg}}(\overline{n}, z))$.
- Na mocy reguły generalizacji mamy: $PA \vdash \forall y\, (\neg\, \underline{\mathsf{sam}}(\overline{n}, y) \rightarrow \exists z\, (z \underline{\leqslant} y \wedge \underline{\mathsf{samneg}}(\overline{n}, z)))$.
- Otrzymaliśmy więc: $PA \vdash ros$, co (łącznie z przypuszczeniem dowodu nie wprost) przeczy założeniu o niesprzeczności PA.
- Ostatecznie, odrzucamy przypuszczenie dowodu nie wprost i mamy: $PA\ non \vdash \neg\, ros$.

Lemat przekątniowy

- Oba powyższe twierdzenia (oraz szereg dalszych) można udowodnić, odwołując się do pewnego wyniku dotyczącego ***dowodów przekątniowych***.
- W dalszym ciągu przyjmujemy we wszystkich twierdzeniach założenie: ***PA jest niesprzeczna***.

- **Lemat Przekątniowy**. *Dla dowolnej formuły języka PA* $\varphi(x)$ *o jednej zmiennej wolnej istnieje zdanie* ψ *tego języka takie, że:*
  $PA \vdash \psi \equiv \varphi(\overline{\ulcorner \psi \urcorner})$.

Tak więc, dla każdej własności (liczb) wyrażalnej w PA znajdziemy zdanie ψ, stwierdzające, że jego numer gödlowski $\ulcorner \psi \urcorner$ ma tę własność.

Dowód Lematu Przekątniowego

Dowód. Przypomnijmy, że dla termu α, zmiennej x oraz formuły ϕ mamy:

- $\mathrm{Sub}(\ulcorner\phi\urcorner, \ulcorner x\urcorner, \ulcorner\alpha\urcorner) = \ulcorner\phi(x/\alpha)\urcorner$ (= numer gödlowski formuły otrzymanej przez podstawienie termu α za zmienną x w formule ϕ).
- Sub jest funkcją rekurencyjną, a więc istnieje co najmniej jedna formuła języka PA, która ją reprezentuje w PA. Niech $\underline{\mathrm{Sub}}(x, y, u, v)$ będzie taką formułą.

- Niech $\mathrm{Subst}(x, y, z) = \mathrm{Sub}(x, y, \mathrm{num}(z))$ i niech $\underline{\mathrm{Subst}}$ będzie formułą reprezentującą w PA funkcję Subst.
- Rozważmy formułę *ref*(x) o postaci: $\forall y\ (\underline{\mathrm{Subst}}(x, \overline{2}, x, y) \rightarrow \varphi(y))$.
- W powyższym (i dalej) zakładamy, że x to zmienna x_1, y to zmienna x_2, z to zmienna x_3. Wtedy $\ulcorner x\urcorner = \ulcorner x_1\urcorner = 2$.

Dowód Lematu Przekątniowego

- Niech $m = \ulcorner ref(x) \urcorner$.
- Niech ψ będzie zdaniem $ref(\overline{m})$.
- Wtedy w PA można udowodnić równoważność następujących zdań (co daje dowód Lematu Przekątniowego):

- ψ
- $ref(\overline{m})$
- $\forall y\ (\underline{\mathrm{Subst}}(\overline{m}, \overline{2}, \overline{m}, y) \rightarrow \varphi(y))$
- $\forall y\ (\underline{\mathrm{Subst}}(\overline{\ulcorner ref(x) \urcorner}, \overline{2}, \overline{m}, y) \rightarrow \varphi(y))$
- $\varphi(\overline{\ulcorner ref(\overline{m}) \urcorner})$
- $\varphi(\overline{\ulcorner \psi \urcorner})$.

Lemat Przekątniowy: komentarz dydaktyczny

- Istniejące na mocy Lematu Przekątniowego zdanie ψ stwierdza samo o sobie, że (jego numer gödlowski) ma własność φ.
- Precyzyjne sformułowanie tego faktu stało się możliwe dzięki procedurze arytmetyzacji składni.

- Unikamy przy tym wszelkich niebezpieczeństw, które stwarzają zdania samozwrotne w językach etnicznych.

Przypominamy, że np. zdanie:

Zdanie napisane w tej ramce jest fałszywe.

prowadzi do antynomii. Powstaje ona na skutek pomieszania języka przedmiotowego i metajęzyka.

I Twierdzenie Gödla jako konsekwencja LP

- Niech φ_G będzie zdaniem takim, że $PA \vdash \varphi_G \equiv \neg \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner})$.
- Zdanie φ_G istnieje na mocy Lematu Przekątniowego.

I Twierdzenie Gödla.

Niech φ_G będzie określonym powyżej zdaniem. Wtedy:

1. *PA non $\vdash \varphi_G$.*
2. *Jeżeli dla dowolnego zdania ψ zachodzi implikacja:*
   (∗) jeśli $PA \vdash \underline{Tw}(\overline{\ulcorner \psi \urcorner})$, to $PA \vdash \psi$,
   to PA non $\vdash \neg\varphi_G$.

I Twierdzenie Gödla jako konsekwencja LP

Dowód.

- Dla dowodu nie wprost punktu 1, przypuśćmy, że $PA \vdash \varphi_G$.
- Wtedy $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner})$, a stąd $PA \vdash \neg\varphi_G$.
- To oznacza, że PA jest sprzeczna, wbrew założeniu.
- Przypuszczenie $PA \vdash \varphi_G$ trzeba więc odrzucić. Ostatecznie, $PA\ non \vdash \varphi_G$.

- Dla dowodu nie wprost punktu 2, przypuśćmy, że $PA \vdash \neg\varphi_G$.
- Wtedy $PA \vdash \neg\neg\, \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner})$, a stąd $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner})$.
- Na mocy $(*)$ mamy wtedy $PA \vdash \varphi_G$, wbrew 1.
- Przypuszczenie dowodu nie wprost zatem odrzucamy i mamy ostatecznie $PA\ non \vdash \neg\varphi_G$.

I Twierdzenie Gödla jako konsekwencja LP: komentarz

- Nie zakładano ω-niesprzeczności PA, a tylko jej niesprzeczność oraz warunek $(*)$.
- Z ω-niesprzeczności wynika warunek $(*)$.
- W dowodzie wykorzystywano warunek $(*)$ tylko dla zdania φ_G.
- Zdanie φ_G ma postać: $\neg\exists x\ \underline{Dow}(x, \overline{\ulcorner\varphi_G\urcorner})$.
- Zdanie φ_G jest zatem równoważne zdaniu ogólnemu: $\forall x \neg\ \underline{Dow}(x, \overline{\ulcorner\varphi_G\urcorner})$.
- Można pokazać, że dla dowolnej liczby naturalnej n zachodzi $PA \vdash \neg\ \underline{Dow}(x, \overline{n})$.

Warunki dowodliwości

- Ponieważ relacja Dow jest rekurencyjna, więc istnieje co najmniej jedna formuła języka PA, która ją mocno reprezentuje. Niech $\underline{\mathrm{Dow}}(x, y)$ będzie taką formułą (o dwóch zmiennych wolnych).
- Niech $\underline{\mathrm{Tw}}(y)$ będzie formułą $\exists x\ \underline{\mathrm{Dow}}(x, y)$.
- Wtedy dla dowolnej formuły ψ języka PA: jeśli $PA \vdash \psi$, to $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \psi \urcorner})$.

Warunkami dowodliwości nazywamy następujące trzy warunki, dla dowolnych zdań φ i ψ:

- (D1) Jeśli $PA \vdash \varphi$, to $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi \urcorner})$.
- (D2) $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi \urcorner}) \ \rightarrow\ \underline{\mathrm{Tw}}(\overline{\ulcorner \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi \urcorner}) \urcorner})$.
- (D3) $PA \vdash (\underline{\mathrm{Tw}}(\overline{\ulcorner \varphi \urcorner}) \wedge \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi \rightarrow \psi \urcorner})) \rightarrow \underline{\mathrm{Tw}}(\overline{\ulcorner \psi \urcorner})$.

Warunki dowodliwości

- Jak się okazuje, postać formuły mocno reprezentującej relację Tw jest istotna w dowodach niektórych twierdzeń o PA. To samo dotyczy też postaci formuły mocno reprezentującej relację Dow. Nie możemy mocno reprezentować relacji dowodliwości całkiem dowolnie, chcąc otrzymać te twierdzenia.
- Warunki dowodliwości są właśnie pewnymi ograniczeniami nakładanymi na mocną reprezentację relacji dowodliwości w PA.

- Dowodliwość w PA można interpretować jako modalność.
- Otrzymujemy wtedy pewną logikę modalną, ***logikę dowodliwości*** (***logikę Gödla-Löba***).
- Warunki dowodliwości przekładają się na aksjomaty tej logiki.

II Twierdzenie Gödla (niedowodliwość niesprzeczności)

- Jak można ***wyrazić*** w PA niesprzeczność PA? Wystarczy zapisać, że w PA nie można dowieść sprzeczności.
- Przez Con_{PA} rozumiemy formułę: $\neg\, \underline{\mathrm{Tw}}(\overline{\ulcorner \underline{0} \doteq \overline{1} \urcorner})$.
- Wtedy Con_{PA} wyraża niesprzeczność PA.
- Wszystkie zdania sprzeczne są równoważne na gruncie PA.

II Twierdzenie Gödla. *(Niedowodliwość niesprzeczności PA w PA.)*
Przy założeniach (D1)–(D3): PA non $\vdash Con_{PA}$.

Pokażemy, że $PA \vdash \varphi_G \equiv Con_{PA}$. Na mocy I Twierdzenia Gödla dostaniemy wtedy: *PA non* $\vdash Con_{PA}$.

Dowód II Twierdzenia Gödla

Dowód.

- Przypominamy, że na mocy Lematu Przekątniowego istnieje zdanie φ_G takie, że $PA \vdash \varphi_G \equiv \neg\, \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner})$ (czyli zdanie gödlowskie, stwierdzające swoją własną niedowodliwość w PA).
- Ponieważ dla wszystkich ψ mamy: $PA \vdash (\underline{0} \doteq \overline{1} \rightarrow \psi)$, więc $PA \vdash (\underline{0} \doteq \overline{1} \rightarrow \varphi_G)$.
- Na mocy (D1): $PA \vdash\ \underline{\mathrm{Tw}}(\overline{\ulcorner \underline{0} \doteq \overline{1} \rightarrow \varphi_G \urcorner})$.
- Na mocy (D3): $PA \vdash\ \underline{\mathrm{Tw}}(\overline{\ulcorner \underline{0} \doteq \overline{1} \urcorner}) \rightarrow\ \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner})$.
- Przez kontrapozycję: $PA \vdash\ \neg\, \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner}) \rightarrow \neg\, \underline{\mathrm{Tw}}(\overline{\ulcorner \underline{0} \doteq \overline{1} \urcorner})$.
- Z definicji φ_G mamy: $PA \vdash \varphi_G \rightarrow \neg\, \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner})$.
- Z powyższego mamy: $PA \vdash \varphi_G \rightarrow \neg\, \underline{\mathrm{Tw}}(\overline{\ulcorner \underline{0} \doteq \overline{1} \urcorner})$, czyli $PA \vdash \varphi_G \rightarrow Con_{PA}$. Trzeba jeszcze udowodnić implikację odwrotną.

Dowód II Twierdzenia Gödla

- Na mocy (D2): $(\dagger)$ $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner}) \to \underline{\mathrm{Tw}}(\overline{\ulcorner \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner}) \urcorner})$.
- Z definicji φ_G mamy: $PA \vdash \varphi_G \to \neg\, \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner})$.
- Przez kontrapozycję: $PA \vdash \neg\, \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner}) \to \neg \varphi_G$.
- Na mocy (D1) oraz (D3) otrzymujemy odpowiednio:
  $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner}) \to \neg \varphi_G \urcorner})$
  $(\ddagger)$ $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner}) \urcorner}) \to \underline{\mathrm{Tw}}(\overline{\ulcorner \neg \varphi_G \urcorner})$.
- Z $(\dagger)$ oraz $(\ddagger)$ mamy: $(\heartsuit)$ $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner}) \to \underline{\mathrm{Tw}}(\overline{\ulcorner \neg \varphi_G \urcorner})$.
- Mamy także: $PA \vdash \varphi_G \to (\neg \varphi_G \to (\varphi_G \wedge \neg \varphi_G))$.
- Na mocy (D1) oraz (D3) mamy:
  $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner}) \to (\underline{\mathrm{Tw}}(\overline{\ulcorner \neg \varphi_G \urcorner}) \to \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \wedge \neg \varphi_G \urcorner}))$.
- Mamy więc też:
  $(\clubsuit)$ $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \neg \varphi_G \urcorner}) \to (\underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner}) \to \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \wedge \neg \varphi_G \urcorner}))$.

Dowód II Twierdzenia Gödla

- Podstawiamy w prawie KRZ $(p \to (q \to r)) \to ((q \to p) \to (q \to r))$: $\underline{\mathrm{Tw}}(\overline{\ulcorner \neg\varphi_G \urcorner})$ za p; $\underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner})$ za q; $\underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \wedge \neg\varphi_G \urcorner})$ za r i otrzymujemy:
  $(\spadesuit)$ $PA \vdash (\underline{\mathrm{Tw}}(\overline{\ulcorner \neg\varphi_G \urcorner}) \to (\underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner}) \to \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \wedge \neg\varphi_G \urcorner}))) \to$
  $((\underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner}) \to \underline{\mathrm{Tw}}(\overline{\ulcorner \neg\varphi_G \urcorner})) \to (\underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner}) \to \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \wedge \neg\varphi_G \urcorner})))$
- Z $(\spadesuit)$, $(\clubsuit)$ oraz $(\heartsuit)$ dostajemy:
  $(\diamondsuit)$ $\underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner}) \to \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \wedge \neg\varphi_G \urcorner})$.
- Ponieważ $PA \vdash (\varphi_G \wedge \neg\varphi_G) \equiv (\underline{0} \doteq \bar{1})$, więc $PA \vdash (\varphi_G \wedge \neg\varphi_G) \to (\underline{0} \doteq \bar{1})$.
- Na mocy (D1) i (D3) mamy:
  $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \wedge \neg\varphi_G \urcorner}) \to \underline{\mathrm{Tw}}(\overline{\ulcorner \underline{0} \doteq \bar{1} \urcorner})$.
- A stąd oraz z $(\diamondsuit)$ mamy: $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner}) \to \underline{\mathrm{Tw}}(\overline{\ulcorner \underline{0} \doteq \bar{1} \urcorner})$.

Dowód II Twierdzenia Gödla

- Przez kontrapozycję mamy: $PA \vdash \neg \underline{\mathrm{Tw}}(\overline{\ulcorner \underline{0} \doteq \overline{1} \urcorner}) \rightarrow \neg \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi_G \urcorner})$.
- Na mocy definicji zdania φ_G oraz formuły Con_{PA} otrzymujemy stąd potrzebną implikację: $PA \vdash Con_{PA} \rightarrow \varphi_G$.
- Udowodniliśmy obie implikacje: $PA \vdash \varphi_G \rightarrow Con_{PA}$ oraz $PA \vdash Con_{PA} \rightarrow \varphi_G$, a więc mamy: $PA \vdash \varphi_G \equiv Con_{PA}$.
- Ponieważ (I Twierdzenie Gödla) mamy $PA\ non \vdash \varphi_G$, więc mamy również: $PA\ non \vdash Con_{PA}$, co kończy dowód II Twierdzenia Gödla.

Przy założeniach (D1)–(D3) każde zdanie wyrażające swoją własną niedowodliwość jest równoważne zdaniu Con_{PA} wyrażającemu niesprzeczność PA. Tak więc, przy tych założeniach dowolne dwa zdania gödlowskie są dowodliwie równoważne na gruncie PA: jeśli $PA \vdash \varphi \equiv \neg \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi \urcorner})$ oraz $PA \vdash \psi \equiv \neg \underline{\mathrm{Tw}}(\overline{\ulcorner \psi \urcorner})$, to $PA \vdash \varphi \equiv \psi$.

Twierdzenie Löba

Twierdzenie Löba.

Dla dowolnego zdania φ języka PA następujące warunki są równoważne:

1. $PA \vdash \underline{Tw}(\overline{\ulcorner \varphi \urcorner}) \rightarrow \varphi$
2. $PA \vdash \varphi$.

- Niech ψ_H będzie ***zdaniem Henkina*** (zdaniem stwierdzającym swoją własną dowodliwość), czyli takim, iż: $PA \vdash \psi_H \equiv \underline{\mathrm{Tw}}(\overline{\ulcorner \psi_H \urcorner})$.
- Z Twierdzenia Löba wynika, że zdanie Henkina jest dowodliwe w PA: $PA \vdash \psi_H$.

Dowód Twierdzenia Löba

Dowód. Implikacja $2 \Rightarrow 1$ jest oczywista, na mocy aksjomatu $p \to (q \to p)$.

Dowód implikacji $1 \Rightarrow 2$.

- Załóżmy, że $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner\varphi\urcorner}) \to \varphi$.
- Na mocy Lematu Przekątniowego istnieje zdanie ψ takie, że: $PA \vdash (\psi \equiv (\underline{\mathrm{Tw}}(\overline{\ulcorner\psi\urcorner}) \to \varphi))$.
- Na mocy warunków (D1) oraz (D3) mamy:
  $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner\psi\urcorner}) \equiv \underline{\mathrm{Tw}}(\overline{\ulcorner\underline{\mathrm{Tw}}(\overline{\ulcorner\psi\urcorner}) \to \varphi)\urcorner})$
  $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner\psi\urcorner}) \to (\underline{\mathrm{Tw}}(\overline{\ulcorner\underline{\mathrm{Tw}}(\overline{\ulcorner\psi\urcorner})\urcorner}) \to \underline{\mathrm{Tw}}(\overline{\ulcorner\varphi\urcorner}))$.
- Na mocy warunku (D2) mamy:
  $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner\psi\urcorner}) \to \underline{\mathrm{Tw}}(\overline{\ulcorner\underline{\mathrm{Tw}}(\overline{\ulcorner\psi\urcorner})\urcorner})$.

Dowód Twierdzenia Löba

- Korzystamy teraz z Prawa Fregego:
  $(p \to (q \to r)) \to ((p \to q) \to (p \to r))$ i otrzymujemy:
  $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \psi \urcorner}) \to \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi \urcorner})$.
- Stąd oraz z założenia $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \varphi \urcorner}) \to \varphi$ mamy:
  $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \psi \urcorner}) \to \varphi$.
- Ponieważ PA dowodzi równoważności ψ z $\underline{\mathrm{Tw}}(\overline{\ulcorner \psi \urcorner}) \to \varphi$, więc $PA \vdash \psi$.
- Z $PA \vdash \psi$ otrzymujemy, na mocy (D1): $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner \psi \urcorner})$.
- Na mocy reguły odrywania mamy ostatecznie: $PA \vdash \varphi$.

- Inny jeszcze dowód Twierdzenia Löba można otrzymać wykorzystując II Twierdzenie Gödla.
- Z Twierdzenia Löba wynika, że każde dwa zdania Henkina są równoważne na gruncie PA.

Twierdzenie Tarskiego

Twierdzenie Tarskiego. *(Niedefiniowalność prawdy arytmetycznej w PA.) Nie istnieje formuła* $alf(x)$ *języka PA taka, że dla dowolnego zdania* φ *tego języka:* $PA \vdash \varphi \equiv alf(\overline{\ulcorner \varphi \urcorner})$.

- Przypuśćmy, dla dowodu nie wprost, że istnieje taka formuła alf.
- Na mocy Lematu Przekątniowego istnieje zdanie ψ takie, że $PA \vdash \psi \equiv \neg\, alf(\overline{\ulcorner \psi \urcorner})$.
- Ponieważ z przypuszczenia dowodu nie wprost mamy: $PA \vdash \psi \equiv alf(\overline{\ulcorner \psi \urcorner})$, więc otrzymujemy $PA \vdash alf(\overline{\ulcorner \psi \urcorner}) \equiv \neg\, alf(\overline{\ulcorner \psi \urcorner})$.
- To oznacza, że PA jest sprzeczna, wbrew założeniu.
- Przypuszczenie dowodu nie wprost musimy więc odrzucić.
- Ostatecznie, nie istnieje taka formuła alf.

Twierdzenie Tarskiego

- Aksjomaty PA są prawdziwe w modelu standardowym $\mathfrak{N}_0$.
- Wszystkie twierdzenia PA są prawdziwe w modelu standardowym.

- Konsekwencją Twierdzenia Tarskiego jest zatem to, że nie istnieje formuła *alf* języka PA taka, iż dla dowolnego zdania φ tego języka: $\mathfrak{N}_0 \models \varphi$ dokładnie wtedy, gdy $\mathfrak{N}_0 \models \mathit{alf}(\overline{\ulcorner\varphi\urcorner})$.
- To z kolei oznacza, że w języku PA nie istnieje definicja zbioru tych zdań tego języka, które są prawdziwe w modelu standardowym.

Można udowodnić, że definicja tego zbioru wykracza poza (omówioną w poprzednim wykładzie) hierarchię arytmetyczną.

Twierdzenie Rossera jako konsekwencja LP

Jeśli $\underline{\mathrm{Dow}}$ jest formułą mocno reprezentującą w PA relację Dow, to niech $\underline{\mathrm{Dow}^R}$ będzie formułą:

$\underline{\mathrm{Dow}}(x,y) \wedge \forall z \underline{\leqslant} x \forall w\ (\underline{\mathrm{Dow}}(z,w) \rightarrow \neg\ \underline{\mathrm{Neg}}(w,y) \wedge \neg\ \underline{\mathrm{Neg}}(y,w))$,

gdzie $\underline{\mathrm{Neg}}$ jest formułą mocno reprezentującą w PA rekurencyjną relację Neg taką, że:

$\mathrm{Neg}(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$ dokładnie wtedy, gdy φ jest tożsama z $\neg\psi$.

Formuła $\underline{\mathrm{Dow}^R}(x,y)$ stwierdza zatem, że:

- x jest numerem gödlowskim dowodu formuły o numerze gödlowskim y oraz
- nie istnieje dowód negacji formuły o numerze gödlowskim y, który miałby numer gödlowski mniejszy od x.

Twierdzenie Rossera jako konsekwencja LP

- Niech $\underline{\mathrm{Tw}^R}(y)$ będzie formułą $\exists x\ \underline{\mathrm{Dow}^R}(x, y)$.
- Niech Con^R_{PA} będzie zdaniem $\neg\ \underline{\mathrm{Tw}^R}(\overline{\ulcorner \underline{0} \doteq \overline{1} \urcorner})$.

- Dla dowolnego φ: $PA \vdash \neg(\underline{\mathrm{Tw}^R}(\overline{\ulcorner \varphi \urcorner}) \wedge \underline{\mathrm{Tw}^R}(\overline{\ulcorner \neg\varphi \urcorner}))$.
- W szczególności: $PA \vdash \underline{\mathrm{Tw}^R}(\overline{\ulcorner \neg(\underline{0} \doteq \overline{1}) \urcorner}) \rightarrow \neg\ \underline{\mathrm{Tw}^R}(\overline{\ulcorner \underline{0} \doteq \overline{1} \urcorner})$.
- Ponieważ $PA \vdash \neg(\underline{0} \doteq \overline{1})$, więc $PA \vdash \underline{\mathrm{Tw}^R}(\overline{\ulcorner \neg(\underline{0} \doteq \overline{1}) \urcorner})$.
- W konsekwencji: $PA \vdash \neg\ \underline{\mathrm{Tw}^R}(\overline{\ulcorner \underline{0} \doteq \overline{1} \urcorner})$, czyli $PA \vdash Con^R_{PA}$.
- Formuła $\underline{\mathrm{Tw}^R}$ nie może zatem spełniać warunków dowodliwości (D1)–(D3). Dowodzi się, że $\underline{\mathrm{Tw}^R}$ nie spełnia (D2).
- Formuła Con^R_{PA} wyraża własność niesprzeczności PA, ale ***dowodliwą*** na gruncie PA.

Twierdzenie Rossera jako konsekwencja LP

Na mocy Lematu Przekątniowego istnieje zdanie φ_R takie, że:
$PA \vdash \varphi_R \equiv \neg \, \underline{\mathrm{Tw}^R}(\overline{\ulcorner \varphi_R \urcorner})$.

Twierdzenie.

Niech φ_R będzie określonym powyżej zdaniem. Wtedy:

1. *PA non $\vdash \varphi_R$*
2. *PA non $\vdash \neg\varphi_R$.*

- **Dowód punktu 1**. Jeśli PA jest niesprzeczna, to formuły $\underline{\mathrm{Dow}}$ oraz $\underline{\mathrm{Dow}^R}$ mocno reprezentują w PA tę samą relację.
- Zachodzi zatem warunek (D1) dla $\underline{\mathrm{Tw}^R}$, czyli: jeśli $PA \vdash \varphi$, to $PA \vdash \underline{\mathrm{Tw}^R}(\overline{\ulcorner \varphi \urcorner})$, dla wszystkich φ.
- Tak samo jak w (drugim) dowodzie I Twierdzenia Gödla pokazujemy, że: PA *non* $\vdash \varphi_R$.

Twierdzenie Rossera jako konsekwencja LP

Dowód punktu 2.

- Dla dowodu nie wprost przypuśćmy, że $PA \vdash \neg\varphi_R$.
- Niech d będzie numerem gödlowskim jakiegoś dowodu zdania $\neg\varphi_R$.
- Ponieważ $Dow(d, \ulcorner\neg\varphi_R\urcorner)$, a $\underline{\mathrm{Dow}^R}$ mocno reprezentuje w PA relację Dow^R (czyli również relację Dow), więc $PA \vdash \underline{\mathrm{Dow}^R}(\overline{d}, \overline{\ulcorner\neg\varphi_R\urcorner})$.
- Na mocy definicji zdania φ_R oraz przypuszczenia dowodu nie wprost mamy: $PA \vdash \underline{\mathrm{Tw}^R}(\overline{\ulcorner\varphi_R\urcorner})$.
- To oznacza, że: $PA \vdash \exists x\, (\underline{\mathrm{Dow}}(x, \overline{\ulcorner\varphi_R\urcorner}) \wedge \forall z \underline{\leqslant} x \forall w\, (\underline{\mathrm{Dow}}(z, w) \rightarrow (\neg\underline{\mathrm{Neg}}(w, \overline{\ulcorner\varphi_R\urcorner}) \wedge \neg\underline{\mathrm{Neg}}(\overline{\ulcorner\varphi_R\urcorner}, w))))$.
- Oznaczmy przez $\gamma(x)$ podformułę powyższej formuły, będącą zasięgiem kwantyfikatora $\exists x$. Mamy wtedy:

Twierdzenie Rossera jako konsekwencja LP

- $PA \vdash \exists x\, ((x \underline{\leqslant} \overline{d} \vee x \underline{\geq} \overline{d}) \wedge \gamma(x))$.
- $PA \vdash \exists x\, ((x \underline{\leqslant} \overline{d} \wedge \gamma(x)) \vee (x \underline{\geq} \overline{d} \wedge \gamma(x)))$.
- $PA \vdash \exists x\, (x \underline{\leqslant} \overline{d} \wedge \gamma(x)) \vee \exists x\, (x \underline{\geq} \overline{d} \wedge \gamma(x))$.
- Ponieważ $PA \vdash \underline{\mathrm{Dow}^R}(\overline{d}, \overline{\ulcorner \neg \varphi_R \urcorner})$, więc drugi składnik powyższej alternatywy jest sprzeczny.
- Mamy więc: $PA \vdash \exists x\, (x \underline{\leqslant} \overline{d} \wedge \gamma(x))$
- Ponieważ $PA \vdash (x \underline{\leqslant} \overline{d} \equiv (x \doteq \underline{0} \vee x \doteq \overline{1} \vee \ldots \vee x \doteq \overline{d}))$, więc mamy: $PA \vdash \underline{\mathrm{Dow}}(\underline{0}, \overline{\ulcorner \varphi_R \urcorner}) \vee \underline{\mathrm{Dow}}(\overline{1}, \overline{\ulcorner \varphi_R \urcorner}) \vee \ldots \vee \underline{\mathrm{Dow}}(\overline{d}, \overline{\ulcorner \varphi_R \urcorner})$.
- To jest sprzeczne z $PA \vdash \underline{\mathrm{Dow}^R}(\overline{d}, \overline{\ulcorner \neg \varphi_R \urcorner})$.
- Przypuszczenie dowodu nie wprost musimy więc odrzucić i mamy ostatecznie: $PA\ non \vdash \neg \varphi_R$.

Inne zdania samozwrotne

Twierdzenie. *Załóżmy, że PA niesprzeczna oraz że formuła* $\underline{Tw}$ *spełnia warunki (D2), (D3) i warunek (D1′): PA* $\vdash \varphi$ *dokładnie wtedy, gdy PA* $\vdash \underline{Tw}(\overline{\ulcorner\varphi\urcorner})$. *Wtedy:*

1. *Jeśli* φ *takie, że PA* $\vdash \varphi \equiv (\neg\, \underline{Tw}(\overline{\ulcorner\varphi\urcorner}) \wedge \neg\, \underline{Tw}(\overline{\ulcorner\neg\varphi\urcorner}))$, *to PA* $\vdash \neg\varphi$ *oraz PA non* $\vdash \varphi$. *[Tu* φ *stwierdza własną niesprzeczność.]*
2. *Jeśli* φ *takie, że PA* $\vdash \varphi \equiv (\underline{Tw}(\overline{\ulcorner\varphi\urcorner}) \vee \underline{Tw}(\overline{\ulcorner\neg\varphi\urcorner}))$, *to PA* $\vdash \varphi$. *[Tu* φ *stwierdza własną rozstrzygalność.]*
3. *Jeśli* φ *takie, że PA* $\vdash \varphi \equiv \neg\underline{Tw}(\overline{\ulcorner\neg\varphi\urcorner})$, *to PA* $\vdash \neg\varphi$ *(oraz PA non* $\vdash \varphi$*). [Tu* φ *stwierdza własną niesprzeczność z PA.]*
4. *Jeśli* φ *takie, że PA* $\vdash \varphi \equiv \underline{Tw}(\overline{\ulcorner\neg\varphi\urcorner})$, *to PA non* $\vdash \varphi$ *oraz PA non* $\vdash \neg\varphi$. *[Tu* φ *stwierdza własną sprzeczność z PA.]*

Inne zdania samozwrotne

Zarys dowodu.

Każde ze zdań wymienionych w twierdzeniu istnieje na mocy Lematu Przekątniowego.

1. Na mocy definicji φ: $PA \vdash \varphi \rightarrow \neg\underline{\mathrm{Tw}}(\overline{\ulcorner\neg\varphi\urcorner})$, a przez kontrapozycję: $PA \vdash \underline{\mathrm{Tw}}(\overline{\ulcorner\neg\varphi\urcorner}) \rightarrow \neg\varphi$. Na mocy Twierdzenia Löba mamy: $PA \vdash \neg\varphi$. Z niesprzeczności PA: $PA\ non \vdash \varphi$.
2. To konsekwencja poprzedniego punktu.
3. Na mocy definicji φ zdanie $\neg\varphi$ jest zdaniem Henkina, a zatem $PA \vdash \neg\varphi$.
4. Na mocy definicji φ zdanie $\neg\varphi$ jest zdaniem gödlowskim. Tak więc, na mocy I Twierdzenia Gödla: $PA\ non \vdash \varphi$ oraz $PA\ non \vdash \neg\varphi$.

Istotna niezupełność arytmetyki PA

Z podanych w tym wykładzie twierdzeń wynika, że jeśli arytmetyka PA jest niesprzeczna, to jest: niezupełna oraz nierozstrzygalna.

- Przez ***rekurencyjne rozszerzenie arytmetyki PA*** rozumiemy każdą teorię pierwszego rzędu, która jest rozszerzeniem PA o rekurencyjny zbiór aksjomatów i której zbiór numerów gödlowskich symboli pozalogicznych jest rekurencyjny.
- Teoria T (w której możliwa jest arytmetyzacja składni) jest ***istotnie niezupełna***, jeśli T jest (niesprzeczna i) niezupełna oraz każde jej niesprzeczne rozszerzenie rekurencyjne jest niezupełne.

Jeśli arytmetyka PA jest niesprzeczna, to jest istotnie niezupełna.

Uwagi końcowe

Twierdzenia Gödla uważa się za najbardziej doniosłe dokonanie w logice XX wieku. Istnieje na jego temat olbrzymia literatura. Na stronie tych wykładów umieszczono kilkadziesiąt plików dotyczących tej problematyki, znalezionych w sieci.

Szczególnie polecamy lekturę dwóch monografii:

- Krajewski, S. 2003. *Twierdzenie Gödla i jego interpretacje filozoficzne. Od mechanicyzmu do postmodernizmu.* Wydawnictwo Instytutu Filozofii i Socjologii PAN, Warszawa.
- Murawski, R. 2000^3. *Funkcje rekurencyjne i elementy metamatematyki. Problemy zupełności, rozstrzygalności, twierdzenia Gödla.* Wydawnictwo Naukowe UAM, Poznań. [Na tej monografii oparte są niniejsze wykłady: 7, 8 i 9.]

Wykorzystywana literatura

- Barwise, J. (ed.) 1977. *Handbook of mathematical logic.* North-Holland Publishing Company, Amsterdam, New York Oxford.
- Bell, J.L., Machover, M. 1977. *A course in mathematical logic.* North Holland Publishing Company, Amsterdam, New York, Oxford.
- Cutland, N. 1980. *Computability. An introduction to recursive function theory.* Cambridge University Press, Cambridge.
- Hinman, P.G. 2005. *Fundamentals of Mathematical Logic.* A K Peters, Wellesley.
- Krajewski, S. 2003. *Twierdzenie Gödla i jego interpretacje filozoficzne. Od mechanicyzmu do postmodernizmu.* Wydawnictwo Instytutu Filozofii i Socjologii PAN, Warszawa.
- Mendelson, E. 1997. *Introduction to mathematical logic.* Chapman & Hall, London.

Wykorzystywana literatura

- Murawski, R. 2000[3]. *Funkcje rekurencyjne i elementy metamatematyki. Problemy zupełności, rozstrzygalności, twierdzenia Gödla.* Wydawnictwo Naukowe UAM, Poznań.
- Odifreddi, P.G. 1989. *Classical recursion theory.* North-Holland Publishing Company, Amsterdam.
- Shoenfield, J.R. 1967. *Mathematical logic.* Addison-Wesley, Reading, MA.
- Smorynski, C. 1977. *The incompleteness theorems.* W: J. Barwise (ed.) 1977. *Handbook of mathematical logic.* North-Holland Publishing Company, Amsterdam, New York Oxford, 821–866.
- Smullyan, R. 1992. *Gödel's incompleteness theorems.* The Clarendon Press, Oxford University Press, New York.
- Smullyan, R. 1993. *Recursion theory for metamathematics.* Oxford University Press.

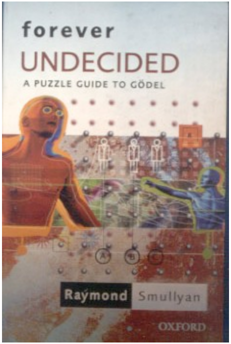
Plan

Pokazujemy kilka twierdzeń z naszego tłumaczenia książki Raymonda Smullyana ***Forever Undecided. A Puzzle Guide to Gödel***, które ukazało się w 2007 roku nakładem *Książki i Wiedzy*, pod tytułem ***Na Zawsze Nierozstrzygnięte. Zagadkowy Przewodnik Po Twierdzeniach Gödla.***

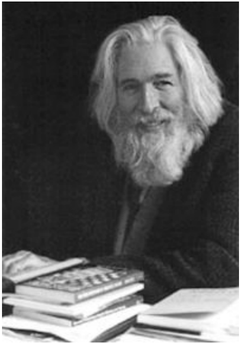
Obok zagadek o Rycerzach (mówiących zawsze prawdę) oraz Łotrach (mówiących zawsze fałsz), książka zawiera zagadki logiczne, w których w formie popularnej przedstawia się ***logikę dowodliwości*** (***logice Gödla-Löba***).

Proszę traktować niniejszy dodatek jako rozrywkę. Chciałbym przede wszystkim zwrócić uwagę na mistrzostwo Smullyana w popularyzowaniu wiedzy logicznej.

Forever Undecided



Forever Undecided



Raymond Smullyan

Kilka książek o logice modalnej

- Boolos, G. 1993. *The Logic of Provability.* Cambridge University Press.
- Smullyan, R. 1992. *Gödel's Incompleteness Theorems.* Oxford University Press.

- Jacek Hawranek: *Aspekty algebraiczne systemu modalnego Gödla–Löba.* Wydawnictwo Uniwersytetu Wrocławskiego, Wrocław, 1994.
- Andrzej Indrzejczak: *Hybrydowe systemy dedukcyjne w logikach modalnych.* Wydawnictwo Uniwersytetu Łódzkiego, Łódź, 2006.
- Jerzy Perzanowski: *Logiki modalne a filozofia.* Uniwersytet Jagielloński, Rozprawy Habilitacyjne nr 156, Kraków, 1989.
- Kazimierz Świrydowicz: *Podstawy logiki modalnej.* Wydawnictwo Naukowe UAM, Poznań, 2004.

Książki z zagadkami logicznymi Raymonda Smullyana

- *Jaki jest tytuł tej książki? Tajemnica Drakuli, zabawy i łamigłówki logiczne.* Warszawa 1993. Przełożył: Bohdan Chwedeńczuk. Trzy wydania polskie.
- *Dama czy tygrys oraz inne zagadki logiczne.* Warszawa 1995, 2004. Przełożył: Bohdan Chwedeńczuk.
- *Szatan, Cantor i nieskończoność oraz inne łamigłówki.* Warszawa 1998. Przełożyli z angielskiego: Anna i Krzysztof Wójtowicz.
- *Przedrzeźniać Przedrzeźniacza. Oraz Inne Zagadki Logiczne Łącznie z Zadziwiającą Przygodą w Krainie Logiki Kombinatorycznej.* Warszawa 2007. Przekład z języka angielskiego: Jerzy Pogonowski.
- *Forever Undecided. A Puzzle Guide to Gödel.* Oxford University Press, 1988. Z angielskiego przełożył Jerzy Pogonowski. Ukazało się w 2007 jako: *Na zawsze nierozstrzygnięte. Zagadkowy Przewodnik po Twierdzeniach Gödla*.

Aby cieszyć się wędrówką po Szczytach Metalogiki. . .

. . . najpierw musimy ominąć przepaście.

Plan na dziś

Plan na dziś:

- **Systemy przekonań.** Kto jest prostaczkiem logicznym?
- **Poziomy samoświadomości.** Kto jest szczęściarzem epistemicznym?
- **II Twierdzenie Gödla.**
  Czy możesz wiedzieć, że twój system przekonań jest niesprzeczny, bez popadnięcia przy tym w sprzeczność?
- **Twierdzenie Löba i samospełniające się przekonania.**
  Kiedy *wishful thinking* ma wartość?
- **I Twierdzenie Gödla i Twierdzenie Rossera (o niezupełności).**
  Czy łatwy jest los *Besserwissera*?
- **Twierdzenie Tarskiego.**
  Czy *dictum*: ***Doctrina multiplex, veritas una!*** jest mrzonką?

Kurt Gödel

Kurt Gödel

Logik i Fizyk

Logik rozwiązał równania Fizyka, otrzymując Rotacyjny Model Wszechświata, w którym możliwe są podróże w czasie. Z rozwiązania tego korzystał ostatnio JM Rektor UAM (Zarządzenie Rektora nr 72/2006/2007 z dnia 15 III 2007 roku).

Systemy przekonań

Notacja. Operatory doksastyczne i epistemiczne to np.:

- $\boldsymbol{B}$ — zdanie $\boldsymbol{B}p$ czytamy: *(rozważany podmiot) wierzy, że p*;
- $\boldsymbol{K}$ — zdanie $\boldsymbol{K}p$ czytamy *(rozważany podmiot) wie, że p*.

(gdzie p jest dowolnym zdaniem rozważanego języka). Zwykle zakłada się, że $\boldsymbol{K}p \equiv (p \wedge \boldsymbol{B}p)$.

Systemy epistemiczne są interesujące same przez się — w opisie systemów przekonań, w szczególności: racjonalnych świadomych przekonań. Mają one także interesującą i ważną interpretację metalogiczną: $\boldsymbol{B}p$ można interpretować jako *zdanie p jest dowodliwe w arytmetyce PA*.

Uwaga. Angielski termin reasoner oddaję przez polski neologizm myślak.

Systemy przekonań

Przypuśćmy, że jesteś racjonalną, samoświadomą Istotą. Jak to przypuszczenie przełożyć na język logiki epistemicznej? Oto propozycja. Nazwiemy szczęściarzem epistemicznym każdą osobę S, której system przekonań spełnia warunki:

- (1a) S wierzy we wszystkie tautologie klasycznego rachunku zdań;
- (1b) system przekonań S jest domknięty na regułę *modus ponens*: jeśli S wierzy w p oraz wierzy w $p \rightarrow q$, to wierzy także w q;
- (2) dla dowolnych p oraz q, S wierzy w $(\boldsymbol{B}p \wedge \boldsymbol{B}(p \rightarrow q)) \rightarrow \boldsymbol{B}q$;
- (3) dla dowolnego p, jeśli S wierzy w p, to wierzy w $\boldsymbol{B}p$;
- (4) dla dowolnego p, S wierzy w $\boldsymbol{B}p \rightarrow \boldsymbol{B}\boldsymbol{B}p$.

Uwaga: rozważamy tylko osoby, które albo zawsze mówią prawdę, albo zawsze mówią fałsz.

Poziomy samoświadomości

Każdą osobę, która spełnia jedynie warunki (1a) i (1b) nazwiemy prostaczkiem logicznym. Zatem, jeśli S jest prostaczkiem logicznym, to jego/jej system przekonań zawiera klasyczną logikę zdaniową, ale S może być tego nieświadom(a).

Powiemy, że osoba S jest:

- normalna, gdy jeśli wierzy w p, to wierzy też w $\boldsymbol{B}p$;
- regularna, gdy jeśli wierzy w $p \rightarrow q$, to wierzy też w $\boldsymbol{B}p \rightarrow \boldsymbol{B}q$;
- sprzeczna, gdy do jej systemu przekonań należy jakaś para zdań wzajem sprzecznych, lub — co na jedno wychodzi — *fałsz logiczny*, który oznaczamy przez $\perp$.

Uwaga. Może bardziej właściwe byłoby mówienie o własnościach systemów przekonań, a nie osób.

Poziomy samoświadomości

Można udowodnić, że: $(*)$ dowolny szczęściarz epistemiczny S wie, że jeśli uwierzy w jakieś zdanie p oraz w jego negację $\neg p$, to stanie się sprzeczny.

O szczęściarzach epistemicznych można udowodnić wiele innych ciekawych rzeczy. Nie wszystkie z nich będą nam dalej potrzebne. Dodajmy może jedynie, że:

- każdy szczęściarz epistemiczny jest normalny, a nawet wie, że jest normalny;
- każdy szczęściarz epistemiczny jest regularny i o tym także wie;
- wreszcie, każdy szczęściarz epistemiczny jest przekonany o tym, że jest szczęściarzem epistemicznym; a zatem to jego przekonanie jest trafne i, w konsekwencji, każdy szczęściarz epistemiczny wie, że jest szczęściarzem epistemicznym.

Poziomy samoświadomości

Można rozważać pięć typów myślaków, o wstępujących poziomach samoświadomości:

- Typ 1: prostaczek logiczny.
- Typ 1*: prostaczek logiczny, który, jeśli uwierzył w $p \rightarrow q$, to uwierzy, że jeśli uwierzył w p, to uwierzy w q.
- Typ 2: prostaczek logiczny, który wierzy we wszystkie zdania postaci $(\boldsymbol{B}p \wedge \boldsymbol{B}(p \rightarrow q)) \rightarrow \boldsymbol{B}q$.
- Typ 3: myślak typu 2, który, jeśli wierzy w p, to wierzy w $\boldsymbol{B}p$.
- Typ 4: szczęściarz epistemiczny, tj. normalny i regularny prostaczek logiczny, który wierzy we wszystkie zdania postaci $\boldsymbol{B}p \rightarrow \boldsymbol{B}\boldsymbol{B}p$, czyli wierzy, że jest normalny.

Uwaga. Terminy: prostaczek logiczny oraz szczęściarz epistemiczny nie występują w Forever Undecided; wprowadzamy je na użytek tej prezentacji.

Poziomy samoświadomości

Z podanych definicji wynika, że:

- Każdy prostaczek logiczny jest myślakiem typu 1^*.
- Każdy myślak typu 1^* jest regularnym prostaczkiem logicznym (i *vice versa*).
- Każdy myślak typu 2 wie, że jest typu 1^*.
- Myślaki typu 3 to dokładnie normalne myślaki typu 2.
- Dla $1 \leqslant n < 4$: każdy myślak typu n jest też myślakiem typu $n + 1$.
- $1 < n \leqslant 4$: każdy myślak typu n wierzy, że jest myślakiem typu $n - 1$.

Uwaga. Ponieważ każdy szczęściarz epistemiczny wie, że jest szczęściarzem epistemicznym, więc stanowi on zwieńczenie hierarchii samoświadomych myślaków. Inaczej mówiąc, gdybyśmy chcieli zdefiniować myślaka typu 5 jako takiego, który jest typu 4 i wierzy, iż jest typu 4, to otrzymalibyśmy jedynie myślaka typu 4.

Zapraszam na szczyt

Możemy już rozpocząć wyprawę na kilka Szczytów Metalogiki.

II Twierdzenie Gödla

Za chwilę dowiesz się czegoś naprawdę frapującego o swoim systemie przekonań. Udowodnimy mianowicie:

Twierdzenie 1.
Przypuśćmy, że normalny prostaczek logiczny S wierzy w zdanie postaci $p \equiv \neg \boldsymbol{B} p$. Wtedy:

- (a) Jeśli S kiedykolwiek uwierzy w p, to stanie się sprzeczny.
- (b) Jeśli S jest szczęściarzem epistemicznym, to wie, iż jeśli kiedykolwiek uwierzy w p, to stanie się sprzeczny — tj. uwierzy w $\boldsymbol{B} p \rightarrow \boldsymbol{B} \perp$.
- (c) Jeśli S jest szczęściarzem epistemicznym i wierzy, że nie może być sprzeczny, to stanie się sprzeczny.

II Twierdzenie Gödla

Dowód Twierdzenia 1.

(a) Przypuśćmy, że S wierzy w p.

Będąc normalnym, uwierzy w $\boldsymbol{B}p$.

Nadto, ponieważ wierzy w p oraz wierzy w $p \equiv \neg \boldsymbol{B}p$,
więc musi uwierzyć w $\neg \boldsymbol{B}p$
(bo jest prostaczkiem logicznym).

A więc uwierzy jednocześnie w $\boldsymbol{B}p$ oraz w $\neg \boldsymbol{B}p$,
a stąd stanie się sprzeczny.

II Twierdzenie Gödla

(b) Przypuśćmy, że S jest szczęściarzem epistemicznym. Ponieważ jest wtedy prostaczkiem logicznym i wierzy w $p \equiv \neg \boldsymbol{B} p$, więc musi także wierzyć w $p \rightarrow \neg \boldsymbol{B} p$.
Nadto, S jest regularny, a stąd uwierzy w $\boldsymbol{B} p \rightarrow \boldsymbol{B} \neg \boldsymbol{B} p$. Wierzy też w $\boldsymbol{B} p \rightarrow \boldsymbol{B} \boldsymbol{B} p$ (ponieważ wie, że jest normalny).
Zatem S uwierzy w $\boldsymbol{B} p \rightarrow (\boldsymbol{B} \boldsymbol{B} p \wedge \boldsymbol{B} \neg \boldsymbol{B} p)$, które jest logiczną konsekwencją ostatnich dwóch zdań.
Wierzy również w $(\boldsymbol{B} \boldsymbol{B} p \wedge \boldsymbol{B} \neg \boldsymbol{B} p) \rightarrow \boldsymbol{B} \perp$ (na mocy $(*)$, ponieważ dla dowolnego zdania X, S wierzy w $(\boldsymbol{B} X \wedge \boldsymbol{B} \neg X) \rightarrow \boldsymbol{B} \perp$, a więc wierzy w jego szczególny przypadek, gdzie X jest zdaniem $\boldsymbol{B} p$).
Gdy S już uwierzy jednocześnie w $\boldsymbol{B} p \rightarrow (\boldsymbol{B} \boldsymbol{B} p \wedge \boldsymbol{B} \neg \boldsymbol{B} p)$ oraz w $(\boldsymbol{B} \boldsymbol{B} p \wedge \boldsymbol{B} \neg \boldsymbol{B} p) \rightarrow \boldsymbol{B} \perp$, będzie musiał uwierzyć w $\boldsymbol{B} p \rightarrow \boldsymbol{B} \perp$ (ponieważ jest prostaczkiem logicznym).

II Twierdzenie Gödla

(c) Ponieważ S wierzy w $\boldsymbol{B}p \rightarrow \boldsymbol{B}\perp$ (jak właśnie udowodniliśmy), więc wierzy także w $\neg\boldsymbol{B}\perp \rightarrow \neg\boldsymbol{B}p$.

Załóżmy teraz, że S wierzy w $\neg\boldsymbol{B}\perp$
(wierzy, że nie może być sprzeczny).

Ponieważ wierzy też w $\neg\boldsymbol{B}\perp \rightarrow \neg\boldsymbol{B}p$ (jak właśnie widzieliśmy), więc uwierzy w $\neg\boldsymbol{B}p$.

A ponieważ wierzy również w $p \equiv \neg\boldsymbol{B}p$,
więc uwierzy w p, a stąd stanie się sprzeczny, na mocy (a).

Co właściwie udowodniliśmy?

Wolisz być Prostaczkiem Logicznym czy Szczęściarzem Epistemicznym?

II Twierdzenie Gödla

Udowodniliśmy przed chwilą nie byle co, bo modalną (epistemiczną) wersję **II Twierdzenia Gödla** (o niedowodliwości niesprzeczności arytmetyki w samej arytmetyce).

Oczywiście był to dowód w postaci wielce uproszczonej — precyzyjny dowód wymagałby, powiedzmy, jednosemestralnego wykładu wstępnego.

Korzystaliśmy z rozdziału 12 tłumaczenia książki Raymonda Smullyana *Forever Undecided*.

Poddajemy ocenie audytorium, czy ten sposób popularyzacji wiedzy (meta)logicznej można uznać za dydaktycznie przydatny.

Przykład teologiczny

Przykład.
Przypuśćmy, że jesteś studentką teologii i że Twój Ulubiony Profesor teologii mówi do Ciebie:
Bóg istnieje wtedy i tylko wtedy, gdy nigdy nie uwierzysz, że Bóg istnieje.

Jeśli wierzysz profesorowi, to wierzysz w zdanie $g \equiv \neg \boldsymbol{B} g$, gdzie g jest zdaniem stwierdzającym, że Bóg istnieje.
Wtedy, zgodnie z Twierdzeniem 1, nie możesz wierzyć w swoją własną niesprzeczność bez popadnięcia w sprzeczność.

Oczywiście, możesz wierzyć we własną niesprzeczność, bez popadnięcia przy tym w sprzeczność — wystarczy, że przestaniesz ufać Twojemu Ulubionemu Profesorowi.
Coś za coś.

Modalna interpretacja dowodliwości

Przy modalnej interpretacji dowodliwości nie mamy jednak takiej możliwości ucieczki, jak w powyższym przykładzie.

Wiadomo, że formuła *god*, stwierdzająca swoją własną niedowodliwość w PA, jest prawdziwa, lecz dowodu w PA nie posiada.

Można pokazać, że twierdzeniem stosownego systemu modalnego (w którym reprezentujemy dowodliwość w PA) jest:

$$god \equiv \neg \boldsymbol{B} god.$$

Navigare necesse est

Schodzimy ze szczytu Gödla. Przed nami pasmo Gór Löba.

Wishful thinking

Pokażemy teraz, co wystarcza, aby każda z obecnych tu Uroczych Pań została — powiedzmy — **Miss World 2010**. Będzie to przykład samospełniającego się przekonania.

Martin Hugo Löb

Samospełniające się przekonania

Przypuśćmy, że:

- jesteś szczęściarą epistemiczną;
- osoby, które rozważamy albo zawsze mówią fałsz, albo zawsze mówią prawdę (i Ty wiesz, że tak jest);
- wierzysz swojemu chłopakowi, który prawdziwie (!) mówi: (†) *Jeśli uwierzysz, że zostaniesz Miss World 2010, to zostaniesz Miss World 2010.*
- wierzysz też mnie (JP), który mówi: (‡) *Jeśli wierzysz, że ja zawsze mówię prawdę, to zostaniesz Miss World 2010.*

Twierdzenie 2.
Przy powyższych założeniach *zostaniesz* **Miss World 2010**. Cieszysz się?

Samospełniające się przekonania

Dla skrótu, przyjmijmy oznaczenia:

- k zastępuje zdanie stwierdzające, iż ja (JP) zawsze mówię prawdę;
- α zastępuje zdanie stwierdzające, że zostaniesz Miss World 2010.

Dowód składa się z dwóch części.
1. W pierwszej pokazujemy, że nasze założenia implikują $\boldsymbol{B}\alpha$. Jest to dowód założeniowy, dostępny dla każdej szczęściary epistemicznej.
Mamy udowodnić formułę:

$$(\bigstar) \quad ((\boldsymbol{B}\alpha \rightarrow \alpha) \wedge (k \equiv (\boldsymbol{B}k \rightarrow \alpha))) \rightarrow \boldsymbol{B}\alpha.$$

Uwaga. Zdanie k stwierdza, iż JP zawsze mówi prawdę; a więc prawdą jest, że JP wypowiada (‡) dokładnie wtedy, gdy prawdziwe jest $k \equiv (\ddagger)$, czyli dokładnie wtedy, gdy prawdziwe jest $k \equiv (\boldsymbol{B}k \rightarrow \alpha)$.

1.	$(\boldsymbol{B}\alpha \rightarrow \alpha) \wedge (k \equiv (\boldsymbol{B}k \rightarrow \alpha))$	założenie
2.	$\boldsymbol{B}\alpha \rightarrow \alpha$	OK: 1
3.	$k \equiv (\boldsymbol{B}k \rightarrow \alpha)$	OK: 1
4.	$k \rightarrow (\boldsymbol{B}k \rightarrow \alpha)$	OR: 3
5.	$(\boldsymbol{B}k \rightarrow \alpha) \rightarrow k$	OR: 3
6.1.	k	założenie dodatkowe
6.2.	$\boldsymbol{B}k \rightarrow \alpha$	MP: 4, 6.1.
6.3.	$\boldsymbol{B}k$	6.1. i warunek **(3)**
6.4.	α	MP: 6.2., 6.3.
7.	$k \rightarrow \alpha$	6.1.→6.4.
8.	$\boldsymbol{B}(k \rightarrow \alpha)$	7 i warunek **(3)**
9.	$\boldsymbol{B}k \rightarrow \boldsymbol{B}\alpha$	8 i warunki **(1a)** i **(2)**
10.	$\boldsymbol{B}k \rightarrow \alpha$	2, 9 i warunki **(1b)**, **(1a)** (prawo sylog. hipotet.)
11.	k	MP: 5, 10
12.	$\boldsymbol{B}k$	11 i warunek **(3)**
13.	α	MP: 10, 12
14.	$\boldsymbol{B}\alpha$	13 i warunek **(3)**.

Samospełniające się przekonania

2. Ponieważ proroctwo (†) Twojego chłopaka (tj. zdanie $\boldsymbol{B}\alpha \rightarrow \alpha$) jest z założenia prawdziwe, a powyższy dowód formuły (★) pokazuje, iż nasze założenia implikują $\boldsymbol{B}\alpha$, więc na mocy reguły odrywania otrzymujemy α, czyli tezę.

Zostaniesz Miss World 2010!!!
Cieszysz się???

Uwaga. Powyższy dowód był przykładem dowodu wprost. Aby pokazać, że zostaniesz Miss World 2010 nie musieliśmy odwoływać się do absurdu.
Cieszysz się?

Wędrujemy dalej?

Jeśli mamy: czas, siły oraz ochotę, to możemy wrócić w Góry Gödlowskie.

I Twierdzenie Gödla

Myślak jest nazywany **stabilnym**, jeśli dla każdego zdania p, jeśli wierzy on w $\boldsymbol{B}p$, to wierzy też w p.

Powiemy, że system przekonań myślaka jest **niezupełny**, jeśli istnieje co najmniej jedno zdanie p takie, że myślak nigdy nie uwierzy w p ani też nigdy nie uwierzy w $\neg p$ (pozostanie na zawsze niezdecydowany, czy p jest prawdziwe, czy fałszywe).
Systemy przekonań, które nie są niezupełne, nazywamy zupełnymi. Osoby, które władają takimi systemami przekonań, są dość uciążliwe w kontaktach społecznych — każda taka osoba jest Besserwisserem, kimś kto na każdy pogląd ma wyrobione zdanie, pozbawiony jest wątpliwości.

Gdy zajmujemy się systemami twierdzeń raczej niż zespołami przekonań, to systemami typu 1 nazwiemy te, które spełniają warunki 1a i 1b podane wyżej.

I Twierdzenie Gödla

Normalny prostaczek logiczny przybywa na Wyspę Rycerzy i Łotrów i wierzy w reguły wyspy. (To, czy reguły wyspy rzeczywiście obowiązują, czy nie, jest bez znaczenia.)
Spotyka tubylca, który mówi:
„Nigdy nie uwierzysz, że jestem rycerzem."
Udowodnimy, że zachodzi wtedy:

Twierdzenie 3.
Jeśli myślak jest jednocześnie niesprzeczny i stabilny, to jego system przekonań jest niezupełny. Dokładniej mówiąc, znajdziemy zdanie p takie, że zachodzą następujące dwa warunki:

- (a) Jeśli myślak jest niesprzeczny, to nigdy nie uwierzy w p.
- (b) Jeśli myślak jest jednocześnie niesprzeczny i stabilny, to nigdy nie uwierzy w $\neg p$.

I Twierdzenie Gödla

Zdanie p o które chodzi jest po prostu zdaniem k — zdaniem stwierdzającym, że tubylec jest rycerzem.

Tubylec wygłosił $\neg \boldsymbol{B}k$, a więc myślak uwierzy w $k \equiv \neg \boldsymbol{B}k$.

(a) Przypuśćmy, że myślak wierzy w k. Wtedy, będąc normalnym, uwierzy w $\boldsymbol{B}k$. Uwierzy też w $\neg \boldsymbol{B}k$ (ponieważ wierzy w k oraz wierzy w $k \equiv \neg \boldsymbol{B}k$ i jest prostaczkiem logicznym), a stąd stanie się sprzeczny. Zatem, jeśli jest niesprzeczny, to nigdy nie uwierzy w k.

(b) Przypuśćmy, że myślak jest prostaczkiem logicznym i wierzy w $k \equiv \neg \boldsymbol{B}k$, wtedy wierzy też w $\neg k \equiv \boldsymbol{B}k$. Przypuśćmy teraz, że kiedykolwiek uwierzy on w $\neg k$. Wtedy uwierzy w $\boldsymbol{B}k$. Jeśli jest stabilny, to uwierzy w k i stąd stanie się sprzeczny (ponieważ wierzy w $\neg k$). Zatem, jeśli jest jednocześnie stabilny i niesprzeczny, to nigdy nie uwierzy w $\neg k$.

Podsumowując, jeśli jest on jednocześnie stabilny i niesprzeczny, to nigdy nie uwierzy że tubylec jest rycerzem i nigdy nie uwierzy, że tubylec jest łotrem.

I Twierdzenie Gödla

To samo rozumowanie, którego użyto w rozwiązaniu powyższego problemu, gdy zastosować je do systemów matematycznych raczej niż do myślaków, ustanawia następującą postać Pierwszego Twierdzenia Gödla o Niezupełności:

Twierdzenie 4. Dowolny niesprzeczny, normalny, stabilny system Gödlowski musi być niezupełny. Dokładniej, jeśli S jest normalnym systemem typu 1, a p jest zdaniem takim, że $p \equiv \neg \boldsymbol{B} p$ jest dowodliwe w S, to jeśli S jest niesprzeczny, to p nie jest dowodliwe w S, a jeśli S jest dodatkowo stabilny, to $\neg p$ również nie jest dowodliwe w S.

Zdanie p nazywamy **nierozstrzygalnym** w systemie S, jeśli ani p ani jego negacja $\neg p$ nie jest dowodliwe w S. Zatem Pierwsze Twierdzenie Gödla o Niezupełności mówi nam, że dla dowolnego niesprzecznego, normalnego, stabilnego systemu Gödlowskiego S, musi zawsze istnieć co najmniej jedno zdanie p, które, choć *wyrażalne* w języku S, nie jest *rozstrzygalne* w S — nie można w S udowodnić ani tego zdania, ani jego negacji.

I Twierdzenie Gödla

Dla dowolnej własności P liczb, zdanie stwierdzające, że istnieje co najmniej jedna liczba n mająca własność P zapisujemy: $\exists nP(n)$.

Przypuśćmy, że mamy system matematyczny i własność P taką, że zdanie $\exists nP(n)$ jest dowodliwe w systemie, a jednak dla każdego poszczególnego n zdanie $\neg P(n)$ jest dowodliwe — to jest, wszystkie z nieskończenie wielu zdań $\neg P(0)$, $\neg P(1)$, $\neg P(2), \ldots, \neg P(n), \ldots$ są dowodliwe.

Oznacza to, z jednej strony, że w systemie można udowodnić zdanie stwierdzające, że jakaś liczba ma własność P, a jednak o każdej *poszczególnej* liczbie n można udowodnić, że liczba ta owej własności nie posiada!

Systemy takie nazywane są *ω-sprzecznymi*.

I Twierdzenie Gödla

Pojęcie ω-sprzeczności zostało kiedyś zabawnie scharakteryzowane przez matematyka Paula Halmosa, który zdefiniował ω-sprzeczną matkę jako taką, która mówi swojemu dziecku: „Jest coś, co możesz robić, ale nie możesz robić tego, nie możesz robić tamtego, nie możesz robić owego, ..." Dziecko pyta: „Ale, mamusiu, czy jest *cokolwiek* co mógłbym robić?" Matka odpowiada: „O tak, ale nie jest to to, ani tamto, ani owo, ..."

System jest nazywany ω-niesprzecznym, jeśli nie jest on ω-sprzeczny. Tak więc dla systemu ω-niesprzecznego, jeśli $\exists n P(n)$ jest dowodliwe, to istnieje co najmniej jedna liczba n taka, że zdanie $\neg P(n)$ *nie jest* dowodliwe.

Sprzeczny system typu 1 jest również ω-sprzeczny, ponieważ w sprzecznym systemie typu 1 wszystkie zdania są dowodliwe.

I Twierdzenie Gödla

We wszystkich dotąd rozważanych problemach, *kolejność* w której myślak wierzył w różnorakie zdania nie odgrywała roli. W pozostałych problemach w tej części, kolejność ta odgrywa rolę pierwszorzędną.

Myślak przybywa na Wyspę Rycerzy i Łotrów pewnego dnia, który nazwiemy dniem numer 0. Następny dzień jest dniem numer 1, następny dniem numer 2, i tak dalej.

Dla każdej liczby naturalnej n mamy więc dzień numer n (n-ty dzień) i zakładamy, że myślak jest nieśmiertelny i ma przed sobą nieskończenie wiele dni.

I Twierdzenie Gödla

Dla każdej liczby naturalnej n i dowolnego zdania p niech $\boldsymbol{B}_n p$ będzie zdaniem stwierdzającym, że myślak uwierzył w p w jakimś momencie n-tego dnia.

Zdanie $\boldsymbol{B}p$ jest, jak zwykle, zdaniem stwierdzającym, że myślak uwierzy w p tego lub innego dnia, lub, co na jedno wychodzi, zdaniem $\exists n \boldsymbol{B}_n p$ (istnieje n takie, że myślak uwierzy w p n-tego dnia).

Nazwiemy myślaka *ω-sprzecznym*, jeśli istnieje co najmniej jedno zdanie p takie, że myślak (kiedyś) wierzy w $\boldsymbol{B}p$, a jednak dla każdego n wierzy on (kiedyś) w $\neg \boldsymbol{B}_n p$.

Myślaka nazywamy *ω-niesprzecznym*, jeśli nie jest on ω-sprzeczny.

I Twierdzenie Gödla

Rozważmy teraz myślaka, który spełnia następujące trzy warunki.

- **Warunek C_1.** Jest on typu prostaczkiem logicznym.
- **Warunek C_2.** Dla dowolnej liczby naturalnej n i dowolnego zdania p: (a) jeśli myślak wierzy w p n-tego dnia, to (prędzej czy później) uwierzy w $\boldsymbol{B}_n p$; (b) jeśli nie wierzy on w p n-tego dnia, to (prędzej czy później) uwierzy w $\neg \boldsymbol{B}_n p$. (Oddajemy w ten sposób, że myślak śledzi to, w jakie zdania wierzył, a w jakie nie wierzył we wszystkich dniach poprzednich.)
- **Warunek C_3.** Dla dowolnych n oraz p myślak wierzy w zdanie $\boldsymbol{B}_n p \rightarrow \boldsymbol{B} p$ (które, oczywiście, jest zdaniem prawdziwym).

Następujący problem jest bardzo zbliżony do oryginalnego sformułowania Gödla jego Pierwszego Twierdzenia o Niezupełności.

I Twierdzenie Gödla

Myślak spełniający powyższe trzy warunki przybywa na Wyspę Rycerzy i Łotrów i wierzy w reguły wyspy. Spotyka tubylca, który mówi mu:

„Nigdy nie uwierzysz, że jestem rycerzem."
Udowodnimy, że zachodzi wtedy:

Twierdzenie 5.

- (a) Jeśli myślak jest (prosto) niesprzeczny, to nigdy nie uwierzy, że tubylec jest rycerzem.
- (b) Jeśli myślak jest ω-niesprzeczny, to nigdy nie uwierzy, że tubylec jest łotrem.

Zatem jeśli myślak jest ω-niesprzeczny (a stąd także prosto niesprzeczny), to pozostanie na zawsze niezdecydowany co do tego, czy tubylec jest rycerzem, czy też łotrem.

I Twierdzenie Gödla

Najłatwiejszym sposobem rozwiązania obecnego problemu będzie pokazanie, że dowolny myślak spełniający warunki C_1, C_2 oraz C_3 musi być normalny, a jeśli jest ω-niesprzeczny, to musi być też stabilny.

(a) Pokazujemy, że jest on normalny.
Przypuśćmy, że wierzy on w p.
Wtedy dla pewnego n, wierzy on n-tego dnia w p.
Wtedy, na mocy punktu (a) z warunku 2, uwierzy w $\boldsymbol{B}_n p$.
Wierzy także w $\boldsymbol{B}_n p \rightarrow \boldsymbol{B} p$ (na mocy warunku 3), a więc będąc typu 1 (warunek 1) uwierzy w $\boldsymbol{B} p$.
Zatem jest normalny.

I Twierdzenie Gödla

(b) Przypuśćmy teraz, że jest on ω-niesprzeczny.
Pokażemy, że jest stabilny.
Przypuśćmy, że wierzy on w $\boldsymbol{B}p$.
Jeśli nigdy nie uwierzy w p, to dla każdej liczby n, nie wierzy on w p n-tego dnia, a stąd na mocy punktu (b) z warunku 2, dla każdego n wierzy on w $\neg\boldsymbol{B}_n p$.
Ale ponieważ wierzy on w $\boldsymbol{B}p$, więc stanie się wtedy ω-sprzeczny.
Zatem, jeśli jest on ω-niesprzeczny i wierzy w $\boldsymbol{B}p$, to musi wierzyć w p tego lub innego dnia.
Dowodzi to, że jeśli jest on ω-niesprzeczny, to musi być stabilny (zakładając, że spełnia on warunki C_1, C_2, C_3 — lub nawet tylko (b) z warunku C_2).
Zatem, na mocy Twierdzenia 4, pozostanie on na zawsze niezdecydowany.

Chcemy iść dalej?

Stąd już niedaleko na Połoniny Rosserowskie.

Twierdzenie Rossera

Dla dowolnych zdań p oraz q, powiemy, że myślak uwierzył w p *wcześniej niż* (zanim) uwierzył w q, jeśli jest taki dzień, w którym wierzy on w p, a jeszcze nie uwierzył w q. Jeśli myślak *nigdy* nie uwierzy w q, ale uwierzył w p (tego lub innego dnia), to uznajemy, iż *prawdziwe* jest, że uwierzył w p wcześniej, niż uwierzył w q. (Innymi słowy, nie musi on wcale kiedykolwiek uwierzyć w q, aby uwierzyć w p wcześniej niż uwierzyć w q.) Niech $\boldsymbol{B}p < \boldsymbol{B}q$ będzie zdaniem stwierdzającym, że myślak uwierzył w p wcześniej niż uwierzył w q. Jeśli $\boldsymbol{B}p < \boldsymbol{B}q$ jest prawdziwe, to oczywiście $\boldsymbol{B}q < \boldsymbol{B}p$ jest fałszywe.

Zdefiniujemy myślaka *Rosserowskiego* jako prostaczka logicznego, dla którego zachodzi następujący warunek:

Warunek R. Dla dowolnych zdań p oraz q, jeśli myślak uwierzył w p pewnego dnia, w którym jeszcze nie uwierzył w q, to (wcześniej czy później) uwierzy on w $\boldsymbol{B}p < \boldsymbol{B}q$ oraz w $\neg(\boldsymbol{B}q < \boldsymbol{B}p)$.

Twierdzenie Rossera

Myślak Rosserowski przybywa na Wyspę Rycerzy i Łotrów i wierzy w reguły wyspy. Spotyka tubylca, który mówi mu:

„Nigdy nie uwierzysz wcześniej, że jestem rycerzem, niż uwierzysz, że jestem łotrem.''

(Oddając to symbolicznie, tubylec wygłasza zdanie $\neg(\boldsymbol{B}k < \boldsymbol{B}\neg k)$.) Udowodnimy:

Twierdzenie 6.
Jeśli myślak jest *po prostu* niesprzeczny, to musi na zawsze pozostać niezdecydowany, czy tubylec jest rycerzem, czy łotrem.

Twierdzenie Rossera

Ponieważ tubylec stwierdził $\neg(\boldsymbol{B}k < \boldsymbol{B}\neg k)$, więc myślak uwierzy w $k \equiv \neg(\boldsymbol{B}k < \boldsymbol{B}\neg k)$. Przypuśćmy, że myślak jest (prosto) niesprzeczny. Mamy pokazać, że nigdy nie uwierzy w k i nigdy nie uwierzy w $\neg k$.

(a) Przypuśćmy, że kiedyś uwierzył w k. Ponieważ jest niesprzeczny, więc nigdy nie uwierzy w $\neg k$, a stąd uwierzy w k wcześniej niż uwierzy w $\neg k$. Stąd, uwierzy w $\boldsymbol{B}k < \boldsymbol{B}\neg k$ (na mocy warunku R). Ale wierzy też w $k \equiv \neg(\boldsymbol{B}k < \boldsymbol{B}\neg k)$, a więc uwierzy w $\neg k$, a wierząc już w k stanie się sprzeczny! Tak więc, jeśli jest niesprzeczny, to nigdy nie uwierzy w k.

(b) Przypuśćmy, że kiedyś uwierzył w $\neg k$. Będąc niesprzecznym, nigdy nie uwierzy w k, a stąd uwierzy w $\neg k$ wcześniej niż uwierzy w k, a stąd na mocy warunku R uwierzy w $\neg(\boldsymbol{B}k < \boldsymbol{B}\neg k)$. Ale wierzy on w $k \equiv \neg(\boldsymbol{B}k < \boldsymbol{B}\neg k)$, a więc uwierzy wtedy w k i stanie się sprzeczny. A zatem, jeśli jest niesprzeczny, to nie może także uwierzyć w $\neg k$.

We mgle Rosserowskich Połonin

Jak możemy interpretować Twierdzenie Rossera?

Twierdzenie Rossera

Dowodliwe zdania systemów matematycznych są dowodliwe na różnych etapach.

Moglibyśmy myśleć o systemie matematycznym jako o komputerze zaprogramowanym tak, aby dowodzić różnorakich zdań *kolejno*.

Powiemy, że p jest dowodliwe *wcześniej* (zanim) niż q (w danym systemie matematycznym), jeśli p zostało udowodnione na pewnym etapie, na którym q jeszcze nie zostało udowodnione (q może być lub też nie być udowodnione na jakimś późniejszym etapie).

Twierdzenie Rossera

Dla dowolnych zdań p oraz q wyrażalnych w systemie, zdanie $\boldsymbol{B}p < \boldsymbol{B}q$ (p jest dowodliwe wcześniej niż q) również jest wyrażalne w systemach typu tych rozpatrywanych przez Gödla, a Rosser pokazał, że jeśli p *jest* dowodliwe wcześniej niż q, to zdania $\boldsymbol{B}p < \boldsymbol{B}q$ oraz $\neg(\boldsymbol{B}q < \boldsymbol{B}p)$ są oba dowodliwe w systemie.

Rosser znalazł także zdanie p takie, że $p \equiv \neg(\boldsymbol{B}p < \boldsymbol{B}\neg p)$ jest dowodliwe w systemie. (Takie zdanie p odpowiada tubylcowi z pierwszego rozważanego w tej części problemu, który mówi: „Nigdy nie uwierzysz wcześniej, że jestem rycerzem, niż uwierzysz, że jestem łotrem.”)
Wtedy, na mocy rozumowania z rozwiązania wspomnianego problemu, jeśli p jest dowodliwe, to system jest sprzeczny, a jeśli $\neg p$ jest dowodliwe, to system także jest sprzeczny.
A zatem, jeśli system jest niesprzeczny, to zdanie p jest nierozstrzygalne w systemie.

Twierdzenie Rossera

Zdanie Gödlowskie może zostać sparafrazowane jako:

„Nie jestem dowodliwe na żadnym etapie.”

Bardziej wyszukane zdanie Rossera może zostać sparafrazowane jako:

„Nie mogę być dowiedzione na żadnym etapie, chyba że moja negacja została już wcześniej udowodniona.”

Zdanie Gödla, chociaż prostsze, wymaga założenia ω-niesprzeczności dla przeprowadzenia rozumowania. Zdanie Rossera, chociaż bardziej skomplikowane, dostarcza szukanego rezultatu przy słabszym założeniu prostej niesprzeczności.

Wędrujemy, dopóki czynny jest horyzont

A na horyzoncie Masyw Tarskiego.

Twierdzenie Tarskiego

Przypuśćmy, że mamy myślaka — nazwijmy go Paul — który jest zawsze *ścisły* w swoich przekonaniach (nigdy nie wierzy w zdania fałszywe). Nie musi on być prostaczkiem logicznym, ani normalnym, nie jest też konieczne, aby rzeczywiście odwiedzał Wyspę Rycerzy i Łotrów. Wszystko, co musimy o nim wiedzieć to to, że jest ścisły.

Pewnego dnia tubylec mówi o nim:

„Paul nigdy nie uwierzy, że jestem rycerzem."
Wtedy logicznie wynika stąd:

Twierdzenie 7.
System przekonań Paula jest niezupełny.

Twierdzenie Tarskiego

Jeśli Paul kiedykolwiek uwierzy, że tubylec jest rycerzem, to sfalsyfikuje to tym samym to, co powiedział tubylec, czyniąc tubylca łotrem, a tym samym czyniąc Paula nieścisłym z powodu jego wiary, że tubylec jest rycerzem.

Ale powiedziano nam, że Paul jest ścisły, a więc nigdy nie uwierzy on, że tubylec jest rycerzem.

Stąd, to co powiedział tubylec jest prawdziwe, a więc tubylec rzeczywiście jest rycerzem.

Wtedy, ponieważ Paul jest ścisły, nigdy nie będzie żywił fałszywego przekonania, że tubylec jest łotrem.

A zatem Paul nigdy nie dowie się, czy tubylec jest rycerzem, czy łotrem.

Twierdzenie Tarskiego

Komentarz. Treść matematyczna powyższej zagadki jest następująca. W systemach rozważanych przez Gödla mamy nie tylko pewne zdania nazywane zdaniami *dowodliwymi*, lecz również obszerniejszą klasę zdań nazywanych zdaniami *prawdziwymi* systemu.

W klasie zdań prawdziwych systemu obowiązują reguły tabliczek prawdziwościowych dla spójników logicznych.
Nadto, dla każdego zdania p systemu, zdanie $\boldsymbol{B}p$ jest *prawdziwym* zdaniem systemu wtedy i tylko wtedy, gdy p jest zdaniem *dowodliwym* systemu.

Gödel znalazł godne uwagi zdanie g takie, że zdanie $g \equiv \neg \boldsymbol{B}g$ było zdaniem *prawdziwym* systemu (było ono nawet faktycznie dowodliwe w systemie, ale ten mocniejszy fakt nie jest potrzebny dla obecnego rozumowania).

Twierdzenie Tarskiego

Gdyby g było fałszywe, to $\boldsymbol{B}g$ byłoby prawdziwe, a stąd g byłoby dowodliwe, a stąd prawdziwe, i mielibyśmy sprzeczność.

Zatem g jest prawdziwe, a stąd $\neg\boldsymbol{B}g$ jest prawdziwe, czyli g nie jest dowodliwe w systemie.

Tak więc, g jest prawdziwe, ale niedowodliwe w systemie.

Ponieważ g jest prawdziwe, więc $\neg g$ jest fałszywe, a stąd także niedowodliwe w systemie (ponieważ wszystkie dowodliwe zdania są prawdziwe).

A zatem g jest nierozstrzygalne w systemie.

Czas pożegnać się ze Szczytami Metalogiki

Byliśmy tylko na kilku. A jest ich nieskończenie wiele.

Dawniejsza opozycja filozoficzna wobec logiki modalnej była osadzona w przybliżeniu w trzech różnych (i nieporównywalnych) przekonaniach. Po pierwsze, są tacy, którzy są przekonani, że wszystko, co jest prawdziwe jest koniecznie prawdziwe, a stąd nie ma żadnej różnicy między prawdą a prawdą konieczną. Po drugie, są tacy, którzy wierzą, że nic nie jest koniecznie prawdziwe, a stąd dla dowolnego zdania p, zdanie $\boldsymbol{N}p$ (p jest koniecznie prawdziwe) jest po prostu fałszywe! A po trzecie, są i tacy, którzy twierdzą, że słowa „koniecznie prawdziwe" nie niosą jakiegokolwiek sensu. Tak więc, każde z tych nastawień filozoficznych odrzuca logikę modalną ze swoich własnych powodów. W istocie, pewien bardzo znany filozof wsławił się sugestią, że nowoczesna logika modalna została poczęta w grzechu. Na co Boolos bardzo stosownie odpowiedział: „Jeśli nowoczesna logika modalna została poczęta w grzechu, to została wybawiona przez Gödlowskość". [W oryginale: *If modern modal logic was conceived in sin, then it has been redeemed through Gödliness.*]

Trzeba już schodzić. . .

Góry i Matematyka uczą pokory [Kazimierz Głazek].

Koniec Dodatku A

Dodatek nie rości sobie pretensji do kompletności:

- ani jako przedstawienie wszystkich treści *Forever Undecided*,
- ani jako wprowadzenie do logiki dowodliwości.

Staraliśmy się jedynie pokazać próbkę możliwości popularyzacji wiedzy o logice modalnej i jej zastosowaniach.

Zachęcamy do lektury książki!

Czy wiesz, jak wysoko byłeś?

Maszyny logiczne Smullyana

Smullyan skonstruował cały szereg maszyn logicznych, które drukują zdania „mówiące" coś o nich samych.
Maszyny: Craiga, Fergussona i McCullocha, przedstawione w *Jaki jest tytuł tej książki?* oraz *Dama czy tygrys?* są już znane polskiemu czytelnikowi.
Tu przedstawimy pewną maszynę Smullyana, opisaną w *Forever Undecided.*

Dla pełnego zrozumienia jej działania potrzebna jest znajomość logiki dowodliwości (logiki Gödla-Löba).
Zakładamy u audytorium znajomość tego materiału (patrz Dodatek A).

Maszyny logiczne Smullyana

Malcolm Fergusson, gdy usłyszał o twierdzeniach Gödla i Löba, z miejsca zabrał się za konstrukcję maszyny, którą z zachwytem pokazał swoim przyjaciołom.
Ku ich zadowoleniu udowodnił, że maszyna jest niesprzeczną i stabilną maszyną typu G, a szczególne upodobanie znalazł w demonstracji, że maszyna, chociaż niesprzeczna, nigdy nie może dowieść własnej niesprzeczności!
Maszyna ilustruje w niezwykle prosty i pouczający sposób podstawowe idee zawarte w Pierwszym oraz Drugim Twierdzeniu Gödla jak również w Twierdzeniu Löba.

Niżej podajemy opis działania maszyny Fergussona oraz pewne ważne fakty jej dotyczące.
Opis pochodzi z rozdziału 26 *Forever Undecided*. W rozdziale tym znajdujemy też opis dwóch innych maszyn, który tu pominiemy.

Maszyna drukuje różnorakie zdania zbudowane z siedemnastu symboli. Pierwsze siedem z tych symboli to następujące:

$$\begin{array}{ccccccc} P & \perp & \rightarrow & (&) & d & , \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array}$$

Pod każdym z tych symboli podpisano jego *numer Gödlowski*.

Pozostałe dziesięć symboli to znane cyfry 1, 2, 3, 4, 5, 6, 7, 8, 9, 0. Tym cyfrom przyporządkowujemy numery Gödlowskie w następujący sposób. Numerem Gödlowskim cyfry 1 jest 89 (8 po której następuje jedna 9); numerem Gödlowskim cyfry 2 jest 899 (8 po której następują dwie 9); i tak dalej, aż do cyfry 0, której numerem Gödlowskim jest 89999999999 (8 po której następuje dziesięć 9).

Tak więc, każdy z siedemnastu symboli uzyskuje numer Gödlowski.

Dla danego wyrażenia złożonego, odnajdujemy jego numer Gödlowski przez zastąpienie każdego symbolu jego numerem Gödlowskim — dla przykładu, numerem Gödlowskim wyrażenia $(P \perp \rightarrow \perp)$ jest 412325. Inny przykład: numerem Gödlowskim $P35$ jest 18999899999.

Dla dowolnego wyrażenia E, przez $\overline{E}$ rozumiemy numer Gödlowski E (zapisany jako ciąg cyfr 1, 2, ..., 0).

Nie każda liczba jest numerem Gödlowskim jakiegoś wyrażenia (na przykład, 88 nie jest numerem Gödlowskim żadnego wyrażenia).

Jeśli n *jest* numerem Gödlowskim jakiegoś wyrażenia, to będziemy czasem odwoływać się do tego wyrażenia jako do n-tego wyrażenia. (Dla przykładu, Pd jest szesnastym wyrażeniem, $\perp$ jest drugim wyrażeniem.)

Maszyna jest *samoodnosząca się* (do siebie) w tym sensie, że wyrażenia drukowane przez maszynę stwierdzają, co maszyna może, a czego nie może wydrukować. Wyrażenie nazywamy *drukowalnym*, jeśli maszyna może je wydrukować.

Symbol „P" oznacza „drukowalne" i dla dowolnego wyrażenia E zbudowanego z podanych siedemnastu symboli, jeśli chcemy zapisać zdanie stwierdzające, że E jest drukowalne, to piszemy nie PE, lecz $P\overline{E}$ (tj., P po którym następuje numer Gödlowski E).
Dla przykładu, zdaniem stwierdzającym, że $(P \perp\rightarrow\perp)$ jest drukowalne jest $P\overline{(P \perp\rightarrow\perp)}$ — tj. $P412325$.

Dla dowolnych wyrażeń X oraz Y, Fergusson zdefiniował *diagonalizację X względem Y* jako wyrażenie $(X(\overline{X}, \overline{Y}) \rightarrow Y)$.
Symbol „d" jest skrótem dla „diagonalizacja" — i dla dowolnych wyrażeń X oraz Y, wyrażenie $Pd(\overline{X}, \overline{Y})$ jest zdaniem stwierdzającym, że diagonalizacja X względem Y jest drukowalna.

Zdefiniujemy teraz, co to znaczy, że wyrażenie jest *zdaniem (maszynowym)* i co to znaczy, że zdanie jest *prawdziwe*.

- (1) $\perp$ jest zdaniem i $\perp$ jest fałszywe.
- (2) Dla dowolnego wyrażenia X, wyrażenie $P\overline{X}$ jest zdaniem i jest ono prawdziwe wtedy i tylko wtedy, gdy wyrażenie X jest drukowalne.
- (3) Dla dowolnych wyrażeń X oraz Y, wyrażenie $Pd(\overline{X}, \overline{Y})$ jest zdaniem i jest ono prawdziwe wtedy i tylko wtedy, gdy wyrażenie $(X(\overline{X}, \overline{Y}) \rightarrow Y)$ — które jest diagonalizacją X względem Y — jest drukowalne.
- (4) Dla dowolnych zdań X oraz Y, wyrażenie $(X \rightarrow Y)$ jest zdaniem i jest ono prawdziwe wtedy i tylko wtedy, gdy albo X nie jest prawdziwe, albo Y jest prawdziwe.

Rozumie się, że żadne wyrażenie nie jest zdaniem (maszynowym), jeśli nie zostało otrzymane zgodnie z powyższymi regułami. Spójniki logiczne $\neg, \wedge, \vee, \equiv$ są definiowane z $\rightarrow$ oraz $\perp$ w znany sposób.

Podamy teraz reguły ustalające, co maszyna może wydrukować. Maszyna jest zaprogramowana do kolejnego drukowania nieskończonej listy zdań. Pewne zdania, nazywane *aksjomatami* mogą zostać wydrukowane na każdym etapie tego procesu. Wśród aksjomatów są wszystkie tautologie. (tak więc, dla dowolnej tautologii X, maszyna może wydrukować X kiedy tylko chce, niezależnie od tego, co dotąd wydrukowała lub czego nie wydrukowała w poprzednich etapach.)

Dalej, maszyna jest zaprogramowana tak, że dla dowolnych zdań X oraz Y, jeśli na pewnym etapie maszyna wydrukowała już X oraz $X \rightarrow Y$, to może wydrukować Y. Tak więc, maszyna jest typu 1 (w tym sensie, że zbiór zdań drukowalnych jest typu 1).

Ponieważ jest prawdą, że jeśli X oraz $X \rightarrow Y$ są oba drukowalne, to Y też jest drukowalne, to zdanie $(P\overline{X} \wedge P\overline{(X \rightarrow Y)}) \rightarrow P\overline{Y}$ jest prawdziwe; lub, co na jedno wychodzi, zdanie $P\overline{(X \rightarrow Y)} \rightarrow (P\overline{X} \rightarrow P\overline{Y})$ jest prawdziwe. Maszyna „wie'' zatem o prawdziwości wszystkich zdań postaci $P\overline{(X \rightarrow Y)} \rightarrow (P\overline{X} \rightarrow P\overline{Y})$ i przyjmuje je jako aksjomaty. Tak więc, maszyna jest typu 2.

Następnie, jeśli maszyna kiedykolwiek wydrukuje zdanie X, to „wie" ona, że wydrukowała X i prędzej czy później wydrukuje prawdziwe zdanie $P\overline{X}$. (Zdanie $P\overline{X}$ *jest* prawdziwe, ponieważ X zostało wydrukowane.) A więc maszyna jest normalna, a stąd jest typu 3.
Ponieważ maszyna jest normalna, więc dla dowolnego zdania X, zdanie $P\overline{X} \rightarrow P\overline{P\overline{X}}$ jest prawdziwe. Czyli maszyna jest początkowo „świadoma" prawdziwości wszystkich takich zdań oraz przyjmuje je jako aksjomaty. Zatem maszyna jest typu 4.

Jest jeszcze jedna rzecz, którą maszyna potrafi robić, a jest to rzecz dość istotna. Dla dowolnych wyrażeń X oraz Y, zdanie $Pd(\overline{X}, \overline{Y})$ jest prawdziwe wtedy i tylko wtedy, gdy $(X(\overline{X}, \overline{Y}) \rightarrow Y)$ jest drukowalne, co z kolei zachodzi wtedy i tylko wtedy, gdy zdanie $P\overline{(X(\overline{X}, \overline{Y}) \rightarrow Y)}$ jest prawdziwe. Zatem następujące zdanie jest prawdziwe: $Pd(\overline{X}, \overline{Y}) \equiv P\overline{(X(\overline{X}, \overline{Y}) \rightarrow Y)}$. Maszyna „wie" o prawdziwości wszystkich takich zdań i przyjmuje je jako aksjomaty. Te aksjomaty nazywane są *aksjomatami przekątniowymi.*

Aksjomaty i reguły maszyny

Aksjomaty:

- *Grupa 1.* Wszystkie tautologie.
- *Grupa 2.* Wszystkie zdania postaci $P\overline{(X \to Y)} \to (P\overline{X} \to P\overline{Y})$.
- *Grupa 3.* Wszystkie zdania postaci $P\overline{X} \to P\overline{P\overline{X}}$.
- *Grupa 4 (aksjomaty przekątniowe).* Wszystkie zdania postaci $Pd(\overline{X},\overline{Y}) \equiv P\overline{(X(\overline{X},\overline{Y}) \to Y)}$, gdzie X oraz Y są dowolnymi wyrażeniami (niekoniecznie zdaniami).

Reguły operowania:

- (1) Aksjomaty mogą zostać wydrukowane na każdym etapie.
- (2) Dla dowolnych już wydrukowanych zdań X oraz $(X \to Y)$, maszyna może wydrukować Y.
- (3) Dla dowolnego wydrukowanego już zdania X, maszyna może wydrukować $P\overline{X}$.

Rozumie się, że jedynym sposobem wydrukowania przez maszynę jakiegoś zdania X na pewnym etapie jest zastosowanie się do powyższych reguł. Zatem, X jest drukowalne na danym etapie *tylko* wtedy, gdy zachodzi jeden z następujących trzech warunków: (1) X jest aksjomatem; (2) istnieje zdanie Y takie, że Y oraz $(Y \rightarrow X)$ zostały już wydrukowane na etapie wcześniejszym; (3) istnieje zdanie Y takie, że X jest zdaniem $P\overline{Y}$ oraz Y zostało już wydrukowane na etapie wcześniejszym.

Uwagi. Dla każdego zdania X, niech $\boldsymbol{B}X$ będzie zdaniem $P\overline{X}$. Symbol „$\boldsymbol{B}$" *nie* należy do języka maszyny; używamy go do *mówienia o* maszynie. Używamy „$\boldsymbol{B}$" jako odpowiadającego operacji, która przyporządkowuje każdemu zdaniu X zdanie $P\overline{X}$.

Gdy mówimy, że maszyna jest typu 4, rozumiemy przez to, że jest ona typu 4 ze względu na tę operację $\boldsymbol{B}$. W istocie, bez aksjomatów przekątniowych, system aksjomatyczny tej maszyny jest systemem modalnym K_4. Zobaczymy wkrótce, że dodanie aksjomatów przekątniowych daje nam pełną moc systemu modalnego G.

Dowodliwość. Zdefiniowaliśmy dla każdego zdania maszyny co to znaczy, że zdanie to jest *prawdziwe*, a więc każde zdanie maszyny wyraża określone zdanie, które może być prawdziwe lub może być fałszywe.

Uwaga. Dotąd *proposition* oddawaliśmy zawsze jako *zdanie*. Teraz mamy:

- *zdania (maszyny)* (w oryginale *sentences*) — zdania języka przedmiotowego,
- oraz zdania metajęzyka (w oryginale *propositions*), tj. języka, w którym *mówimy o* maszynie, jej zdaniach (maszynowych), itp.

W przypadkach, gdy mogłoby to prowadzić do nieporozumień, w dalszym ciągu będziemy dodawać określenie *maszynowe*, gdy mowa będzie o zdaniach drukowanych przez maszynę.

Powiemy, że maszyna *dowodzi* danego zdania, gdy drukuje ona zdanie maszynowe, które wyraża to dane zdanie. Dla przykładu, zdanie maszynowe $\neg P2$ wyraża zdanie stwierdzające, że maszyna jest niesprzeczna (ponieważ 2 jest numerem Gödlowskim $\perp$), a więc jeśli maszyna wydrukowała $\neg P2$, to udowodniła swoją własną niesprzeczność. Gdyby maszyna wydrukowała $P2$, to udowodniłaby swoją własną *sprzeczność*.

Powiemy, że maszyna jest *ścisła*, jeśli wszystkie zdania dowodliwe przez maszynę są prawdziwe.
Powiemy, że maszyna jest *niesprzeczna*, jeśli nie może ona dowieść $\perp$, oraz że jest *stabilna*, jeśli dla każdego zdania (maszynowego) X, jeśli $P\overline{X}$ jest drukowalne, to drukowalne jest też X.

Zwrotność. Przechodzimy teraz do dowodu, że maszyna jest Gödlowska, a faktycznie, zwrotna.

Ważne własności maszyny Fergussona

- (1) Znajdziemy zdanie G takie, że zdanie $G \equiv \neg P\overline{G}$ — tj. zdanie $G \equiv (P\overline{G} \rightarrow \perp)$ — jest drukowalne.

- (2) Pokażemy, że dla dowolnego zdania Y istnieje zdanie X takie, że zdanie $X \equiv (P\overline{X} \rightarrow Y)$ jest drukowalne.

Uwaga. Problem 1 jest szczególnym przypadkiem problemu 2, a więc najpierw rozwiążemy problem 2.

Przypomnijmy, że:

- warunek wspomniany w problemie (2) nazywamy zwrotnością;
- systemem typu G nazywamy system modalny typu 4, w którym dowodliwe są wszystkie zdania postaci $\boldsymbol{B}(\boldsymbol{B}p \rightarrow q) \rightarrow \boldsymbol{B}p$.

Niech Y będzie dowolnym zdaniem. Dla dowolnego wyrażenia Z, zdanie $Pd(\overline{Z}, \overline{Y}) \equiv P\overline{(Z(\overline{Z}, \overline{Y}) \rightarrow Y)}$ jest drukowalne (ponieważ jest jednym z aksjomatów przekątniowych).

Weźmiemy za Z wyrażenie Pd i otrzymujemy wtedy, że $Pd(\overline{Pd}, \overline{Y}) \equiv P\overline{(Pd(\overline{Pd}, \overline{Y}) \rightarrow Y)}$ jest drukowalne.

Ponieważ maszyna jest typu 1, więc wynika stąd, że następujące zdanie jest drukowalne:

$$(Pd(\overline{Pd}, \overline{Y}) \rightarrow Y) \equiv (P\overline{(Pd(\overline{Pd}, \overline{Y}) \rightarrow Y)} \rightarrow Y)$$

Tak więc, zdanie $X \equiv (P\overline{X} \rightarrow Y)$ jest drukowalne, gdzie X jest zdaniem $(Pd(\overline{Pd}, \overline{Y}) \rightarrow Y)$.

Problem 1 jest szczególnym przypadkiem problemu 2, gdy za Y weźmiemy $\perp$. Tak więc, zdaniem Gödla G dla tej maszyny jest $Pd(\overline{Pd}, \overline{\perp}) \rightarrow\perp$ — tj., zdanie $(Pd(16, 2) \rightarrow\perp)$.

Co stwierdza zdanie $Pd(16, 2)$?

Mówi ono, że diagonalizacja szesnastego wyrażenia względem drugiego wyrażenia jest drukowalna. Wyrażeniem szesnastym jest Pd, a wyrażeniem drugim jest $\perp$, a więc $Pd(16, 2)$ mówi, że diagonalizacja Pd względem $\perp$ jest drukowalna, ale ta diagonalizacja to zdanie $(Pd(16, 2) \rightarrow\perp)$ — tj. właśnie samo zdanie G!

A więc $Pd(16, 2)$ mówi, że G *jest* drukowalne, a stąd $(Pd(16, 2) \rightarrow\perp)$ — które jest zdaniem G — mówi, że G *nie jest* drukowalne (lub, co na jedno wychodzi, że drukowalność G implikuje fałsz logiczny). Tak więc, G mówi, że G nie jest drukowalne; G jest *prawdziwe* wtedy i tylko wtedy, gdy G nie jest drukowalne.

Zatem G stwierdza swoją własną niedrukowalność. Oto, w miniaturce, pomysłowa idea Gödla otrzymywania samoodniesienia.

Zdanie $G \equiv \neg P\overline{G}$ — tj. zdanie $G \equiv (P\overline{G} \rightarrow \bot)$ — jest nie tylko prawdziwe, ale także drukowalne (jest ono jednym z aksjomatów przekątniowych).

Ponieważ maszyna jest normalna i jest typu 1, wynika stąd na mocy Pierwszego Twierdzenia Gödla o Niezupełności, że jeśli maszyna jest niesprzeczna, to G nie jest drukowalne, a jeśli maszyna jest dodatkowo stabilna, to również $\neg G$ nie jest drukowalne.

A więc, *jeśli* maszyna jest jednocześnie niesprzeczna i stabilna, to zdanie G jest nierozstrzygalne w systemie zdań, które maszyna może wydrukować.

Maszyna jest faktycznie typu 4, a ponieważ jest Gödlowska — zdanie $G \equiv \neg P\overline{G}$ jest drukowalne — więc z Drugiego Twierdzenia Gödla o Niedowodliwości Niesprzeczności wynika, że jeśli maszyna jest niesprzeczna, to nie może ona dowieść swojej własnej niesprzeczności — tj. nie może wydrukować zdania $\neg P2$.

Nadto, jeśli maszyna jest niesprzeczna, to zdanie $\neg P2$ jest *prawdziwe*, a stąd jest innym przykładem zdania prawdziwego, którego maszyna nie może wydrukować.

Co więcej, maszyna jest zwrotna (problem 2), a będąc typu 4, musi być Löbowska (na mocy Twierdzenia Löba), a więc dla dowolnego zdania X, jeśli $P\overline{X} \rightarrow X$ jest drukowalne, to drukowalne jest X. Ponieważ każdy zwrotny Löbowski system typu 4 jest typu G, więc wynika stąd, że maszyna jest typu G.

Czy Maszyna Fergussona jest niesprzeczna?

Poprawność, ścisłość i niesprzeczność Maszyny Fergussona.

Pokazaliśmy, że *jeśli* maszyna Fergussona jest niesprzeczna, to nie może udowodnić swojej własnej niesprzeczności.

Ale skąd wiemy, czy maszyna jest, czy nie jest niesprzeczna?

Udowodnimy teraz, że maszyna jest nie tylko niesprzeczna, ale że jest też całkowicie ścisła — tj., że każde zdanie wydrukowane przez maszynę jest prawdziwe.

Pokazaliśmy już, że wszystkie *aksjomaty* maszyny są prawdziwe, ale prześledźmy uważnie to rozumowanie.

Aksjomaty Grupy 1 są wszystkie tautologiami, a stąd są z pewnością prawdziwe.

Jeśli chodzi o aksjomaty Grupy 2, to powiedzieć, że $P\overline{(X \to Y) \to (P\overline{X} \to \overline{Y})}$ jest prawdziwe to tyle, co powiedzieć, że jeśli oba $P\overline{(X \to Y)}$ oraz $P\overline{X}$ są prawdziwe, to takie jest też $P\overline{Y}$, czyli to samo, co powiedzieć, że jeśli $(X \to Y)$ oraz X są oba drukowalne, to takie jest też Y.
A tak oczywiście jest, na mocy Operacji 2.

Tak więc, aksjomaty Grupy 2 są wszystkie prawdziwe.

Jeśli chodzi o aksjomaty Grupy 3, powiedzieć, że $P\overline{X} \to P\overline{P\overline{X}}$ jest prawdziwe, to tyle, co powiedzieć, że jeśli $P\overline{X}$ jest prawdziwe, to takie jest też $P\overline{P\overline{X}}$.

To z kolei jest tym samym, co powiedzenie, że jeśli X jest drukowalne, to takie jest też $P\overline{X}$ — a tak jest rzeczywiście, na mocy Operacji 3.

Jeśli chodzi o aksjomaty przekątniowe, to $Pd(\overline{X}, \overline{Y})$ jest prawdziwe wtedy i tylko wtedy, gdy $(X(\overline{X}, \overline{Y}) \to Y)$ jest drukowalne, a tak jest wtedy i tylko wtedy, gdy $P\overline{(X(\overline{X}, \overline{Y}) \to Y)}$ jest prawdziwe.

Zatem $Pd(\overline{X}, \overline{Y}) \equiv P\overline{(X(\overline{X}, \overline{Y}) \to Y)}$ jest prawdziwe.

Wiemy teraz, że wszystkie aksjomaty maszyny są prawdziwe, ale musimy pokazać, że wszystkie zdania *drukowalne* są prawdziwe.

Przypomnijmy, że maszyna drukuje zdania na pewnych *etapach*. Chcemy teraz ustanowić następujący lemat, twierdzenie i wniosek:

- **Lemat.** Jeśli X jest zdaniem wydrukowanym na pewnym etapie i wszystkie zdania wydrukowane do tego etapu są prawdziwe, to X jest prawdziwe.
- **Twierdzenie.** Każde zdanie wydrukowane przez maszynę jest prawdziwe.
- **Wniosek.** Maszyna jest jednocześnie niesprzeczna i stabilna.

Dowody.

Najpierw udowodnimy lemat. Załóżmy, że wszystkie dotąd wydrukowane zdania są prawdziwe; mamy pokazać, że X jest prawdziwe.

Przypadek 1. X jest aksjomatem. Wtedy X jest prawdziwe (jak już udowodniliśmy).
Przypadek 2. Istnieje zdanie Y takie, że Y oraz $(Y \rightarrow X)$ zostały już wydrukowane. Wtedy z przyjętego założenia Y oraz $(Y \rightarrow X)$ są oba prawdziwe, a więc X jest prawdziwe.
Przypadek 3. X jest postaci $P\overline{Y}$, gdzie Y jest zdaniem, które już zostało wydrukowane. Ponieważ Y zostało wydrukowane, więc $P\overline{Y}$ jest prawdziwe — tj. X jest prawdziwe.

To kończy dowód lematu.

Dowód Twierdzenia.

Maszyna jest zaprogramowana tak, aby wydrukować wszystkie drukowalne zdania w jakimś określonym ciągu $X_1, X_2, \ldots, X_n, \ldots$ Przez X_n rozumiemy zdanie wydrukowane na etapie n.
Pierwsze zdanie wydrukowane przez maszynę (zdanie X_1) musi być aksjomatem (ponieważ dotąd maszyna nie wydrukowała żadnych zdań), a stąd X_1 musi być prawdziwe.
Jeśli powyższa lista zawierałaby jakiekolwiek zdanie fałszywe, to musiałaby istnieć *najmniejsza* liczba n taka, że X_n jest fałszywe — to jest, musiałoby istnieć *pierwsze* zdanie fałszywe wydrukowane przez maszynę. Wiemy, że n nie jest równe 1 (ponieważ X_1 jest prawdziwe), a zatem n jest większe od 1. Znaczy to, że maszyna drukuje zdanie fałszywe na etapie n, ale na wszystkich wcześniejszych etapach drukowała wyłącznie zdania prawdziwe. Przeczy to jednak lematowi.

Zatem maszyna nigdy nie może wydrukować jakichkolwiek zdań fałszywych.

Dowód Wniosku.

Ponieważ maszyna jest ścisła (na mocy Twierdzenia), więc $\perp$ nigdy nie może zostać wydrukowane, ponieważ $\perp$ jest fałszywe. Zatem maszyna jest niesprzeczna.

Następnie, przypuśćmy, że $P\overline{X}$ jest drukowalne. Wtedy $P\overline{X}$ jest prawdziwe (na mocy Twierdzenia), co oznacza, że X jest drukowalne. Zatem maszyna jest stabilna.

Widzimy teraz, że maszyna Fergussona *jest* niesprzeczna, ale nigdy nie potrafi dowieść swojej niesprzeczności. Tak więc i ty i ja (równie dobrze jak Fergusson) wiemy, że maszyna jest niesprzeczna, ale biedna maszyna wiedzy tej nie ma!

Koniec dodatku B

O dalszych wynikach związanych z „maszynową" interpretacją twierdzeń metalogicznych traktuje rozdział 28 *Forever Undecided*.

W szczególności, podane są związki między maszynami logicznymi a samostosowalnymi systemami modalnymi.