Twierdzenia Gödla – pogadanka

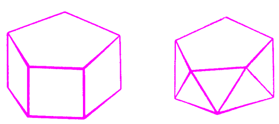
*Leszek KOŁODZIEJCZYK**

28 kwietnia br. minęła setna rocznica urodzin Kurta Gödla, jednego z wielkich matematyków ubiegłego wieku, a zarazem „największego logika od czasów Arystotelesa”, wedle powszechnie podzielanej opinii von Neumanna. Warto z tej okazji przypomnieć najznakomitsze (choć bynajmniej nie jedyne!) osiągnięcie matematyczne Gödla, czyli dwa twierdzenia o niezupełności silnych teorii aksjomatycznych.

Aby lepiej zrozumieć znaczenie twierdzeń o niezupełności i wpływ, jaki wywarły, dobrze jest uświadomić sobie kontekst historyczny, w którym się pojawiły. Koniec XIX w. i pierwsze dziesięciolecia XX w. to w matematyce okres burzliwego rozwoju nowych metod. Coraz śmielej stosowano zwłaszcza abstrakcyjne techniki rozwiniętej przez Georga Cantora teorii mnogości (teorii zbiorów). Metody teoriomnogościowe szybko doprowadziły do uściślenia niejasnych dotąd pojęć i uzyskania fascynujących wyników. Wyniki te bywały jednak nieintuicyjne, a czasem wysoce nieprzyjemne (np. istnienie zbioru na prostej, któremu nie da się w sensowny sposób przypisać miary). Co gorsza, beztroskie używanie zbiorów prowadziło do paradoksów (z których najlepiej znanym jest tzw. paradoks Russella: czy zbiór wszystkich zbiorów, które nie są elementami samych siebie, jest elementem samego siebie?). Stopniowo rodziło się poczucie, że matematyka przeżywa kryzys swoich podstaw.

W odpowiedzi na kryzys liczni matematycy, w tym tak wybitni jak Poincaré, Brouwer czy Weyl, sugerowali ograniczenia (nieraz radykalne) zestawu dopuszczalnych w matematyce środków dowodowych. Zupełnie inny sposób wyjścia z kryzysu proponował Dawid Hilbert, bezwzględnie przeciwny wszelkim próbom okrajania matematyki. Sednem tzw. *programu Hilberta*, sformułowanego w latach 20. ubiegłego wieku, było wyodrębnienie pewnej klasy metod uznanych za bezwzględnie pewne. Całkowicie bezpieczne miały być metody finitystyczne, czyli, w uproszczeniu, odwołujące się wyłącznie do konkretnych, skończonych obiektów (takich jak liczby naturalne). Metody infinitystyczne – odwołujące się do zbiorów nieskończonych – wymagały natomiast usprawiedliwienia. Hilbert miał przy tym pomysł na to, jak je usprawiedliwić.

Według Hilberta należało, po pierwsze, sformalizować całą matematykę w postaci jednej teorii aksjomatycznej. Teoria miała być zupełna, tj. umożliwiać rozstrzygnięcie każdego problemu dającego się wyrazić w jej języku. Po drugie zaś, należało pokazać, że teoria ta jest niesprzeczna, czyli że nie można w niej udowodnić zdania wewnętrznie sprzecznego (np. $0 \neq 0$). Co więcej, dowód niesprzeczności miał być całkowicie pewny, powinien zatem używać jedynie środków finitystycznych; zwróćmy uwagę, że jest to *a priori* możliwe, jako że dowody w systemach formalnych to konkretne, skończone obiekty kombinatoryczne (skończone ciągi napisów konstruowane według pewnych reguł).



nieskończone serie to $(4, 4, n)$ – czyli graniastosłupy i $(3, 3, 3, n)$ – czyli antygraniastosłupy. Na rysunku oczywiście $n = 5$.

Dowód taki gwarantowałby przynajmniej tyle, że używanie wszelkich środków ówczesnej matematyki, w tym infinitystycznych, nie doprowadzi do katastrofy, jaką byłaby sprzeczność. W istocie zaś oznaczałby nawet, że wszystkie „najbardziej konkretne zdania” (w żargonie logicznym: Π^0_1-zdania), mające dowód w ogóle, mają też dowód finitystyczny.

Hilbert pokładał w swym programie wielkie nadzieje, licząc na to, że pozwoli na „rozwiązanie problemu podstaw matematyki raz na zawsze”. Udało się zresztą poczynić pewne drobne kroki na drodze do realizacji programu. W 1931 r. ukazała się jednak praca Gödla *Über formal unentscheidbare Sätze der ‘Principia Mathematica’ und verwandter Systeme*, która nadzieje Hilberta zniweczyła.

Praca zawierała dowód tzw. pierwszego twierdzenia o niezupełności: każda dostatecznie silna teoria aksjomatyczna jest niezupełna (nie może zatem istnieć

*Instytut Matematyki, Uniwersytet Warszawski

zupełna teoria obejmująca całą matematykę). Co gorsza, Gödel zaanonsował też drugie twierdzenie o niezupełności: żadna dostatecznie silna teoria aksjomatyczna nie dowodzi własnej niesprzeczności (skoro zaś niesprzeczność silnej teorii nie jest nawet dowodliwa w niej samej, tym bardziej nie ma co marzyć o dowodzie finitystycznym).

Zanim powiemy kilka słów o dowodzie twierdzeń o niezupełności, sformułujmy same twierdzenia nieco dokładniej (acz, z różnych względów, nie całkiem dokładnie).

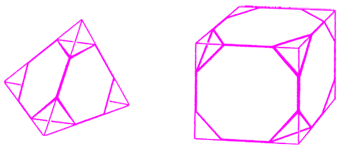
Pierwsze twierdzenie o niezupełności
Niech T będzie niesprzeczną teorią aksjomatyczną zawierającą pewien ustalony fragment arytmetyki. Załóżmy ponadto, że własność bycia aksjomatem T jest algorytmicznie rozpoznawalna. Wtedy T jest niezupełna: istnieje zdanie φ w języku teorii T nierozstrzygalne w T, czyli takie, że $T \nvdash \varphi$ i $T \nvdash \neg\varphi$.

Zapis $T \vdash \psi$ oznacza, że T dowodzi zdania ψ, a $\neg\varphi$ oznacza zaprzeczenie φ.

Drugie twierdzenie o niezupełności
Przy założeniach pierwszego twierdzenia, jeśli Con_T jest zdaniem w języku T wyrażającym w sposób naturalny niesprzeczność T, to $T \nvdash \mathrm{Con}_T$.

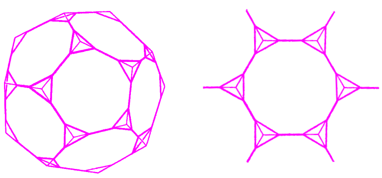
Niektóre zwroty użyte w powyższych sformułowaniach wymagają zapewne wyjaśnień.

Po pierwsze, co to znaczy, że T ma zawierać pewien fragment arytmetyki? Wystarczy tyle, by w teorii T można było udowodnić pewne fakty dotyczące arytmetyki liczb naturalnych. Dziś wiemy, że może to być raptem kilka prostych faktów typu: dodawanie liczb naturalnych jest łączne, każda liczba naturalna jest zerem lub jest postaci $n+1$, itp. Nie trzeba przy tym zakładać, że język T zawiera pojęcia arytmetyczne (wystarczy, że dają się one wewnątrz T zdefiniować).



z czworościanu i sześcianu wycinamy (3, 6, 6) i (3, 8, 8)

Dalej, jaki jest sens założenia o algorytmicznej rozpoznawalności aksjomatów T i czy nie jest to istotne ograniczenie? Wymagamy tyle, by istniała mechaniczna procedura rozpoznawania, czy dany ciąg znaków jest aksjomatem T (pojęcie „mechanicznej procedury" można uściślić, używając np. teorii maszyn Turinga, ale jego intuicyjny sens powinien być w epoce komputerów dość jasny). Założenie to, spełnione przez wszystkie teorie aksjomatyczne spotykane w praktyce, jest istotne z punktu widzenia matematycznego, ale w pełni naturalne z punktu widzenia „ludzkiego": jeśli nie wiadomo nawet, jak rozstrzygnąć, co jest aksjomatem danej teorii, a co nie, to taka teoria jest dla nas bezużyteczna.

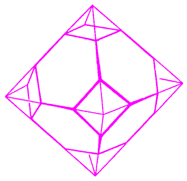


z dwunastościanu i parkietażu sześciokątnego otrzymujemy (3, 10, 10) i (3, 12, 12)

Kolejna rzecz: zakładamy tylko tyle, że T „wie" trochę na temat liczb naturalnych. W szczególności, język T może zawierać np. wyłącznie pojęcia arytmetyczne. Jakże zatem wyrazić niesprzeczność T, czyli pewien fakt dotyczący dowodów formalnych, a nie liczb, w tymże języku? Idea jest całkiem naturalna. Obiekty syntaktyczne, takie jak zmienne, formuły, dowody w teorii aksjomatycznej itp., to skończone ciągi znaków – możemy je np. zakodować jako skończone ciągi zerojedynkowe. Ciąg zerojedynkowy można z kolei bez trudu utożsamić z liczbą naturalną (np. dopisując z przodu jedynkę i traktując go jako zapis liczby w systemie dwójkowym). Tak więc, każdemu obiektowi syntaktycznemu można w naturalny sposób przyporządkować liczbę (tzw. numer Gödla). W dodatku (co jest już faktem wymagającym nietrywialnego uzasadnienia) własności, takie jak „x jest numerem Gödla formuły zaczynającej się od kwantyfikatora ogólnego" albo „x jest numerem ciągu będącego dowodem zdania o numerze y" można wyrazić formułami języka T, i to w taki sposób, że da się w teorii T o tych własnościach rozumować. Nam szczególnie potrzebna będzie formuła $\mathrm{Pr}(x)$: „x jest numerem zdania dowodliwego w T". Niesprzeczność T, czyli Con_T, możemy teraz sformułować np. jako $\neg\,\mathrm{Pr}(\ulcorner 0 \neq 0 \urcorner)$.

$\ulcorner\varphi\urcorner$ to konwencjonalne oznaczenie numeru Gödla zdania φ

Możemy teraz przystąpić do naszkicowania dowodu obu twierdzeń Gödla. Kluczową rolę w dowodzie gra odpowiednio zmodyfikowany paradoks kłamcy (czy zdanie „to zdanie jest fałszywe" jest prawdziwe?), w którym prawdziwość



z ośmiościanu wycinamy czternastościan (4, 6, 6) (można go też wyciąć z sześcianu); w *Delcie* pisaliśmy o nim w numerze 9/1996 i 2/2005

zostaje zastąpiona dowodliwością. Aby można było go sformułować, potrzebny jest następujący fundamentalny lemat.

Lemat przekątniowy

Niech $\varphi(x)$ będzie formułą w języku teorii T z jedną zmienną wolną x. Istnieje takie zdanie ψ, że $T \vdash (\psi \iff \varphi(\ulcorner\psi\urcorner))$.

Mówiąc swobodnie: dla każdej własności wyrażalnej w języku teorii T istnieje zdanie, które „orzeka o sobie", że ma tę własność.

Innym ważnym elementem dowodu, zwłaszcza drugiego twierdzenia, jest obserwacja, że jeśli T jest dostatecznie silna, to wprowadzona powyżej formuła $\mathrm{Pr}(x)$, wyrażająca dowodliwość w teorii T, spełnia trzy warunki, zwane tradycyjnie *warunkami wywodliwości*:

(D1) jeśli $T \vdash \varphi$, to $T \vdash \mathrm{Pr}(\ulcorner\varphi\urcorner)$,

(D2) $T \vdash (\mathrm{Pr}(\ulcorner\varphi\urcorner) \to \mathrm{Pr}(\ulcorner\mathrm{Pr}(\ulcorner\varphi\urcorner)\urcorner))$,

(D3) $T \vdash (\mathrm{Pr}(\ulcorner\varphi \to \psi\urcorner) \,\&\, \mathrm{Pr}(\ulcorner\varphi\urcorner) \to \mathrm{Pr}(\ulcorner\psi\urcorner))$.

Intuicyjnie: (D1) – jeśli jakieś zdanie jest dowodliwe w T, to T dowodzi tego faktu; co więcej, (D2) – samo (D1) jest dowodliwe w T; wreszcie, (D3) – dowodliwość w T jest (dowodliwie w T!) zamknięta na pewną standardową regułę wnioskowania (tzw. *modus ponens*).

W istocie nasz argument działa przy założeniu nieco wzmocnionego warunku niesprzeczności T, tzw. ω-niesprzeczności; jeśli chce się przeprowadzić dowód wyłącznie przy założeniu zwykłej niesprzeczności, trzeba nieco zmodyfikować zdanie γ.

Gdy mamy już do dyspozycji lemat przekątniowy i warunki wywodliwości, reszta dowodu twierdzeń Gödla jest już stosunkowo nietrudnym rachunkiem. Rozważmy bowiem dane przez lemat takie zdanie γ, że $T \vdash (\gamma \iff \neg\,\mathrm{Pr}(\ulcorner\gamma\urcorner))$. Zdanie γ „mówi" zatem, że samo nie jest dowodliwe w teorii T. Okazuje się, że $T \nvdash \gamma$ oraz $T \nvdash \neg\gamma$. Tak więc, γ jest zdaniem, o którym mowa w pierwszym twierdzeniu Gödla.

Pokażmy, że $T \nvdash \gamma$. Jeśli $T \vdash \gamma$, to na mocy warunku (D1) jest także $T \vdash \mathrm{Pr}(\ulcorner\gamma\urcorner)$. Ale $\mathrm{Pr}(\ulcorner\gamma\urcorner)$ jest równoważne $\neg\gamma$. A zatem $T \vdash \gamma$ i $T \vdash \neg\gamma$, wbrew założeniu niesprzeczności T.

Formalny dowód, że również $T \nvdash \neg\gamma$, pominiemy, zauważając jedynie, że $T \nvdash \neg\gamma$ musi zachodzić, jeśli tylko teoria T nie dowodzi żadnego zdania *fałszywego*. Otóż zdanie $\neg\gamma$ jest właśnie fałszywe: mówi bowiem, że istnieje dowód zdania γ w teorii T. Ale przed chwilą pokazaliśmy, że taki dowód nie istnieje!

Aby udowodnić drugie twierdzenie Gödla, wystarczy pokazać, że $T \vdash (\mathrm{Con}_T \to \gamma)$. W tym celu użyjemy warunków wywodliwości (D1)–(D3). Mamy $T \vdash (\gamma \iff \neg\,\mathrm{Pr}(\ulcorner\gamma\urcorner))$, a zatem, na mocy (D1) i (D3), także

$$T \vdash (\mathrm{Pr}(\ulcorner\gamma\urcorner) \iff \mathrm{Pr}(\ulcorner\neg\,\mathrm{Pr}(\ulcorner\gamma\urcorner)\urcorner)).$$

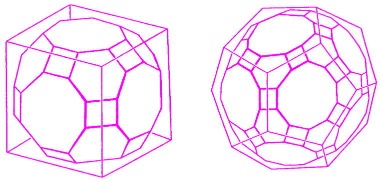
Z drugiej strony, (D2) daje

$$T \vdash (\mathrm{Pr}(\ulcorner\gamma\urcorner) \to \mathrm{Pr}(\ulcorner\mathrm{Pr}(\ulcorner\gamma\urcorner)\urcorner)).$$

Ale dla zupełnie dowolnego zdania ψ

$$T \vdash (\mathrm{Pr}(\ulcorner\psi\urcorner) \,\&\, \mathrm{Pr}(\ulcorner\neg\psi\urcorner) \to \mathrm{Pr}(\ulcorner 0 \neq 0\urcorner))$$

(co wynika z (D3) oraz faktu, że T dowodzi wszystkich podstawień tautologii rachunku zdań). Ostatecznie zatem dostajemy $T \vdash (\mathrm{Pr}(\ulcorner\gamma\urcorner) \to \mathrm{Pr}(\ulcorner 0 \neq 0\urcorner))$, czyli $T \vdash (\mathrm{Con}_T \to \gamma)$. Szkic dowodu twierdzeń o niezupełności został więc zakończony.



z sześcianu i dwunastościanu wycinamy, choć w bardziej skomplikowany sposób (4, 6, 8) i (4, 6, 10)

Jak już wspominaliśmy, wyniki Gödla obaliły nadzieje (oddaliły obawy?), że całą matematykę da się sprowadzić do rozumowania wewnątrz jednego ustalonego systemu aksjomatycznego. Zrodziły także dalsze pytania. Przykładowo: czy dałoby się może zaksjomatyzować całą „spotykaną w praktyce" matematykę bądź „matematykę stosowaną w naukach przyrodniczych"? Albo: jak dalece trzeba wykroczyć poza daną teorię aksjomatyczną, aby móc udowodnić jej niesprzeczność? Te i inne pytania inspirowały następnie bujny rozwój nowych gałęzi logiki matematycznej. To już jednak tematy na inne pogadanki.