Semina □ Nr 3
Scientiarum 2004

Maria Piesko

Twierdzenie Gödla i marzenie Leibniza[1]

Podczas jednego z seminariów z kognitywistyki prowadzący je prof. J. Perzanowski wypowiedział sąd, który pozostał dla mnie przez jakiś czas zagadkowy, mimo iż nie zupełnie obca była mi filozofia Leibniza i sławny artykuł Gödla „Über formal unentscheidbare Sätze…”[2]. Profesor stwierdził mianowicie, że niesłuszne jest mniemanie, jakoby twierdzenie Gödla o niezupełności matematyki obaliło projekt Leibniza znalezienia *scientia universalis* — uniwersalnej nauki posługującej się *lingua characteristica* — idealnym językiem opisującym całość rzeczywistości w tak doskonały sposób, że wszelkie rozumowania (w szczególności filozoficzne) można by sprowadzić do odpowiednich obliczeń przeprowadzanych w specjalnym rachunku (*calculus ratiocinator*). „Gödel odkrył po prostu zdanie przypadkowe *a priori*”. Odkrył zatem, jak tłumaczył profesor Perzanowski, że nie wszystkie prawdy matematyczne są konieczne. Cóż to znaczy? Pytanie okazało się tym bardziej niepokojące, że konkluzję, iż twierdzenia Gödla (a dokładniej późniejszy dowód Churcha nierozstrzygalności rachunku predykatów, którego konsekwencją jest niezupełność tego rachunku) pokazały

[1]Za przejrzenie tego artykułu i cenne uwagi pragnę podziękować Panu Profesorowi Jerzemu Perzanowskiemu, Księdzu Doktorowi Adamowi Olszewskiemu i Panu Doktorowi Bartoszowi Brożkowi. Nie ponoszą oni oczywiście odpowiedzialności za treść tego tekstu (niestety nie udało mi się zgodzić ze wszystkimi ich twierdzeniami).

[2]K. Gödel, „Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I”, *Monatshefte für Mathematik und Physik*, 38 (1931), ss. 173–198; angielski przekład: „On Formally Undecidable Propositions of Principia Mathematica and Related Systems I”, *The Undecidable. Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, M. Davies (ed.), Raven Press, New York, 1965, ss. 5–38.

nierealizowalność marzenia Leibniza, znalazłam w tak znakomitej książce, jaką jest *Classical Recursion Theory* P. Odifreddiego[3].

Wydaje się, że wynik Gödla godzi w Leibnizjański plan stworzenia *lingua characteristica* — możliwości systemów formalnych okazały się ograniczone. Nadzieję na *calculus ratiocinator* zdają się ostatecznie przekreślać twierdzenia o nierozstrzygalności i teza Churcha–Turinga. W poniższym tekście zamierzam rozważyć, czy tak jest rzeczywiście.

Przyjrzyjmy się najpierw twierdzeniu Gödla. Głosi ono, iż jeśli teoria T jest niesprzeczna, „zawiera" elementarną arytmetykę (tzw. „słabą arytmetykę") i jest oparta o rozstrzygalny zbiór aksjomatów, to jest ona niezupełna. Zatem w języku teorii T można utworzyć zgodnie z regułami formacji zdanie G takie, że ani zadanie G ani jego zaprzeczenie $\neg G$ nie są twierdzeniami teorii T (nie istnieje dla nich dowód w T).

Na pierwszy rzut oka można więc stwierdzić, że rzeczywiście wynik ten zdaje się zagrażać planom Leibniza, jego sławnemu postulatowi *Calculemus!* Jeśli filozofowie pokłóciliby się o prawdziwość zdania G, ich sporu nie można by rozstrzygnąć przez proste odwołanie się do rachunków.

Jak wiadomo, nie zawsze to, co wydaje się oczywiste na pierwszy rzut oka, jest prawdziwe. Przyjrzyjmy się zatem twierdzeniu uważniej. Do rozpatrzenia pozostają:

- założenia,
- wniosek,
- dowód.

[3] P. Odifreddi, *Classical Recursion Theory. The Theory of Function and Sets of Natural Numbers*, Elsevier, Amsterdam – New York – Tokyo, s. 164.

Dowód

Dowód w różnych, bardziej i mniej technicznych wersjach, został przytoczony w tym numerze czasopisma. Nadto, żeby wskazać na wagę założeń, które pragnę podkreślić, należałoby prześledzić dość żmudne, techniczne szczegóły, które stosunkowo łatwo znaleźć w książkach dotyczących teorii rekursji. Pozwolę sobie zatem pominąć dowód, odsyłając Czytelnika do literatury przedmiotu[4] i prosząc, by zechciał uwierzyć (lub sprawdzić), że założenia, które poniżej rozważam, są w dowodzie twierdzeń Gödla rzeczywiście konieczne.

Założenia

Najmniej interesujące z naszego punktu widzenia jest założenie, by teoria T „zawierała” arytmetykę, choć jest to ważne i ciekawe wymaganie. Wyrażalność arytmetyki w teorii T umożliwia arytmetyzację języka i metajęzyka tej teorii i dalej pozwala na diagonalizację, która jest jednym z podstawowych narzędzi teorii rekursji wykorzystywanym w dowodach nierozwiązywalności różnych problemów. Niemniej, skoro według Leibniza rozwiązywanie problemów miałoby się dokonywać w oparciu o obliczenia, niecelowe wydaje się zubażanie poszukiwanego rachunku o tak podstawowe narzędzie jak arytmetyka i to nawet w znacznie okrojonej wersji (można w przybliżeniu powiedzieć, że tzw. „słaba arytmetyka” jest jedynie nieskończoną tabliczką dodawania i mnożenia).

Niezbyt rozsądne wydaje się również podważenie podstawowej zasady rozumowania (także u Leibniza) — zasady niesprzecz-

[4] Por. np. cytowaną wyżej pracę Odifreddiego albo w języku polskim: R. Murawski, *Funkcje rekurencyjne i elementy matematyki*, Wyd. UAM, Poznań 2000.

ności[5]. Wprawdzie sprzeczna teoria jest zupełna[6], ale też, skoro wszystkiego da się w niej dowieść, to i wszystkie dowody wydają się bezwartościowe. Nie pozwalają na odróżnienie tego, co uzasadnione od tego, co bezpodstawne. Nadto rachunek Leibniza miał za zadanie opisywać rzeczywistość, a ta zdaniem autora *Monadologii*, jest niesprzeczna.

[5]Tu warto by dodać, że we współczesnej logice rozwijane są badania rachunków z różnymi „rodzajami sprzeczności" (por. B. Brożek, „Nauka w poszukiwaniu logiki", *Semina Scientiarum* 1, 2002, ss. 2–14). Często jednak w dyskusjach na temat różnych logik nieklasycznych, także niemonotonicznych, podkreśla się, że logika klasyczna (z zasadą niesprzeczności) wydaje się pełnić wobec nich wyróżnioną rolę — to w niej na przykład przeprowadza się analizę rachunków niestandardowych. Przywodzi to na myśl (dość luźne) skojarzenia ze zwykłym sposobem rozumowania ludzi, którzy niejednokrotnie uznają sprzeczne tezy, ale kiedy tylko zdają sobie z tego sprawę, starają się tę sprzeczność usunąć. Oczywiście, że powyższe zestawienie jest dość powierzchowne, sądzę jednak, że za wskazaną odległą analogią może się ukrywać głębsza prawidłowość — trudno rozstrzygnąć czy ontologiczna czy psychologiczna. To ciekawe zagadnienie w oczywisty sposób wykracza jednak poza przedmiot tego artykułu. Zrezygnowanie z założenia niesprzeczności wydaje się być sprzeczne (!) z filozofią Leibniza.

[6]Warto w tym punkcie zauważyć rzecz nieco zaskakującą na terenie logiki, że pojęcie „zupełności" jest wieloznaczne. Na szczęście każde spośród jego różnych znaczeń jest ściśle określone. W *Małej Encyklopedii Logiki* W. Marciszewski wśród różnych możliwych znaczeń „zupełności systemu" wyróżnia trzy:

1) System S jest zupełnym zbiorem zdań zawierającym terminy stałe $P_1, \ldots P_n$ wtedy i tylko wtedy, gdy dla każdego zdania A zawierającego jako symbole stałe jedynie wyrażenia spośród $P_1, \ldots P_n$ prawdą jest, że $A \in S$ lub $\neg A \ \in S$.
2) System jest zupełny wtedy i tylko wtedy, gdy każda poprawnie zbudowana formuła bądź jest twierdzeniem systemu, bądź po dołączeniu do jego aksjomatów wprowadzi doń sprzeczność (tzw. zupełność w sensie Posta).
3) Trzecie pojęcie zupełności wyraża się w polskim piśmiennictwie raczej słowem pełność (definicję pełności podaję poniżej).

Por. *Mała encyklopedia logiki*, red. W. Marciszewski, Ossolineum, Wrocław–Warszawa–Kraków, 1970.
System sprzeczny jest zupełny w sensie (1).

Pozostaje do rozważenia trzecie założenie — o rozstrzygalności zbioru aksjomatów. Wydaje się ono bardzo dobrze uzasadnione: jeśli nie umielibyśmy rozstrzygnąć, co jest a co nie jest aksjomatem, tym bardziej trudno oczekiwać, byśmy potrafili stwierdzić, co jest twierdzeniem teorii (które przecież jest z definicji zdaniem, dla którego istnieje dowód, czyli skończony ciąg zdań, które albo są aksjomatami, albo też są zdaniami uzyskanymi za pomocą reguł dowodzenia z aksjomatów lub podobnie uzyskanych zdań występujących uprzednio w dowodzie).

To wskazuje na jeszcze jedno założenie — nie wyrażone *explicite*, gdyż przyjmowane jako oczywiste i powszechnie akceptowane, czyli wypracowaną w szkole Hilberta technikę uzasadniania zdań matematyki — technikę formalnych dowodów. Zwróćmy przede wszystkim uwagę na podstawowe wymaganie wobec dowodu: ma być on skończony. Fakt ten jest ważny i wykorzystywany w dowodzie twierdzenia Gödla.

Na związek tego poniekąd ukrytego założenia z rozumowaniem Leibniza zwrócił uwagę prof. Perzanowski na kolejnym seminarium, przypominając, klasyczny podział sądów przeprowadzony przez Leibniza:

- sądy konieczne (oparte o zasadę niesprzeczności),
- sądy przypadkowe (oparte o zasadę racji dostatecznej).

Można pokazać (i jest to czasem przyczyną rozczarowania studentów filozofii zapoznających się z dziełami Leibniza), że wnioskowanie w oparciu o zasadę racji dostatecznej sprowadza się *de facto* do odwołania do zasady niesprzeczności. Chciałoby się powiedzieć: „Cóż to za podział”, skoro jedna z proponowanych kategorii zawiera się całkowicie w drugiej? A jednak wskazane przez Leibniza kryterium rozróżniania tych dwóch rodzajów sądów jest istotne. Według Leibniza bowiem do sądów koniecznych można dojść, wychodząc od zasady niesprzeczności za pomocą skończonej liczby kroków wnioskowania, sądy przypadkowe zaś można wprawdzie otrzymać na mocy samej tylko zasady niesprzeczności, ale przeprowadzając dowody o nieskończonej liczbie kroków. Twierdzenie Gödla pokazuje, jak ważne (na pociechę studentom filozofii) jest to rozróżnienie — za pomocą dowodów o skończonej długości, wychodząc od rozstrzygalnego zbioru aksjomatów, nie możemy pewnych twierdzeń dowieść. Jakiego rodzaju są to twierdzenia?

Wniosek

Przyjrzyjmy się krótko wnioskowi twierdzenia Gödla. Żeby go lepiej zrozumieć zestawmy go z innym twierdzeniem genialnego logika — o pełności rachunku predykatów pierwszego rzędu. System dedukcyjny jest pełny wtedy i tylko wtedy, gdy z jego aksjomatów dadzą się wywieść wszystkie zdania będące zdaniami prawdziwymi w każdym modelu. Rachunek predykatów jest zatem takim systemem, który jest pełny (każde zdanie prawdziwe w każdym modelu jest w nim dowiedlne), ale nie jest zupełny (można w nim wyrazić zdania, które ani nie są dowiedlne, ani nie są dowiedlne ich zaprzeczenia). Jak to możliwe? Otóż zdania niedowiedlne w rachunku predykatów, to takie zdania, które ani nie są prawdziwe we wszystkich modelach, ani też nie są we wszystkich modelach fałszywe (wówczas ich zaprzeczenie byłoby prawdziwe we wszystkich modelach, a więc dowiedlne). Są to zatem zdania przypadkowe. W szczególności zdanie Gödla jest takim zdaniem przypadkowym (bo niekoniecznym; nie jest prawdziwe we wszystkich możliwych światach — modelach). Jest ono mimo to zdaniem *a priori*, gdyż o tej jego przypadkowości nie wnioskujemy na podstawie doświadczenia, możemy ją wykazać teoretycznie przytaczając dowód Gödla.

***

Można spekulować, że jeśli potrafilibyśmy podać system (*lingua characteristica*?), który adekwatnie opisywałby zamierzony model (rzeczywistość) do ostatniego szczegółu, w taki sposób, że jednoznacznie wyróżniałby (z dokładnością do izomorfizmu) jeden jedyny model, dla którego byłby prawdziwy i pełny, to wówczas byłby on także zupełny. Kolejne twierdzenia limitacyjne mówią, że opis taki przekracza możliwości języków elementarnych[7].

Mimo tego, zgodnie z twierdzeniem Lindenbauma, każdą niesprzeczną teorię można w prosty sposób rozszerzyć do teorii niesprzecznej oraz zupełnej. Nie podważa to bynajmniej wyniku Gödla, gdyż teoria uzyskana „sposobem Lindenbauma" nie spełnia po prostu jednego z założeń jego twierdzenia — założenia o rozstrzygalności zbioru aksjomatów.

Zagadnieniu rozstrzygalności warto przyjrzeć się uważniej. Wiadomo, że wynik Gödla można otrzymać z dowiedzionego 6 lat później przez Churcha i niezależnie przez Turinga twierdzenia o nierozstrzygalności rachunku predykatów. Jeśli bowiem teoria (o rozstrzygalnym zbiorze aksjomatów) byłaby zupełna, to dla każdej formuły istniałby albo jej dowód, albo dowód jej zaprzeczenia. Wówczas o każdym zdaniu moglibyśmy za pomocą dowodów rozstrzygać[8], czy przynależy ono do teorii czy też nie. Zatem skoro teoria jest nierozstrzygalna, to jest także niezupełna. I znowu przyjrzyjmy się technicznemu pojęciu zaangażowanemu w dowód a nawet w wysłowienie twierdzenia Churcha: „rozstrzygalności". Otóż z definicji „teoria jest rozstrzygalna, jeśli istnieje metoda pozwalająca o każdym wyrażeniu tej teorii rozstrzygnąć za pomocą skończonej liczby prób czy jest ono twierdzeniem danej teorii. Tego typu metody nazywa się efektywnymi"[9].

Po raz kolejny okazuje się, że twierdzenie o nierozstrzygalności teorii wydaje się zrelatywizowane do techniki — tym razem sposobu rozstrzygania, czy coś przynależy do danego zbioru (aksjomatów lub twierdzeń teorii). Podobnie jak metody dowodzenia, tak

[7] Por. twierdzenie Löwenheima — Skolema — Tarskiego.

[8] Algorytm rozstrzygania, czy dana formuła F jest tezą systemu, polegałby na przeprowadzaniu w ustalonej kolejności dowodów, tak długo, aż natrafimy na dowód F albo $\neg F$. Ponieważ założyliśmy, że system jest zupełny, więc na pewno istnieje dowód formuły F lub jej zaprzeczenia.

[9] *Mała encyklopedia logiki*, dz. cyt.

i metody efektywne doczekały się precyzyjnego opracowania teoretycznego. Zaproponowano kilka formalizmów, takich jak funkcje rekurencyjne, maszyny Turinga, reprezentowalność w teorii, skończona definiowalność, itd., które okazały się równoważne. Wyróżniają one tę samą klasę funkcji (np. każda funkcja obliczalna za pomocą maszyn Turinga jest rekurencyjna i na odwrót) i za ich pomocą rozstrzygalne są te same problemy. To, co jest nierozstrzygalne za pomocą jednej spośród tych metod, nierozstrzygalne jest też za pomocą jakiejkolwiek innej.

W odróżnieniu jednak od powszechnie uznawanej teorii dowodu, teza, że wszystkie metody efektywne sprowadzają się do jednej spośród metod ujętych w któryś z tych formalizmów, jest wciąż przedmiotem dyskusji. Dopiero zaś ta teza (zwana tezą Churcha) pozwala na stwierdzenie **absolutnej**[10] nierozstrzygalności danej teorii (np. rachunku predykatów).

Warto podkreślić również, że po raz kolejny rzecz wiąże się z zagadnieniem nieskończoności. W każdym ze wspomnianych wyżej formalizmów zawarte są pewne ograniczenia, wymagania skończoności pewnych parametrów. Dla przykładu przywołajmy jeden z formalizmów — maszynę Turinga.

Najprościej rzecz ujmując, maszyna Turinga to pewien abstrakcyjny byt złożony z podzielonej na pola nieskończonej taśmy i głowicy czytająco–piszącej, wyposażony w odpowiedni zbiór instrukcji. Maszyna może znajdować się w jednym ze skończonej listy stanów wewnętrznych, może zapisywać lub wymazywać w polach taśmy po jednym ze skończonej listy symboli taśmowych. Instruk-

[10] „Absolutna nierozstrzygalność" oznacza, że wynik ten nie jest zrelatywizowany do jakiejś wybranej metody rozstrzygania. Zilustrujmy to przykładem: Turing udowodnił, że rachunek predykatów jest nierozstrzygalny, gdyż w przeciwnym przypadku rozstrzygalny byłby tzw. problem stopu. Wcześniej zaś pokazał, że problemu stopu nie da się rozwiązać za pomocą maszyn Turinga. Żeby wysnuć stąd wniosek, że problem stopu (a co za tym idzie problem bycia tezą rachunku predykatów) nie jest rozstrzygalny za pomocą żadnej efektywnej metody, trzeba przyjąć tezę Churcha; por. G. Boolos, R. Jeffrey, *Computability and Logic*, Cambridge University Press, 1974, s. 49 lub Odifreddi, dz. cyt., s. 103.

cje, zgodnie z którymi postępuje, mogą mieć jedną z następujących trzech postaci:

- $q_a\ S_b\ S_c\ q_d$: maszyna w stanie q_a, czytająca symbol S_b wymazuje go i wpisuje w jego miejsce symbol S_c, po czym zmienia stan na q_d,
- $q_a\ S_b\ R\ q_d$: maszyna w stanie q_a, czytająca symbol S_b przesuwa się o jedno pole w prawo, po czym zmienia stan na q_d,
- $q_a\ S_b\ L\ q_d$: maszyna w stanie q_a, czytająca symbol S_b przesuwa się o jedno pole w lewo, po czym zmienia stan na q_d,

Zbiór instrukcji powinien być skończony i niesprzeczny. Niesprzeczność oznacza, że nie ma takich instrukcji, które miałyby takie same przesłanki $q_a\ S_b$, a żądałyby wykonania różnych działań.

Jak widać, można wyliczyć wiele wymagań co do skończoności różnych parametrów maszyny Turinga: skończona liczba stanów wewnętrznych, symboli taśmowych, instrukcji, głowic, taśm i obserwowanych w jednym kroku pól. Wśród nich szczególnie jedno przywołuje skojarzenie z filozofią Leibniza — założenie o skończonej liczbie stanów wewnętrznych.

Uzasadniając adekwatność swojej teorii jako formalnego ujęcia obliczalności, Turing analizuje działanie rachmistrza przeprowadzającego obliczenia. Uzasadnia między innymi, że sensowne jest założenie, iż głowica maszyny „obserwuje” jedynie skończoną liczbę pól taśmy [11], bo także człowiek jest w stanie objąć wzrokiem zaledwie skończony zapis. Turing przekonuje, że możemy posługiwać się jedynie skończoną liczbą symboli taśmowych, bo gdyby było ich więcej, to nie umielibyśmy ich w końcu rozróżniać[12].

[11]Łatwo pokazać, że maszyna obserwująca jedynie jedno pole taśmy jest równoważna takiej, która obserwuje tych pól dowolną skończoną liczbę.

[12]Por. A. Turing, „On Computable Numbers, with an Application to the Entscheidungsproblem”, *Proceedings of the London Mathematical Society*, ser. 2, vol. 42 (1936–7), s. 249.

Wreszcie twierdzi, że na podobnej zasadzie niemożliwa jest nieskończona liczba dopuszczalnych stanów wewnętrznych — odpowiednika stanów umysłu rachmistrza — bo wówczas, podobnie jak w przypadku symboli taśmowych, różniłyby się one nieskończenie mało od siebie, a więc byłyby praktycznie nierozróżnialne. Sądzę, że to ostatnie wnioskowanie jest nadużyciem. Cóż bowiem szkodzi nasza niewiedza, nasza nieumiejętność rozróżniania stanów umysłu ich faktycznie infinityzemalnie małym różnicom?[13] Jako przykład metafizyki, w której twierdzenie o swego rodzaju „dyskretyzacji" stanów umysłu jest nieprawdziwe, można podać chociażby sławną monadologię — zgodnie z nią wszak w monadzie odbija się Wszechświat, który jest nieskończony[14] i w którym zmiany przebiegają w sposób ciągły.

Czy możliwe jest stworzenie rachunku (*calculus ratiocinator*?), który „radziłby sobie" z opisem relacji w takim świecie? Nie wiadomo. Jeśli teza Churcha jest prawdziwa, nie jest to możliwe. O to, czy jest prawdziwa, wciąż toczą się spory.

Jak sądzę, z powyższych rozważań można wysnuć następujące wnioski:

- Samo twierdzenie Gödla nie obaliło programu Leibniza, dzięki niemu jedynie wskazane zostały ograniczenia systemów formalnych spełniających ściśle określone warunki. (Nie widać dlaczego *lingua characteristica* nie miałaby istnieć).

- Choć możliwe jest budowanie systemów opisujących rzeczywistość lepiej niż rachunek predykatów pierwszego rzędu i języki elementarne, to na razie nie dysponujemy odpowiednim dla nich narzędziem wnioskowania. Ani wypracowana

[13]Trzeba przyznać, że Turing rozważa także sposób wyeliminowania słabo określonych „stanów umysłu" z analizy obliczalności, ale moim zdaniem nie udaje mu się to do końca.

[14]Por. założenie o skończonej liczbie obserwowanych pól — twierdzenie to można przyjąć w odniesieniu do potocznego rozumienia, świadomej „obserwacji", ale nie do „działania" i „doznawania" na poziomie monad.

w szkole Hilberta metoda dowodu, ani inne „obliczalne środki” nie są wystarczające.

- Jeśli teza Churcha–Turinga jest prawdziwa, to nie istnieje efektywna metoda pozwalająca na rozstrzyganie wszystkich problemów[15].

- Program Leibniza poszukiwania *calculi ratiocinatoris* pozostaje wciąż aktualny.

- Wydaje się, że jego realizacja wymagałaby wynalezienia skutecznego sposobu „okiełznania nieskończoności”, co zresztą genialny filozof przewidział.

Leibniz przewidywał również, że odkrycie tego rachunku jest w zasięgu ludzkich możliwości. Czy miał rację? Gödel wierzył, że (przynajmniej na terenie matematyki) tak.

[15]Teza Churcha–Turinga nie została dotąd udowodniona i nie wydaje się, by mogła zostać udowodniona za pomocą znanych obecnie technik. Zagadnienie to, szeroko dyskutowane, świadomie pominęłam; najczęściej wskazuje się, że w tezie Churcha porównuje się ściśle określone pojęcie funkcji obliczalnej za pomocą maszyn Turinga (lub funkcji rekurencyjnej, λ–definiowalnej, itd. — znaleziono wiele równoważnych, jak się później okazało, rachunków, co samo w sobie wydaje się przemawiać na korzyść tezy Churcha) z **intuicyjnym** pojęciem funkcji efektywnie obliczalnej. To ostatnie zaś niemal z założenia nie może zostać opisane formalnie (taką formalizację proponuje się właśnie w tezie Churcha, rzecz polega na tym, czy propozycja ta jest uzasadniona), więc i formalny dowód nie jest możliwy.