Semina □ Nr 3
Scientiarum 2004

Anna Tomaszewska

Twierdzenie Gödla a filozofia umysłu. W sprawie pewnej dwuznaczności argumentu Lucasa

1.

Co mają wspólnego artykuł Gödla z 1931 roku o zdaniach formalnie nierozstrzygalnych w systemie *Principia Mathematica* i filozofia umysłu? Z punktu widzenia nie uprawiającego filozofii matematyka czy logika być może niewiele, natomiast z punktu widzenia filozofa umysłu czy epistemologa – wprost przeciwnie: filozofowie uważają, że korzystając przede wszystkim z osiągnięć z zakresu metamatematyki (choć w filozoficznych rozważaniach uwzględnia się przeważnie w równym stopniu odkrycia z zakresu nauk przyrodniczych), mogą dowodzić twierdzeń o charakterze wyłącznie filozoficznym, a nawet metafizycznym. Niekiedy żywią przekonanie, że rozstrzygnięcia dotyczące zagadnień matematycznych czy metamatematycznych, wykorzystane w filozoficznych sporach, umożliwią ostateczne rozwiązanie niejako z definicji nierozwiązywalnych problemów filozoficznych albo – co wydaje się niezwykle istotne – „oczyszczenie" filozofii z ogromnej liczby błędnych mniemań, „metafizycznych iluzji", „nonsensów" itp. Przykładowo, Kazimierz Ajdukiewicz, korzystając z twierdzenia Gödla o niezupełności arytmetyki, dowodził fałszywości idealizmu transcendentalnego w ujęciu Rickerta[1]. John R. Lucas, w oparciu o to

[1]Por. K. Ajdukiwewicz, *Problemat transcendentalnego idealizmu w sformułowaniu semantycznym* [w:] *Język i poznanie*, Warszawa 1985, t. I.

samo twierdzenie, usiłował wykazać fałszywość mechanicyzmu[2], tj. poglądu, zgodnie z którym można skonstruować model umysłu ludzkiego nie różniący się istotnie od modelu skonstruowanego dla maszyny. W niniejszej pracy zajmę się argumentem zaproponowanym przez Lucasa.

Zanim jednak zostanie przedstawiona struktura tego argumentu, należałoby przybliżyć treść pierwszego twierdzenia Gödla. Artykuł Gödla[3] zaczyna się nawiązaniem do programu formalizacji matematyki Hilberta. Jeden z postulatów tego programu wyrazić można w najprostszy sposób tak: zdanie, które jest prawdziwe w danym systemie formalnym, daje się w nim w skończonej liczbie kroków dowieść przy użyciu środków, jakich ten system dostarcza (czyli zestawu znaków i procedur konstruowania z nich wyrażeń, aksjomatów systemu i obowiązujących w nim reguł inferencji) i na odwrót: zdanie, które daje się w danym systemie formalnym udowodnić, jest w nim prawdziwe, czyli jest tezą tego systemu. Tym samym po udowodnieniu powyższych zależności ustalona zostałaby koekstensjonalność odnoszących się do zdań predykatów „prawdziwy" i „mający dowód", przy czym tak rozumiana prawdziwość przestaje być pojęciem semantycznym, lecz zostaje zredukowana do własności syntaktycznej. Procedurę rozstrzygania, czy dane zdanie jest, czy też nie jest tezą danego systemu formalnego, można nazwać mechaniczną. Sposób postępowania, mający na celu przedłożenie dowodu tego zdania, zostaje bowiem ściśle wyznaczony przez określenie, jakiego rodzaju operacje można wykonywać na pewnych wyrażeniach, przy czym nie bierze się pod uwagę ich znaczenia, a tylko poprawność syntaktyczną (czyli uwzględnia się fakt, czy są poprawnie zbudowane).

---

[2]Por. J. R. Lucas, *Minds, Machines and Gödel*, Philosophy 36, 1961; artykuł można znaleźć także na stronie internetowej: `<http://users.ox.ac.uk/~jrlucas/Godel/mmg.html>`.

[3]Opieram się na angielskiej wersji tego artykułu: *On Formally Undecidable Propositions of 'Principia Mathematica' and Related Systems I* [w:] *Goedel's Theorem in Focus*, ed. G. S. Shanker, London 1988.

Ideę pierwszego twierdzenia Gödla można wyrazić tak: dla każdego niesprzecznego systemu formalnego SF, zawierającego elementarną arytmetykę liczb naturalnych z operacją mnożenia i dodawania, istnieje zdanie G, wyrażone w języku systemu formalnego J_{SF}, które jest dla tego systemu prawdziwe, ale nie daje się w nim dowieść; jest więc tak, że ani zdanie G, ani jego negacja nie należy do zbioru konsekwencji tez SF. Zakładamy na początku, ze zdanie prawdziwe to takie, dla którego istnieje dowód w SF. Niech zdanie G[4] brzmi: „To zdanie nie ma dowodu w SF". Jeśli jest ono fałszywe, to znaczy, że jego negacja – przy metalogicznym założeniu dwuwartościowości logiki – jest prawdziwa, a głosi ona, że zdanie to ma dowód w SF, a skoro tak, to należy do zbioru tez SF – bo, zgodnie z naszym założeniem, każde zdanie, dające się dowieść w SF, jest zdaniem prawdziwym, co w rezultacie przeczy założeniu o jego fałszywości. Jeśli natomiast jest prawdziwe, to na mocy przyjętego wcześniej założenia, ma dowód w SF, a więc zdanie: „To zdanie nie ma dowodu w SF" jest fałszywe, co przeczy założeniu o jego prawdziwości. Aby rozwiązać tę antynomię, przyjmujemy, że jest tak, jak to zdanie mówi: zdanie G jest prawdziwe, czyli – jak „samo" stwierdza – nie ma ono dowodu. Z tego oraz z założenia o niesprzeczności naszego systemu formalnego wynika jego niezupełność, czyli nie jest tak, że dla dowolnego zdania albo ono samo, albo jego negacja jest tezą SF.

[4] Należące do języka J_{SF}.

2.

Zdaniem J. R. Lucasa, filozoficzną konsekwencją[5] rozumowania Gödla jest teza, iż umysł ludzki istotnie różni się od maszyny, rozumianej tutaj jako maszyna Turinga czy też równoważność dowolnego systemu formalnego, zawierającego porcję arytmetyki. Różnica, o jaką tu chodzi, polega na tym, iż umysł, co najmniej w zakresie matematyki, potrafi wykonywać operacje, których nie potrafi wykonywać maszyna. Jeśli maszyna (system formalny) potrafi przedłożyć n zdań prawdziwych, my – ludzie – potrafimy zawsze podać ich $n + 1$. Na naszej liście zdań prawdziwych znajduje się bowiem zdanie gödlowskie, dla którego dowód formalny, możliwy do przedstawienia przez maszynę, nie może istnieć.

Można zadać sobie pytanie: co to znaczy, że pojęcie prawdziwości okazuje się, w rezultacie, nie być pojęciem jednoznacznym

---

[5]Kiedy mówi się o filozoficznych konsekwencjach danego twierdzenia matematycznego albo teorii fizycznej, z pewnością nie można mieć na myśli tego rodzaju konsekwencji, o jakiej mówi się w logice. (Zob. w tej sprawie: J. Woleński, *Metamatematyka a epistemologia*, Wyd. Naukowe PWN, Warszawa 1993, s. 10nn). W tym ostatnim przypadku jest tak, że jeśli z jakiegoś zbioru zdań wynika pewne zdanie, to pod warunkiem, że ów zbiór nie jest sprzeczny, nie może zeń wynikać także negacja owego zdania. W przypadku tzw. konsekwencji filozoficznych nie jest to takie oczywiste: teorie fizyczne czy nawet twierdzenia matematyki mogą być rozmaicie interpretowane – to m.in. od interpretacji zależy wyciągnięcie z nich takich czy innych konsekwencji. Przez interpretację rozumiem funkcję przeprowadzającą według z góry określonych reguł pewien zbiór zdań w inny, nie zawierający się w nim, zbiór zdań.
Operator konsekwencji logicznej ma między innymi taką własność: dla pewnego zdania p i zbioru zdań A, jeśli p należy do A, to p należy także do zbioru konsekwencji logicznych A. W szczególnym przypadku, gdy A jest zbiorem pustym, zbiór konsekwencji tego zbioru będzie się składał z praw logiki. Nie jest natomiast prawdą, jakoby zbiór konsekwencji filozoficznych danego zbioru zdań zawierał się w tym zbiorze. Przykładowo, z twierdzenia Gödla o niedowodliwości niesprzeczności elementarnej arytmetyki liczb naturalnych w niej samej logicznie nie wynika zdanie mówiące, dajmy na to, że poznanie prawdy absolutnej jest niemożliwe. Byłoby raczej tak, że zbiór konsekwencji filozoficznych stanowiłby podzbiór zbioru zdań będących interpretacjami interesujących nas w danym przypadku twierdzeń czy teorii.

z pojęciem dowodliwości, co to znaczy, że prawdziwość jakiegoś zdania nie sprowadza się do wszystkich „mechanicznych” procedur, jakie trzeba wykonać, by prawdziwość tę wykazać? Nieco tajemniczą mogłaby się wydać odpowiedź, że umysłowi ludzkiemu dane jest coś w rodzaju „wglądu w naturę rzeczywistości” czy „wglądu w naturę prawdy”. Umożliwiałby on uznawanie za prawdziwe zdania, którego prawdziwości nie można w żaden sposób dowieść, a w każdym razie nie przy pomocy narzędzi systemu formalnego (ale wtedy powstaje pytanie o innego rodzaju „narzędzia” służące dowodzeniu naszych twierdzeń *nieempirycznych*[6]). Prawdziwość jest niejako dostępna w bezpośrednim intelektualnym ujęciu, widoczna „gołym okiem”, jej „samomanifestowanie się” oznacza dokładnie to samo, co zasadność zdania charakteryzującego się ową prawdziwością. Żywienie takiego przeświadczenia mogłoby się okazać równoznaczne ze zwolnieniem osoby, głoszącej prawdziwość jakiegoś twierdzenia, z odpowiedzialności za podanie dlań stosownego, akceptowalnego uzasadnienia. W ostateczności uzasadnienie dowolnego stwierdzenia typu „Wiem, że A”, gdzie A ma charakter zdania nie mającego empirycznych podstaw uzasadnienia, a więc charakter czegoś w rodzaju zdania *a priori*, przedstawiałoby się nadzwyczaj prosto: „Wiem, ponieważ wiem”.

Jak należałoby rozumieć pojęcie prawdziwości w kontekście argumentacji przedstawionej przez Lucasa? Niestety, sprawa ta nie została dokładnie przez niego wyjaśniona. Co to bowiem znaczy, że zdanie gödlowskie jest prawdziwe? Nie podejmuję się jednoznacznie odpowiedzieć na to pytanie, pozwolę sobie jednak zacytować stosowny fragment artykułu Lucasa. Otóż pisze on tak:

Konstruujemy teraz formułę gödlowską w tym [rozważanym przez Lucasa] *systemie formalnym. Formuła ta nie może być* ***dowiedziona–w–systemie*** [podkr. Lucasa]. *Dlatego maszyna nie może wyprodukować odpowiadającej jej formuły jako prawdziwej. My jednak* ***widzimy, że formuła gödlowska jest prawdziwa***

[6]Problem nie dotyczy wszak dowodzenia twierdzeń w rodzaju „Słońce dzisiaj wzeszło”, lecz takich, które mają charakter aprioryczny.

[podkr. A.T.]: *każda* ***racjonalna istota*** [podkr. A.T.] *mogłaby prześledzić argument Gödla i* ***przekonać się*** [podkr. A.T.], *że formuła gödlowska, choć niedowodliwa–w–systemie, jest jednakże* [...] *prawdziwa.* [...] *dla każdej maszyny istnieje prawda, której nie może ona* [*scil.* maszyna] *wyprodukować jako prawdziwej, ale którą umysł może*[jako prawdziwą przedłożyć]. *To pokazuje, że maszyna nie może być zupełnym i adekwatnym modelem umysłu*[7].

W ostateczności niezbyt dobrze wiadomo, jak należałoby interpretować fakt rozumienia przez nas prawdziwości formuły gödlowskiej. Lucas zdaje się jednak sugerować co następuje: każda racjonalna istota, a więc istota wyposażona w ludzki umysł[8], jest w stanie — jak zakładamy, abstrahując od wszelkiego rodzaju ograniczeń z intelektualnymi włącznie —zrozumieć argumentację Gödla. Jeśli ją zrozumie, da się przekonać co do jej słuszności. A zatem zgodzi się z twierdzeniem, iż dla każdego systemu formalnego istnieje pewne zdanie prawdziwe, którego w systemie tym nie można dowieść. W następstwie tego zaś uzna antymechanistyczną tezę Lucasa, stanowiącą filozoficzną konsekwencję twierdzenia Gödla o niezupełności arytmetyki. Zwróćmy uwagę na pewną szczególną własność prawdziwości, wychodzącą tu na jaw jakby niezamierzenie: otóż prawdziwość jest tym, co ma *moc przekonywania* albo raczej tym, co stwierdza się w odniesieniu do jakiegoś zdania wtedy, gdy zostało się już w dostatecznym stopniu (czyli najczęściej „do końca") przekonanym. Aby zostać przekonanym, trzeba się najpierw poddać działaniu tego, kto przekonuje; mamy zatem sytuację – nazwijmy ją „sytuacją dialektyczną"[9] – w któ-

---

[7]John R. Lucas, *Minds, Machines and Göedel* [w:] *Philosophy* XXXVI, 1961, ss. 112–127. Także na stronie: `<http://users.ox.ac.uk/~jrlucas/Godel/mmg.html>`, tłum. A.T.

[8]Nie będziemy się tutaj wdawać w analizę natury ontologicznej umysłu; prawdopodobnie Lucas nie zgodziłby się z fizykalistycznym redukcjonizmem uznającym tezę o tożsamości umysłu i mózgu, gdyż uznanie tej tezy może łatwo prowadzić do uznania mechanicyzmu.

[9]Lucas mówi o „dialektycznej grze", której protagonistami będą Umysł i Maszyna – „obrazek" odsyłający do artykułu Turinga *Maszyna licząca a in-*

rej dwa umysły wchodzą we wzajemną relację, porozumiewają się ze sobą dla osiągnięcia pewnego *celu*. Być może nie całkiem od rzeczy będzie powiedzieć, że zrozumienie, na czym polega prawdziwość jakiegoś zdania, wymaga, aby przyjrzeć się temu, jak do prawdziwości owego zdania *w praktyce* się dochodzi. Jest to dla filozofa niezmiernie interesująca kwestia, wymaga jednak zajęcia stanowiska maksymalistycznego, niemiłego orędownikom programu Hilberta, pozytywistom, zwolennikom, niewykonalnej zresztą, redukcji semantyki do syntaksy itp. Prawdziwość bowiem w ludzkiej praktyce życiowej[10] jedynie w luźny sposób wiąże się z dowodliwością; aby się o tym przekonać, wystarczy sporządzić listę twierdzeń przyjmowanych niejako „na wiarę”, czyli bez dowodu, których odrzucenie jednakże mogłoby poważnie zachwiać naszą racjonalnością, czyniąc obraz naszego doświadczenia świata wewnętrznie niespójnym czy po prostu sprzecznym. Jednym z takich nie mających dowodu twierdzeń będzie to, które mówi, że jesteśmy istotami niesprzecznymi.

3.

Przedstawiony w poprzedniej części argument Lucasa[11], wykazujący fałszywość mechanistycznych koncepcji umysłu[12], spotkał się z licznymi zarzutami, z których najważniejszy brzmi: Lucas jest sprzeczny. Dlatego też, bez względu na to, kto ma rację: mechanicyści czy antymechanicyści, z matematycznego punktu widzenia, teza Lucasa okazuje się fałszywa. Odwołując się do twierdzenia Gödla, nie możemy zatem pokazać, że umysł istotnie różni się

teligencja, [w:] *Filozofia umysłu. Fragmenty filozofii analitycznej*, Aletheia–Spacja, Warszawa 1995.

[10] Jeśli chcemy być bardziej precyzyjni, możemy mówić o praktyce językowej.

[11] Jak widać, rekonstrukcja argumentu Lucasa nie wypadła imponująco, ale też nie widzę powodów, dla których miałoby być inaczej. Rozumowanie przeprowadzone przez Lucasa jest bowiem bardzo proste, ale jak wkrótce zobaczymy, nie znaczy to jeszcze, iż jest ono zarazem matematycznie poprawne.

[12] Sformułowanie to brzmi nieco zbyt „dziewiętnastowiecznie”, dlatego można by zastąpić je wyrażeniem „komputacjonistyczne modele umysłu”.

od „inteligentnej" maszyny. Przed omówieniem jednak tego najistotniejszego zarzutu warto przyjrzeć się kilku „mniej groźnym" zarzutom, z którymi rozprawia się autor *Minds, Machines and Gödel*.

Zarzut 1. Tworzenie zdania gödlowskiego jest procedurą standardową, można wręcz powiedzieć: mechaniczną, za każdym razem bowiem stosujemy taką samą sztuczkę – przedkładamy prawdziwe zdanie gödlowskie, którego maszyna (system formalny) nie ma na liście swoich tez. Jeśli przedłożone przez nas zdanie zostanie wciągnięte przez maszynę na tę listę (poprzez dołączenie do zbioru jej aksjomatów), to możemy przedłożyć kolejne zdanie, którego maszyna nie ma na liście — i tak w nieskończoność[13]. Skoro zatem „wygödlowywanie" (*out-gödeling* — słowo, którego używa Lucas) jest procedurą standardową, powinno dać się sformalizować — można zaprogramować maszynę tak, by „produkowała" bez żadnych ograniczeń (ograniczeń technicznych ani czasowych nie bierzemy tu pod uwagę) zdania nierozstrzygalne; maszyna posiadałaby wtedy tzw. *Gödelizing operator*.

Lucas odpowiada na ten zarzut, że umysł *zawsze* jest w stanie „wygödlować" maszynę, przedkładając takie zdanie, którego ta nie ma na liście. Ta w istocie niezbyt interesująca procedura przypomina grę, polegającą na wymienianiu coraz większych liczb naturalnych —wygra ten, kto poda taką liczbę, że jego przeciwnik nie będzie potrafił podać większej. Wiadomo, że w takiej grze, która gdyby nie przeszkody natury technicznej, nigdy by się nie skończyła, nie ma wygranej. Według Lucasa, jeśli maszyna jest w stanie podać ω zdań gödlowskich, to umysł zawsze może podać tych zdań $\omega + 1$ i wygra. Skoro wiemy, że arytmetyka jest niezupełna i zakładamy też, że arytmetyka jest „wbudowana" w program maszyny, to zgodnie z twierdzeniem Gödla *musi* dać się zna-

[13]Podobnie w nieskończoność można rozbudowywać hierarchię języków: dla *każdego* języka, prócz języka naturalnego, zawierającego swój własny metajęzyk, istnieje co najmniej jeden metajęzyk, będący też językiem, tyle że wyższego stopnia.

leźć zdanie, którego maszyna nie przedłoży. Można postawić Lucasowi pytanie, czy *nasza* umiejętność „wygödlowania" maszyny nie oznacza, że jesteśmy „zupełni"[14], a więc potrafimy efektywnie rozstrzygnąć wartość logiczną każdego zdania[15]? I na czym by to „efektywne rozstrzygnięcie" miało polegać – wszak przecież nie na podaniu formalnego dowodu nierozstrzygalnego dla maszyny zdania – wtedy bowiem z powrotem „lądujemy" w systemie formalnym. Mówi się w takich wypadkach o jakimś „intuicyjnym ujęciu" prawdy czy, jak już wspominałam, o jakimś szczególnego rodzaju „wglądzie", umożliwiającym identyfikację konkretnej prawdy nawet wtedy, gdy prawda ta nie ma dowodu; nie zawsze jest jednak jasne, co się rozumie przez samo pojęcie intuicji czy wglądu. Jak już sygnalizowałam, problem sprowadza się do uznawania bądź nieuznawania procedur innych niż dowodowe za wystarczający warunek uzasadniania naszych twierdzeń — i do tego, czy w ogóle takie „procedury" istnieją. Problem ten staje się jeszcze bardziej ogólny w momencie, gdy rozszerzymy, „odformalizujemy" nasze rozumienie pojęcia dowodu — będziemy wtedy mówić, że jakieś zdanie ma dowód, jeżeli istnieją powszechnie uznane, racjonalne procedury demonstracji prawdziwości tego zdania; innymi słowy, jeśli można *sprawdzić*, czy dane zdanie jest prawdziwe czy nie. Starałam się także wcześniej zasugerować, że takie rozumienie prawdziwości zdań byłoby zbyt restrykcyjne: w praktyce uznajemy za prawdziwe wiele twierdzeń, których prawdziwości z różnych przyczyn nie potrafimy udowodnić, co nie oznacza bezzasadności ich przyjmowania. Żyjemy w konkretnych społecznościach, gdzie zachowanie, sposób postępowania itd. jednostek są wyznaczane przez

[14] A zarazem niesprzeczni, w przeciwnym bowiem wypadku całe to rozumowanie, którego konsekwencją jest odróżnienie umysłów od maszyn, po prostu upada. Jeśli umysł jest sprzeczny, a maszyna – nie, to na liście zdań, które umysł przedkłada maszynie, znajdzie się wiele takich, które nie znajdą się na liście przedkładanej przez maszynę, jak choćby negacje tych ostatnich.

[15] Chodzi oczywiście o takie zdania, które są dobrymi kandydatami na posiadaczy wartości logicznej, których szczególny przypadek stanowią zdania matematyczne.

pewne reguły racjonalności, każące pewne twierdzenia akceptować, pewne zaś odrzucać. Prawdziwość, jaką charakteryzują się tego rodzaju niedowodliwe, zdania, nie będące jednak aksjomatami, nazwałabym *prawdziwością normatywną* (albo *transcendentalną*) – *warunkującą* prawdziwość twierdzeń, które się na ich podstawie wygłasza oraz wewnętrzną spójność systemu, do którego się one stosują.

Zarzut 2[16]. Twierdzenie Gödla zachowuje ważność jedynie dla systemów dedukcyjnych, ale ludzie stosują także inne rodzaje rozumowania, np. indukcyjne, probabilistyczne.

Zarzut ten mówi tyle: nie trzeba twierdzenia Gödla, aby pokazać, że nie jesteśmy maszynami. W pewnym sensie bowiem oczywiście nie jesteśmy nimi, a tylko być może ta część naszego myślenia, którą stanowi matematyka, daje się „zmechanicyzować". Lucasowi jednakże chodzi o to właśnie, by pokazać, że nasze *myślenie matematyczne* jest niemechaniczne, a w konsekwencji niezdeterminowane ściśle określonymi, wewnątrzsystemowymi regułami. Po prostu: nawet wtedy, gdy rozumujemy dedukcyjnie, rezultat naszego rozumowania nie musi być z góry określony; dlatego matematyka jest tak twórczą dziedziną ludzkiej działalności poznawczej.

Zarzut 3 (Hilary Putnam[17]). Nie da się wykluczyć możliwości, że jesteśmy sprzecznymi maszynami.

Lucas uważa to za mało prawdopodobne. Bywamy wprawdzie niekiedy sprzeczni, na przykład wtedy, gdy nasz rozum daje się opętać niezdrowym emocjom[18]. Przyznajmy Lucasowi rację: rzekomą sprzeczność istot ludzkich możemy złożyć na karb ich omylności i ignorancji. Wartość naszych twierdzeń – to, czy są one prawdziwe czy nie – zależy przecież od stanu naszej wiedzy; korzystając z różnych założeń, dysponując na wyjściu raz

[16]Na liście zarzutów rozpatrywanych przez Lucasa był to zarzut trzeci.

[17]O zarzucie tym wspomina Lucas w cytowanym wyżej artykule *Minds, Machines and Goedel.*

[18]Prawdę mówiąc, jeśli określenie „niezdrowy" uznać w tym przypadku za determinujące, nie zaś modyfikujące, to otrzymamy pleonazm, prosiłabym jednak Czytelnika o wybaczenie tej językowej usterki.

takim, a innym razem zgoła odmiennym zasobem informacji, dochodzimy niekiedy do wzajemnie się wykluczających wniosków. Stąd konkluzja, że absolutnie czy też idealnie niesprzeczny podmiot musiałby zarazem być podmiotem wszystkowiedzącym. Ale z drugiego twierdzenia Gödla, mówiącego o niedowodliwości niesprzeczności arytmetyki w niej samej, oraz z rozważań Ajdukiewicza przedstawionych w artykule *Problemat transcendentalnego idealizmu w sformułowaniu semantycznym*[19] wynika, że idea wszystkowiedzącego podmiotu jest wewnętrznie sprzeczna, oczywiście jeśli zgodzić się z postulatem, że „wiedzieć" znaczy tyle, co „dysponować dowodem" dla tego, co się wie. Podmiot ów nie może wtedy wiedzieć wszystkiego – nie wie bowiem tego, czy jest niesprzeczny.

Zauważmy, że założenie niesprzeczności umysłu ma charakter normy, którą nasza racjonalność sama sobie nakłada; norm zaś nie dowodzi się, a tylko się je postuluje lub wyznacza. Można jednakże pójść nieco dalej i postawić pytanie o prawomocność tej normy. Abstrahując od gödlowsko–lucasowskiego kontekstu naszych rozważań, możemy zapytać, co w ogóle czyni zasadę niesprzeczności normą regulującą naszej działalności teoretycznej, a także (w dużym stopniu przynajmniej) praktycznej. Niesprzeczność, jako warunek wszelkiej rozumności, umożliwia racjonalny dyskurs, a więc rozumienie znaczenia zdań, za pomocą których ludzie się komunikują – szacowanie ich prawdziwości względnie fałszywości. Jeśli naraz przyznaję jednemu i temu samemu zdaniu dwie wykluczające się wartości logiczne, zwyczajnie nie wiem, o czym mówię, czyli nie rozumiem znaczenia wypowiadanego przez siebie zdania.

Wracam teraz do wspomnianego wyżej zarzutu Putnama. Otóż jeśli niemechaniczność umysłu należy wykazać przy założeniu jego niesprzeczności, to musimy znać odpowiedź na pytanie, czy jesteśmy sprzeczni, czy nie; uczciwość intelektualna wymaga, abyśmy posiadali dowód naszej niesprzeczności (przyjmujemy tu, że wiedza, iż jakieś zdanie jest prawdziwe, oznacza posiadanie dowodu

[19] K. Ajdukiewicz, *Problemat...* [w:] *Język i poznanie*, Warszawa 1985, t. 1.

na tę jego prawdziwość). Wiemy jednak tylko tyle, że pod warunkiem, iż jesteśmy niesprzeczni, maszyna przegrywa. Ale założenie, że jesteśmy niesprzeczni, przyjmuje się bez uzasadnienia. Z tego, że *wiem, że [jeśli A, to B]* wynika, że *jeśli wiem, że A, to wiem, że B* (rozdzielność operatora epistemicznego względem implikacji); nie wynika natomiast, że *jeśli A, to wiem, że B*. Antymechanicysta zdaje się być w takiej sytuacji: *wie, że [jeśli jest niesprzeczny, to wygrywa]* i *wie, że wygrywa* (ale jeśli wygrywa, to nie znaczy jeszcze, że jest niesprzeczny; może być sprzeczny, również tego nie wiedząc; wtedy też musi wygrać). Nie wie jednak najważniejszego, mianowicie, czy jest niesprzeczny i nie ma nadziei na to, że nawet jeśli byłby niesprzeczny, to mógłby tego dowieść. Argument Lucasa nie może być zatem konkluzywny.

Zarzut 4 (oczywiście nie rozpatrywany przez Lucasa!). Lucas[20] jest sprzeczny[21].

Sprawa przedstawia się w następujący sposób: mechanicysta przedstawia Lucasowi maszynę M, o której twierdzi, że stanowi adekwatny model zdolności matematycznych (czy — szerzej — umysłu) Lucasa. Wtedy ten przedstawia mechanicyście formułę gödlowską G_M odpowiadającą teorii $T(M)$, w której znajdują się zdania arytmetyki, dające się udowodnić przez maszynę M. Mamy dwie możliwości: albo (1) $T(M)$ jest sprzeczna, albo (2) $T(M)$ jest niesprzeczna. Jeśli zachodzi (2), G_M nie daje się dowieść przez maszynę, ale Lucas widzi, że formuła ta jest prawdziwa. Jeśli zachodzi (1), maszyna ma dowód dla każdego zdania, w tym także G_M i takich zdań jak np. $0 + 1 = 0$, czego Lucas nie udowodni, jako że jest niesprzeczny. W obu przypadkach Lucas różni się od maszyny.

---

[20] A ściślej mówiąc, proponowany przez oksfordzkiego filozofa model umysłu, mający istotnie się różnić od modelu maszyny Turinga.

[21] Dowód sprzeczności Lucasa można znaleźć w: Stanisław Krajewski, *Twierdzenie Goedla i jego interpretacje filozoficzne*, Warszawa 2003, a także w: tenże, *Philosophical Consequences of Goedel's Theorem* [w:] *Bulletin of the Section of Logic*, 1983; 12.

Powyższe stwierdzenie dotyczy wszystkich maszyn. Możemy więc sporządzić listę M_1, M_2,... (indeksy oznaczają opisy czy też numery porządkowe maszyn na liście). Istnieje funkcja rekurencyjna g [22], taka że dla każdego n:
(*) jeśli $T(Mn)$ jest niesprzeczna, to $g(n)$ jest numerem gödlowskim zdania gödlowskiego, które jest prawdziwe i nie daje się dowieść w $T(M_n)$.

Wykażemy, że zbiór wartości funkcji g, powiedzmy zbiór $A = \{g(n) : n = 1, 2, \dots\}$, jest sprzeczny. Dochodzimy do tego następująco: $A = T(M_k)$, dla pewnego k. Jeśli $T(M_k)$ jest niesprzeczna, to na mocy (*) $g(k)$ nie daje się dowieść w $T(M_k)$, czyli nie należy do $T(M_k)$. Ale $g(k)$ należy do A, na mocy definicji, a więc też do $T(M_k)$. Otrzymaliśmy sprzeczność.

Ponieważ n oznacza dowolny numer maszyny, więc możemy przyjąć, że $n = k$. Zatem także $T(M_n) = T(M_k)$ i $T(M_n)$ jest sprzeczna. Stąd też funkcja g nie może być rekurencyjna (rekurencyjność tej funkcji stanowiła warunek niesprzeczności $T(M_n)$).

Pozostaje więc albo zgodzić się z uzyskanym rezultatem mówiącym, że Lucas jest sprzeczny, wobec czego należy odrzucić (*), albo uznać możliwość istnienia nierekurencyjnych sposobów dochodzenia do prawdy za jedną z filozoficznych konsekwencji twierdzenia Gödla i utrzymać (*). Pozostaje jednak problem, czy $T(M_n)$ jest sprzeczna czy nie.

4.

Konkluzje z powyższych rozważań wydają się dość zaskakujące: z jednej strony bowiem intuicyjnie, zdroworozsądkowo zgadzamy się z rozumowaniem Lucasa, bez względu na to, czy nasze ideologiczne sympatie lokują się po stronie antymechanicystów czy też ich przeciwników; z drugiej strony można pokazać, że rozumowanie to jest matematycznie niepoprawne. Oczywiście, brak mi

[22] Założenie to jest istotne dla przeprowadzenia przez S. Krajewskiego, w cytowanych w powyższym przypisie pracach, dowodu sprzeczności Lucasa.

kompetencji, aby dokonać „matematycznej apologii" Lucasa. Nie chcę też zarazem przesądzać o prawdziwości owych intuicji, nakazujących z jego rozumowaniem się zgodzić; niemniej faktu, że takie intuicje tkwią w nas dość głęboko, nie należy lekceważyć. I jeśli w jakiś sposób udało mi się zwrócić uwagę Czytelnika na tę rozbieżność – którą można z grubsza scharakteryzować jako rozbieżność między intuicyjnym, potocznym, zdroworozsądkowym a matematycznym rozumieniem i uznawaniem prawdziwości czy nawet poprawności twierdzeń i rozumowań – uważam, iż cel niniejszego artykułu został osiągnięty.