Czasopismo Filozoficzne 6 (2010), s.50-70.
ISSN 1896-5059

Czy twierdzenie Gödla jest sprzeczne z mechanicyzmem?

Mariusz Bodynek

Abstract: The article is an attempt at determining the possibility of the application of Godel's meta-logical statements regarding the philosophy of mind. The first systematical attempt of using these statements in the aforementioned scope was made by J. R. Lucas. He became famous by formulating a thesis that the statements about incompleteness and non-decidability are a basis for a definitive refutation of mechanisms. Lucas' attempt is a part of a practice often met in social and humanistic sciences which consists in proving theses and creating theories on the basis of Godel's statements. However, these attempts are generally unfortunate, arguments based on famous metalogical statements are irrelevant and defective. It results mainly from using Godel's statements in inadequate contexts to their specificity. Therefore, the conditions of the application of the aforementioned statements are first formulated in the article; and then, an analysis of the correctness of using Godel's statements by Lucas is conducted within the context of a dispute between supporters of mechanism and supporters of anti-mechanism. The next parts are dedicated to the examination of Lucas' argumentation and reasoning. In the course of the analysis the reasoning of the British philosopher, apparently simple and brilliant, proves to be burdened by a number of flaws and weaknesses.

Key words: Lucas's argument, Turing machine, human mind, limits of proof, idea of contradiction

I. Uwagi ogólne

Twierdzenia Gödla uchodzą za najdonioślejsze odkrycia XX wieku w dziedzinie logiki matematycznej. Z perspektywy upływu około 80 lat od sformułowania tezy o niezupełności i nierozstrzygalności, stawianie Kurta Gödla w szeregu najwybitniejszych uczonych i myślicieli wydaje się całkowicie uzasadnione. Wielu określa go mianem największego logika od czasów Leibniza czy nawet Arystotelesa[1]. O randze i przełomowości odkryć Gödla zaświadczyć może także autorytet innego wielkiego uczonego XX wieku – Alberta Einsteina. Ten, pod wpływem owych odkryć, miał zmienić swój stosunek do matematyki, którą wcześniej postrzegał jako dyscyplinę pomocniczą i niezdolną do generowania praw tak fundamentalnych jak fizyka[2]. Tymczasem

[1] Na przykład wybitny matematyk John von Neuman. Szerzej na ten temat: S. K r a j e w s k i: *Twierdzenie Gödla i jego interpretacje filozoficzne. Od mechanicyzmu do postmodernizmu*. Warszawa 2003, s. 173.
[2] Zob. tamże, s. 10.

Opublikowano online: 01.12.2010

K. Gödel miał mu uwidocznić, że badania łączące dziedzinę matematyki z logiką mogą służyć nie tylko usprawnianiu aparatu technicznego do wyrażania matematycznych oczywistości czy też umacniania lub budowania fundamentów matematyki, ale mogą też przyczynić się do sformułowania praw rzucających nowe światło na możliwości i ograniczenia ludzkiej wiedzy.

Zasięg naukowego oddziaływania tytułowych twierdzeń jest zdumiewająco rozległy. Choć są to twierdzenia metalogiczne odnoszące się do matematyki, to próby ich zastosowań i interpretacji były podejmowane w dziedzinach zarówno pokrewnych z matematyką czy logiką matematyczną (np. filozofia logiki i matematyki), spowinowaconych (np. fizyka), jak i bardzo odległych (nauki społeczne i humanistyczne). Wśród tych ostatnich szczególne zdziwienie może budzić obecność socjologii kultury, językoznawstwa, kulturoznawstwa (w tym teorii dekonstrukcji tekstów literackich), prawa czy religii[3]. Nie wszędzie próby te były udane, trafne czy po prostu właściwe. Nie wszędzie jednak mogły być udane, gdyż zakres stosowalności twierdzeń Gödla jest ograniczony. Dlatego warto odróżniać konteksty, w których ich stosowanie jest dopuszczalne, choć błędne (z czym możemy mieć do czynienia np. na polu filozofii matematyki czy logiki), od kontekstów, w których stosowanie jest niedopuszczalne, choć pozornie poprawne (np. na polu teorii prawa wysuwanie na podstawie twierdzenia Gödla wniosku o niezupełności systemu prawnego czy teorii socjologicznej o „niedomkniętości" systemu społecznego)[4]. Te ostatnie zaś należy odróżniać od powierzchownych analogii będących wynikiem inspiracji twierdzeniami Gödla, z czym możemy się spotkać między innymi na polu filozofii religii czy teologii (wzmocnienie obrony przed zarzutem o brak dowodu istnienia Boga, wskazując na niedowodliwość niesprzeczności matematyki).

Najwcześniej rozpoznaną i najistotniejszą konsekwencją sformułowania twierdzeń Gödla była konieczność uznania upadku programu Davida Hilberta[5]. Drugie twierdzenie mówiące, że nie można dowieść niesprzeczności niesprzecznej teorii w ramach tej teorii, zniweczyło wszelkie próby udowodnienia niesprzeczności matematyki środkami finitystycznymi. W tym przypadku nie ma wątpliwości co do właściwej interpretacji twierdzenia Gödla jako tezy obalającej przekonanie o możliwości ugruntowania matematyki w sposób zaproponowany przez D. Hilberta. Natomiast poważny problem przedstawia zastosowanie twierdzeń Gödla w ramach filozofii umysłu. Mówiąc

[3] Zob. tamże, s. 253-343.
[4] Kryterium stosowalności zostanie dokładniej sformułowane niżej.
[5] Niemniej jednak pojawiały się próby rewitalizacji programu; więcej na ten temat: R. M u r a w s k i: *Kontekst historyczny i recepcja twierdzenia Gödla o niezupełności*. W: *Filozofia i logika. W stronę Jana Woleńskiego*. Red. J. H a r t m a n. Kraków 2000, s. 414-420.

konkretniej, chodzi o użycie tych twierdzeń przeciwko mechanicyzmowi, przyrównującemu ludzki umysł do maszyny. Pierwszą systematyczną próbę wykorzystania odkryć Gödla do „rozprawienia się" z mechanicyzmem podjął John Randolph Lucas i to właśnie jego wywód zostanie poddany analizie w dalszej części artykułu.

Przedtem należy jednak wrócić do kwestii dopuszczalności zastosowania twierdzeń Gödla i pokazać, dlaczego stosowanie ich w jednych kontekstach jest w pełni uprawnione, choć nie zawsze pomyślne, a w innych z góry skazane na niepowodzenie. Gwoli większej precyzji, należałoby powiedzieć, że chodzi tu o rozstrzyganie kwestii samej możliwości rozważania twierdzeń Gödla w charakterze potencjalnego dowodu jakiejś tezy. Jest to zagadnienie o pierwszorzędnym znaczeniu.

II. Warunki stosowalności twierdzeń Gödla a argument Lucasa

Zostało zasygnalizowane wyżej, że twierdzenia Gödla zostały sformułowane przede wszystkim z myślą o matematyce. Pierwsze twierdzenie głosi, że w każdej niesprzecznej teorii zawierającej arytmetykę (co najmniej elementarną teorię liczb naturalnych) znajduje się para zdań sprzecznych, z których żadne nie ma dowodu. Wynika więc z tego, że istnieją takie formuły A i nie-A w teorii T, których nie można ani dowieść ani obalić w ramach danej teorii. Każda więc niesprzeczna teoria, co najmniej tak bogata jak arytmetyka, jest niezupełna, gdyż zawiera niedowodliwą formułę.

Do uzyskania tego typu własności teorii potrzebny jest skomplikowany aparat techniczny, którego znaczenie jest niebagatelne zarówno dla zrozumienia przedsięwzięcia Gödla, jak i właściwego interpretowania jego wyników w określonych kontekstach. Wydaje się jednak, że można pominąć jego opis formalny bez uszczerbku dla prezentacji istoty zagadnienia i przedłożyć ogólnie jego najważniejsze elementy.

W celu otrzymania wyników Gödla należy dokonać szeregu operacji na matematyce. Po pierwsze matematykę należy ująć jako system aksjomatyczny i poddać ją formalizacji. Chodzi w tym kontekście o formalizację języka arytmetyki, to znaczy o określenie jej alfabetu (skończonego ciągu symboli służących do budowy formuł: np. spójniki logiczne i kwantyfikatory, symbole arytmetyczne, symbole zmiennych itp.), reguł gramatycznych (np. reguł budowania sensownych wyrażeń) oraz podanie skończonego szeregu fundamentalnych tez (aksjomatów) charakteryzujących najważniejsze właściwości syntaktyczne zachodzące między formułami (np.

aksjomaty charakteryzujące spójniki logiczne, symbol identyczności, symbole specyficzne dla arytmetyki jak np. symbol odpowiadający operacji następnika czy dodawania). W konsekwencji dokonania formalizacji matematyki (a precyzyjniej mówiąc arytmetyki) możliwe staje się wyróżnienie trzech poziomów: arytmetyki formalnej, teorii liczb i metamatematyki. Pierwszy poziom wyznacza język, w którym są formułowane twierdzenia należące do rozważanej teorii, czyli poziomu drugiego; natomiast z trzeciego poziomu podejmowane jest badanie symboli systemu formalnego ujmującego określoną teorię, a konkretnie takich ciągów wyrażeń, które stanowią dowód formalny. Następnym etapem Gödlowskiego przedsięwzięcia jest arytmetyzacja języka formalnego matematyki oraz metateorii, która polega na przypisaniu liczb znakom alfabetu, wyrażeniom z nich złożonych, ciągom wyrażeń oraz dowodom formalnym. Zabieg ten jest warunkiem koniecznym, choć niedostatecznym możliwości zinterpretowania wyrażeń z metateorii i teorii na poziomie arytmetycznym, czyli w wymiarze relacji między liczbami. Arytmetyzacja umożliwia więc przełożenie relacji metamatematycznych na relacje arytmetyczne. Jest ona także warunkiem kodowania wyrażeń przy użyciu tak zwanych numerów Gödla, czyli kodów odpowiadających skończonym ciągom liczb przypisanych wyrażeniom języka formalnego arytmetyki oraz metateorii (np. metateoretyczne pojęcie dowodliwości po uprzedniej arytmetyzacji, czyli przełożeniu na operacje na liczbach, przyjmuje postać zbioru numerów Gödla formuł formalnie dowodliwych z określonego zbioru aksjomatów). Używając dopiero numeracji gödlowskiej możemy otrzymać adekwatną interpretację w arytmetyce zdania metateoretycznego. Jest tak dlatego, że każde wyrażenie, jak i ciąg wyrażeń ma określony numer gödlowski, zaś zbiór tych numerów daje klasę relacji pierwotnie rekurencyjnych[6]. Pojęcie funkcji rekurencyjnie pierwotnej od czasu sformułowania przez Gödla jej definicji, należy do kanonu pojęć logiki matematycznej[7]. Badania Churcha pozwalają interpretować zbiory rekurencyjne jako „efektywnie obliczalne", zaś rekurencyjnie przeliczalne jako „efektywnie wypisywalne"[8]. Oznacza to, że rekurencyjność byłaby równoważna obliczalności przez algorytmy, a więc procedurę dającą w skończonej liczbie kroków efektywne rozstrzygnięcie. Pojęcie funkcji rekurencyjnej jest o tyle

[6] Na przykład zbiór numerów formuł, symboli relacyjnych, funkcyjnych, ciągów skończonych formuł stanowiących dowód formalny; wyjątek stanowi zbiór numerów formuł formalnie dowodliwych, które są jedynie rekurencyjnie przeliczalne.

[7] Funkcja pierwotnie rekurencyjna to funkcja otrzymywana z prostych funkcji wyjściowych (funkcji następnika, rzutu czy funkcji stałej), w wyniku stosowania operacji składania i rekursji prostej. Funkcja rekurencyjnie przeliczalna jest zaś rzutem funkcji rekurencyjnych. Szerzej na ten temat: S. K r a j e w s k i: *Twierdzenie Gödla i jego interpretacje filozoficzne...*, s. 40-51; Zob. także: R. M u r a w s k i: *Funkcje rekurencyjne i elementy metamatematyki. Problemy zupełności, rozstrzygalności, twierdzenia Gödla*. Poznań 1990.

[8] Zob. S. K r a j e w s k i: *Twierdzenie Gödla i jego interpretacje filozoficzne...*, s. 97.

istotne, że określa pewną mechaniczną procedurę otrzymywania obiektów matematycznych (np. funkcji, zbiorów), a to z kolei gwarantuje uzyskanie dwóch bardzo pożądanych efektów. Po pierwsze, daje kontrolę nad całością zakodowanego systemu, eliminując jakąkolwiek dwuznaczność w identyfikacji wyrażenia, któremu odpowiada dany numer gödlowski; krótko mówiąc zapewnia efektywną procedurę odkodowania. Po drugie, dzięki rekurencyjnej własności, funkcje i relacje zachodzące między wyrażeniami stają się reprezentowalne (mniej ściśle: definiowalne) w arytmetyce, a to umożliwia zachowanie odpowiedniości czy też ścisłej więzi pomiędzy wspomnianymi poziomami systemu[9].

Powyższy wykaz kilku elementów maszynerii, którą Gödel wykorzystał do wyprowadzenia swoich twierdzeń, stanowi także zestaw kryteriów ich stosowalności. Po pierwsze, twierdzeniom Gödla podlegają tylko takie teorie, które zawierają co najmniej elementarną arytmetykę (liczby naturalne oraz dodawanie i mnożenie). Po drugie teorie są formalizowalne i aksjomatyzowalne. Po trzecie, teorie podlegają wymogowi arytmetyzacji. Po czwarte relacje, funkcje i zbiory określające i składające się na całość teorii muszą mieć charakter rekurencyjny, tzn. otrzymywanie ich wartości musi być wynikiem zastosowania operacji algorytmicznych. Należy zaznaczyć, że powyższe kryteria są tutaj traktowane jako zbiór warunków, z których każde z osobna jest warunkiem koniecznym, a wszystkie razem tworzą warunek dostateczny. Tak więc uprawnione będzie stosowanie twierdzeń Gödla tylko do takich dziedzin wiedzy czy nauki, w ramach których można sformułować teorie spełniające wszystkie wymienione warunki.

Stąd już może wydać się jasne dlaczego twierdzenia Gödla nie dają się zastosować do nauk społecznych czy humanistycznych. Trudno bowiem byłoby utrzymywać możliwość poddania obróbce formalnej (w sensie logiki) problematyki dotyczącej społeczeństwa, państwa, prawa czy tekstów literackich. Dla celów niniejszego artykułu najistotniejsze jest jednak zbadanie stosowalności twierdzenia Gödla na polu filozofii umysłu, a konkretnie w sporze antymechanicyzmu z mechanicyzmem. Sprawa nabiera wielkiej wagi w momencie, gdy uprzytomnimy sobie, że zdaniem Lucasa twierdzenie Gödla nie tyle wzmacnia stanowisko antymechanicyzmu, co dowodzi wyższości umysłu ludzkiego nad maszyną, a w konsekwencji obala mechanicyzm. W pierwszej więc kolejności należy przyjrzeć się czy zastosowanie argumentu używającego twierdzenia Gödla jest w ogóle uprawnione. Dopiero spełnienie tego warunku czyni zasadnym dalsze analizy nad ważnością i prawidłowością przeprowadzonego przez Lucasa wywodu. A zatem czy Lucas miał prawo sądzić, że twierdzenia Gödla mogą wykazać fałszywość

[9] Zob. więcej na temat aparatury pojęciowej i technicznej w: tamże, s. 17-76 oraz A. O l s z e w s k i: *Teza Churcha a Twierdzenie Gödla*. http://www.loic.amu.edu.pl (dostęp dnia: 20.01.2010), s. 1-6.

mechanicyzmu? Jeżeli tak, to argument Lucasa będzie trzeba potraktować bardzo poważnie.

Wygłoszona przez Lucasa teza w pierwszym zdaniu jego głośnego i nośnego tekstu jest następująca: „Twierdzenie Gödla zdaje się, moim zdaniem, dowodzić, że mechanicyzm jest fałszywy, tj. że nie można pojmować umysłów jako maszyn”[10]. Uzasadnieniem tej tezy jest twierdzenie, że maszyna nie może dokonać czegoś, co jest osiągalne dla człowieka, a mianowicie nie może *dostrzec*, że pewne formuły choć niedowodliwe w systemie, są prawdziwe. My zaś ich prawdziwość *widzimy*. Zatem Lucas używa twierdzenia Gödla do pokazania pewnych ograniczeń maszyny, którym człowiek ma nie podlegać. Skoro maszyny mają podlegać pewnym ograniczeniom wynikającym z twierdzeń Gödla, to odpowiedź na zadane w poprzednim akapicie pytanie jest uwarunkowana możliwością ukazania maszyny jako odpowiednika formalnego systemu zawierającego elementarną arytmetykę, w której ponadto zachodzą procesy, które mogą zostać opisane jako funkcje rekurencyjne.

Lucas używa określenia „maszyny cybernetyczne”, przez co rozumie „[...] urządzenie, które wykonuje pewien zestaw operacji zgodnie z określonym zestawem reguł. Zazwyczaj »programujemy« maszynę, tj. wydajemy jej zestaw instrukcji, mówiących co robić w poszczególnych sytuacjach i wprowadzamy początkowe »informacje«, w oparciu o które maszyna ma przeprowadzać swoje obliczenia”[11]. Taka koncepcja maszyny odpowiada tak zwanym maszynom Turinga, czyli matematycznym idealizacjom maszyn[12]. Mają one nieograniczoną pamięć, składającą się z komórek, którym odpowiadają znaki z ustalonego skończonego alfabetu. Określa ją także skończona liczba stanów (mieszczących się między stanem początkowym a końcowym) oraz instrukcji. Te ostatnie określają sekwencje operacji, które odpowiadają pewnym stanom. Tak określonym elementom składowym maszyny możemy przyporządkować elementy systemu formalnego: instrukcje można rozumieć jako zbiór aksjomatów i reguł wnioskowania, stanom przypisalibyśmy formuły języka, a sekwencje operacji to obliczenia maszyny, którym odpowiadają dowody formalne. Jak wiemy w systemie formalnym możemy „zagnieździć” arytmetykę, a w ślad za tym możemy stworzyć maszynę do generowania twierdzeń arytmetycznych. Okazuje się także, że wszelkie operacje, czyli funkcje obliczalne przez maszyny Turinga dają funkcje rekurencyjne. Należy zauważyć, że wszystkie wymienione cechy odnoszą się

---

[10] J. R. Lucas: *Umysły, maszyny i Gödel*. Tłum. M. Zawidzki. http://www.filozof.uni.lodz.pl (dostęp dnia: 18.01.2010), s. 96.

[11] Tamże, s. 98.

[12] Zob. więcej M. Hetmański: *Maszyna Turinga a umysł ludzki*. http://www.kognitywistyka.net (dostęp dnia: 21.01.2010), s. 1-6.

zarówno do maszyn deterministycznych, jak i niedeterministycznych. Te ostatnie (rozważane przez Lucasa jako bliższe mechanicznemu modelowi umysłu), posiadają dodatkowo generator liczb losowych, czyli urządzenie losujące, służące do losowania następnego kroku. Lucas trafnie zauważa, że „[...] jest jednak oczywiste, że urządzenia losującego nie można by zainstalować w maszynie tak, by wybierało dowolną możliwość; dopuszczalny byłby jedynie wybór pomiędzy pewną liczbą dozwolonych możliwości"[13]. Chodzi o to, że program maszyny losującej byłby także wyznaczony przez skończoną liczbę instrukcji (zestaw reguł i aksjomatów) określających skończoną liczbę możliwych typów operacji. A zatem taka maszyna nie mogłaby wykonywać zupełnie dowolnych operacji. W ramach alternatywy dającej jej możliwość wyboru, mogłaby wykonać tylko taką operację, która nie prowadzi do sprzeczności. Uwagi Lucasa są zgodne z ustaleniami Turinga, w myśl których maszyny niedeterministyczne nie mogą zrobić nic więcej niż zwykłe[14]. Maszyny losujące można bowiem „przejrzeć" dzięki opisaniu struktury wszystkich możliwości. Funkcje obliczalne przez niedeterministyczne maszyny są więc także rekurencyjne. Toteż w świetle powyższego powinno stać się jasne, że maszyny w ujęciu Lucasa są równoważne systemom formalnym, a więc spełniają wszystkie warunki, które zostały postawione dla stosowalności twierdzeń Gödla.

III. Granice ważności i stosowalności argumentu Lucasa

Można zatem w sposób uprawniony przyjąć przypuszczenie, że jeżeli argument Lucasa i składające się nań rozumowanie jest poprawne, to dowodzi wyższości umysłu nad maszynami Turinga. Trzeba jednak być ostrożnym. Nie mówimy bowiem o dowolnych maszynach, lecz tylko o takich, które odpowiadają powyższemu ujęciu. Stosowne jest więc pytanie, czy można dopuścić istnienie lub powstanie maszyn różniących się od maszyn Turinga? Jeżeli tak by było, to argument Lucasa, byłby stosowalny tylko do pewnej kategorii maszyn, a w rezultacie dowodziłby jedynie wyższości człowieka nad maszynami Turinga. A to nie wystarczyłoby do obalenia mechanicyzmu.

Minęło już niemal pół wieku od buńczucznej zapowiedzi Minsky'ego – pioniera badań nad sztuczną inteligencją – że zbudowanie maszyny dorównującej człowiekowi pod względem inteligencji jest kwestią kilku lat. Choć postęp w tej dziedzinie od tamtego czasu jest znaczący, to wydaje się, że daleki od tak górnolotnych oczekiwań. Wśród ważnych zdobyczy zwolenników czy ideologów sztucznej inteligencji należy odnotować tak zwane algorytmy uczące się i sieci

[13] J. R. L u c a s: *Umysły...*, s. 99.
[14] Zob. S. K r a j e w s k i: *Twierdzenie Gödla i jego interpretacje filozoficzne...*, s. 99.

neuropodobne. W praktyce przynoszą godne uznania efekty, ale mimo to, nie mają większego przełożenia na refleksję teoretyczną. Po pierwsze dlatego, że sieci neuronowe dają się symulować już na zwykłych komputerach, a po drugie ich stosowanie skutkuje wprowadzeniem funkcji, które nie wykraczają poza zakres funkcji rekurencyjnych, obliczalnych przez maszyny Turinga. Nie można jednak na podstawie dotychczasowych niepowodzeń w powieleniu ludzkiego umysłu wykluczyć możliwości, że w przyszłości pojawią się maszyny o dużo większej mocy niż obecne. Co to jednak miałoby znaczyć? Maszynę zdolną do wnioskowania indukcyjnego, operowania w sferze znaczeniowej[15], a może po prostu ilościowy wzrost struktur składających się na maszynę? Przykładowo Churchlandowie wyrażają nadzieję, że „chińska sala gimnastyczna" przekroczy ograniczenia chińskiego pokoju Searla i zrozumie język chiński. Sam Lucas tego nie wyklucza, twierdząc w nieco marksistowskim duchu, że „Złożoność często pociąga za sobą zmiany jakościowe […] powyżej pewnego poziomu złożoności maszyna przestanie być przewidywalna, nawet z zasady, i zacznie działać na własną rękę"[16]. Zaraz jednak zręcznie dodaje, że wówczas maszyna przestanie być maszyną i uzyska swój własny umysł.

Nie popadając w przesadę i stronniczość można stwierdzić, że wizja wyrafinowanej i rozumiejącej maszyny jest kwestią otwartą, choć dzisiaj zdecydowanie bliższą *science fiction* i beletrystyce futurystycznej niż sferze naukowych hipotez[17]. Toteż wydaje się, że pozostaje nam na razie polegać na analizach obliczalności maszynowej wykonanych przez Turinga i jego kontynuatorów[18]. Te wskazują, że każde dowolnie długie obliczanie w przypadku dowolnej maszyny oznacza ciąg lokalnych, osobnych operacji, którym odpowiada zapis złożony ze skończonej (choć nieograniczonej) liczby klatek pustych bądź wypełnionych symbolem. W każdym momencie więc operacja jest identyfikowalna i wyznaczona przez skończony spis instrukcji. To daje racjonalne podstawy, by sądzić, że nie pojawi się jakaś nowa reguła dowodzenia, która nie będzie rekurencyjna, czyli efektywnie obliczalna. W konsekwencji wydaje się też racjonalne, by przyjąć, że możliwość wykazania wyższości umysłu nad maszyną Turinga obejmuje dowolną maszynę, co (dzisiaj) sprowadza się do urządzenia, którego produktem jest zbiór wyrażeń co

---

[15] J. Searle dopuszcza możliwość zbudowania maszyny zdolnej do semantycznego przetwarzania informacji, ale pod warunkiem, że byłaby zbudowana z innego materiału niż dotychczasowe programy komputerowe. Zob. więcej J. S e a r l e: *Umysł, mózg i nauka*. Tłum. J. Bobryk. Warszawa 1995; podobne stanowisko przyjmuje H. Putnam w pracy *The mental life of some machines* z 1979, zob. więcej na ten temat w: P. K o ł o d z i e j c z y k: *Funkcjonalizm jako filozoficzna podstawa teorii Sztucznej Inteligencji*. http://www.kognitywistyka.net (dostęp dnia: 21.01.2010).

[16] J. R. L u c a s: *Umysły...*, s. 115.

[17] Zob. S. K r a j e w s k i: *Twierdzenie Gödla i jego interpretacje filozoficzne...*, s. 99-101, 103-105; zob. także M. H e t m a ń s k i: *Maszyna Turinga...*, s. 4-18.

[18] Szerzej na ten temat: S. K r a j e w s k i: *Twierdzenie Gödla i jego interpretacje filozoficzne...*, s. 96-98.

najmniej rekurencyjnie przeliczalny.

IV. Błędne koło po raz pierwszy oraz błąd ekwiwokacji

W jednym z poprzednich akapitów został naszkicowany argument Lucasa. Tamto ujęcie może jednak budzić uzasadnione zastrzeżenia, gdyż zostaje w nim użyta w odniesieniu do maszyny kategoria semantyczna; mianowicie posługuje się pojęciem prawdziwości dla opisania mechanicznych operacji. O ile przed następującymi później ustaleniami, zabieg „semantyzacji" maszyny mógł wydawać się niewinny, o tyle po świadomym przyjęciu ograniczenia się do maszyn Turinga, owa swobodna stylistyka przybiera charakter merytorycznej niezręczności. Wynika to z faktu, iż przyjęcie rzeczonego ograniczenia pociąga za sobą konieczność ograniczania się do kategorii syntaktycznych określających relacje wyłącznie między wyrażeniami bez odnoszenia ich do potencjalnie „zakodowanej" w nich rzeczywistości. Sprawa nie jest błaha, ponieważ na pomoc Lucasa w zniwelowaniu tej niezręczności nie ma co liczyć. On sam notorycznie używa pojęcia prawdziwości w odniesieniu do maszyny, a co najistotniejsze, stosuje ten niefortunny zabieg, wówczas, gdy przedstawia istotę swojego argumentu. Oto symptomatyczny fragment: „[...] jeśli daną mamy jakąkolwiek maszynę, która jest niesprzeczna i zdolna generować prostą arytmetykę, istnieje formuła [zdanie Gödla – M.B], której nie jest w stanie przedłożyć jako prawdziwej [oryg. *produce as being true* – M.B], tj. formuła, która jest niedowodliwa-w-systemie, lecz której prawdziwość my widzimy"[19]. Trzeba jednak Lucasowi oddać sprawiedliwość i zauważyć, że stosowanie określenia „prawdziwy" w tym kontekście nie jest w jego przypadku wynikiem przeoczenia lub braku staranności, ale wpisuje się w przemyślaną strategię. W innym fragmencie wyjaśnia nam bowiem powód, dla którego posługuje się takim określeniem: „[...] każdy mechaniczny model umysłu musi zawierać mechanizm, który może oznajmiać prawdy arytmetyki, gdyż jest to coś, co umysły umieją czynić"[20]. Nie można zgodzić się jednak, że jest to satysfakcjonujące załatwienie sprawy i to z dwóch powodów.

Po pierwsze Lucas balansuje na granicy błędnego koła. Można mu zarzucić, że już w założeniu „przemyca" to, czego chce dowieść. Stawianie „maszynom cybernetycznym" wymogu oznajmiania prawd może bowiem czynić zbytecznym posługiwanie się twierdzeniami Gödla do wykazania naszej wyższości. Wystarczy, że zauważymy niezdolność maszyny do operowania w sferze semantycznej i w ten sposób uzyskamy nad nią przewagę. Drugim błędem popełnianym

[19] J. R. L u c a s: *Umysły...*, s. 98.
[20] Tamże, s. 101.

przez Lucasa jest ekwiwokacja. Używa pojęcia „prawdziwości" w dwóch znaczeniach: gdy mówi, że maszyna „produkuje", „przedkłada" jako prawdziwe, utożsamia „prawdziwość" z „dowodliwością"; gdy zaś mówi, że człowiek „widzi" prawdziwość zdania, ma na myśli jakieś rozumienie, uznanie, akceptację. Jeszcze przed wynikami Gödla czymś naturalnym było utożsamianie prawdy z dowodem, ale to właśnie Gödel pokazał, że twierdzenie może być prawdziwe, a mimo to nie posiadać dowodu. Oczywiście trudno przypuszczać, by Lucas o tym nie wiedział. Zapewne zdawał sobie także sprawę z tego, że twierdzenie Gödla można sformułować posługując się wyłącznie kategoriami syntaktycznymi: w każdym systemie niesprzecznym zawierającym co najmniej arytmetykę istnieje formuła G taka, że ani ona ani jej negacja nie jest wywodliwa z aksjomatów i z twierdzeń z owych aksjomatów wywiedzionych. Dlaczego więc używa zbędnego pojęcia prawdziwości narażając się na zarzut o błąd ekwiwokacji? Niektórzy twierdzą, że stosowanie ekwiwokacji między dowodliwością a prawdziwością warunkuje wręcz ustanowienie argumentu w stylu Lucasa. S. Krajewski relacjonuje ten problem następująco: „[...] używając określenia »produkuje jako prawdziwe«, Lucas może mówić jednocześnie o ograniczeniach maszyny i o braku ograniczeń dla możliwości umysłu"[21], a dzięki temu argument Lucasa może „[...] traktować o maszynach i ludziach jednocześnie, a zarazem nie dopuszczać do ich utożsamienia"[22].

Odrzucenie argumentu Lucasa na podstawie stwierdzenia ekwiwokacji byłoby jednak próbą „zamiecenia problemu pod dywan". Widać, że w argumencie Lucasa coś jednak leży na rzeczy: faktycznie jest tak, że na mocy pierwszego twierdzenia Gödla maszyna jako odpowiednik systemu formalnego nie może przedłożyć zdania Gödla (niedowodliwej formuły) jako ...prawdziwego; natomiast my wiemy, że to zdanie przynależy do systemu w sposób, by tak rzec, pełnoprawny, to znaczy widzimy, że jest ono prawdziwe. Może więc zbagatelizować problem ekwiwokacji jako sztuczny? Nie należy go jednak tak traktować. Problem stosowania kwalifikacji dla jakościowej oceny czynności wykonywanych przez maszyny (czyli, jak określić to, co ona przedkłada, produkuje) jest problemem dość poważnym. Bez odpowiednich kwalifikacji nie moglibyśmy bowiem odróżnić jakichś produktów od arytmetycznych produktów maszyny. Innymi słowy, trzeba umieć odróżniać skonstruowanie formuły Gödla przez maszynę (co jest oczywiście możliwe, gdyż formuła Gödla nie tylko jest zbudowana ze znaków skończonego alfabetu pozostającego do dyspozycji maszyny, ale po prostu jest sensowną formułą zbudowaną poprawnie wedle reguł) od

[21] S. K r a j e w s k i: *Twierdzenie Gödla i jego interpretacje filozoficzne...*, s. 102.
[22] Tamże.

przedłożenia jej wśród innych twierdzeń arytmetycznych, co możliwe nie jest. Wydaje się, że skutecznym rozwiązaniem problemu ekwiwokacji jest pomysł wysunięty przez Penrose'a. Jego propozycja sprowadza się do tego, żeby wprowadzić specjalne oznaczenie[23]. Wtedy wyrażenie „przedłożyć twierdzenia jako prawdziwe" można byłoby zastąpić wyrażeniem „przedłożyć twierdzenia w sposób oznaczony specjalnym symbolem". S. Krajewski zaś proponuje „zielone światełko", które zapalałoby się w momencie, gdy maszyna przedkłada twierdzenia arytmetyczne[24]. W następstwie tych modyfikacji otrzymujemy argument, w myśl którego maszyna nie może przedłożyć zdania Gödla wśród twierdzeń oznaczonych specjalnym symbolem zastrzeżonym dla określonej grupy zdań, w której znajdują się twierdzenia arytmetyczne (inaczej: których prawdziwość możemy stwierdzić); ludzie natomiast potrafiliby przedłożyć tą formułę w specjalny sposób, (gdyż „widzą" jej prawdziwość). Wyróżnione operacje maszyny będzie można teraz odczytać nie jako „przedłożenie twierdzenia jako prawdziwego", ale „symulację przedłożenia twierdzenia jako prawdziwego", co czyni zadość specyfice maszyn Turinga.

V. Błędne koło po raz drugi oraz kłopot ze sprzeczną maszyną

Dlaczego właściwie maszyna nie może przedłożyć formuły Gödla w sposób specjalnie oznaczony? Lucas przyjmuje dwa założenia: po pierwsze, maszyna jest niesprzeczna, a po drugie jeżeli jest niesprzeczna, to nie przedkłada zdania Gödla, gdyż w przeciwnym razie popadłaby w sprzeczność. Dlaczego zdaniem Lucasa maszyna musiałaby popaść w sprzeczność? Żeby to zrozumieć trzeba wiedzieć, jakiego rodzaju zdaniem jest zdanie Gödla. Jest to zdanie należące do klasy zdań samozwrotnych[25]. Stanowi przeciwieństwo antynomii kłamcy, które jest fałszywe (czy raczej prowadzi do sprzeczności) przy każdym założeniu; natomiast ono przy każdym założeniu jest prawdziwe[26]. Zdanie Gödla może przybrać postać: „niniejsze zdanie jest niedowodliwe w teorii T". Jeżeli „niniejsze zdanie jest niedowodliwe w teorii T" jest fałszywe to prawdą jest, że to zdanie jest dowodliwe, ale skoro jest dowodliwe, to znaczy, że „niniejsze zdanie jest niedowodliwe w teorii T" jest prawdziwe, gdyż w teorii niesprzecznej wszystko co może zostać dowiedzione jest

---

[23] Zob. R. P e n r o s e: *Cienie umysłu. Poszukiwanie naukowej teorii świadomości*. Tłum. – P. Amsterdamski. Poznań 2000, s. 205

[24] Zob. S. K r a j e w s k i: *Twierdzenie Gödla i jego interpretacje filozoficzne...*, s. 106.

[25] Szerzej na ten temat: W. M a r c i s z e w s k i: *Szkic uzasadnienia Twierdzenia Gödla o nieusuwalnej niezupełności arytmetyki liczb naturalnych*. http://www.logic.amu.edu.pl (dostęp dnia: 20.01.2010), s. 4-5; zdaniem S. Krajewskiego, precyzyjne rozróżnienie poziomów, o których była mowa w drugim paragrafie niniejszego artykułu, może zniwelować efekt samozwrotności. Zob. S. K r a j e w s k i: *Twierdzenie Gödla i jego interpretacje filozoficzne...*, s. 123.

[26] Tautologią logiczną zapewniającą wynikanie prawdziwości z fałszywości jest: (nie-p→p)→p.

prawdziwe. Skoro jednak „niniejsze zdanie jest niedowodliwe w teorii T" jest prawdziwe, to jest tak jak mówi, czyli jest niedowodliwe. Trzeba więc zgodzić się z Lucasem, że gdyby maszyna takie zdanie przedłożyła w sposób zastrzeżony dla zdań dowodliwych, to popadłaby w sprzeczność. Dla maszyny kryterium zaliczenia jakiegokolwiek zdania do grupy zdań mających status aksjomatów lub z nimi spokrewnionych (dzięki więzi wywodliwości), czyli takich, które my uznajemy za prawdziwe, jest dowodliwość. A zdanie Gödla jest prawdziwe dzięki swojej niedowodliwości. Maszyna musiałaby albo zastosować podwójne standardy i to tylko raz specjalnie dla zdania Gödla i/lub doznałaby czegoś w rodzaju „mechanicznego olśnienia", co jednak możemy potraktować jako humorystyczne li tylko hipotezy. Przedkładając maszynie niesprzecznej formułę Gödla dla jej systemu (odpowiadającego jej teorii) możemy ją więc „wygödlować". Próba wygödlowania naturalnie nie może się powieść wtedy, gdy mamy do czynienia z maszyną sprzeczną. W systemie sprzecznym na mocy tautologii A∩nie-A→B wszystko jest dowodliwe, a więc taka maszyna potrafiłaby przedłożyć w sposób specjalny formułę Gödla. Ale właśnie przedłożenie jej przez maszynę byłoby, zdaniem Lucasa, świadectwem jej sprzeczności. Jako taka jest zaś zdyskwalifikowana, gdyż nie może stanowić modelu równoważnego dla umysłu człowieka, który jest, w opinii brytyjskiego filozofa, niesprzeczny. Hilary Putnam wysuwa jednak tezę, że nie możemy wykluczyć, że jesteśmy maszynami sprzecznymi[27]. Co na to Lucas?

Starając się stawić czoło tej brawurowej hipotezie Putnama, Lucas dokonuje ważnego rozróżnienia, które moglibyśmy przedstawić posługując się określeniami „sprzeczność niezbywalna" oraz „sprzeczność okazjonalna". Ten pierwszy rodzaj sprzeczności odpowiadałby jakiejś immanentnej cesze wpisanej w genotyp odpowiadający za rozumowania. Drugi rodzaj byłby równoważny omylności, czy też „[...] przypadkowym awariom maszyny, nie zaś normalnemu schematowi działań"[28]. Wszelkie paradoksy, które dotknęły bądź dotkną teorie naukowe są, zdaniem Lucasa, wynikiem naszych pomyłek, a nie wyrazem sprzeczności naszej natury. Najdobitniejszym świadectwem, iż tak się rzeczy mają, ma być nasza skłonność do niestrudzonego naprawiania wadliwych teorii i bezwzględnego odrzucania tych, które okazały się sprzeczne. Podobnie jest w życiu codziennym. Wprawdzie zdarza się nam zaprzeczać sobie samym, to jednak nie dlatego, że uważamy, że wszystko można udowodnić i wszystko ujdzie. Popadamy w sprzeczność albo wskutek pomyłki intelektualnej albo obłędu: „gdy dana osoba gotowa jest stwierdzić absolutnie wszystko i zaprzeczać sobie bez żadnych skrupułów i odrazy, wówczas uznaje

[27] Por. H. P u t n a m: *Mind and Machines*. http://www.csun.edu (dostęp dnia: 15.01.2010).
[28] J. L u c a s: *Umysły...*, s. 109.

się, że »straciła rozum«”[29]. Tak więc człowiek, eliminując sprzeczności, ma zaświadczać o swojej niesprzeczności, która niczym idea regulatywna kieruje naszymi dążeniami poznawczymi. Czy jednak zabiegi samonaprawcze można uznać za dowód naszej niesprzeczności? Bez wątpienia można je uznać za wyraz racjonalności i rozsądku, za konieczność wynikającą ze społecznego współżycia, za warunek *sine qua non* wszelkiego dyskursu, ale nie za dowód naszej niesprzeczności. Z tego, że nauczyłem się nie tolerować sprzeczności nie wynika bynajmniej, że nie jestem sprzeczny, a jedynie, że sprzeczność uważam z jakichś względów za stan dolegliwy i niepożądany. Tą drogą nie sposób wykluczyć możliwość istnienia jakiejś fundamentalnej sprzeczności tkwiącej w głębokich zasobach naszego umysłu. Takie przypuszczenie zyskuje zresztą na wiarygodności, gdy posłuchamy informatyków. Ci twierdzą, że programy o wysokim poziomie złożoności zawierają w sobie nieusuwalne błędy, jednakże mogące ujawnić się tylko w wyjątkowych okolicznościach[30]. Tak więc skoro komputery działają normalnie i poprawnie pomimo obecności pewnych usterek programowych, to dlaczego z nami miałoby być inaczej?[31] A nawet jeżeli przyjmiemy, że nasze skłonności samonaprawcze i wynikające stąd poczucie stanowią dostateczny dowód naszej niesprzeczności, to będzie to dowód nieformalny oparty na samoświadomości i introspekcji. Lucas zdaje sobie z tego doskonale sprawę. Broni zasadności używania dowodów nieformalnych odwołując się nawet do twierdzeń Gödla mówiąc, że „[...] cały wydźwięk rezultatów Gödla jest taki, że nie powinniśmy szukać całkowitej formalizacji i nie możemy jej osiągnąć”[32]. Nie odmawiając mu prawa do stosowania dowodów nieformalnych, możemy jednak zasadnie takiej linii obrony zarzucić cyrkularność: zakładając u człowieka zdolność do stosowania dowodu nieformalnego opartego na wewnętrznym wglądzie, zakłada zdolności niemechaniczne, których chce dowieść. Mówiąc inaczej: oddala hipotezę, w myśl której jesteśmy sprzecznymi maszynami w sposób niezależny od swojego podstawowego argumentu, w rezultacie czyniąc ten ostatni zbytecznym. Maszyna sprzeczna może przedłożyć zdanie Gödla, a w konsekwencji nie podlega ograniczeniom gödlowskim. W takiej sytuacji, aby odróżnić człowieka od sprzecznej maszyny trzeba wykazać jego niesprzeczność. W celu uniknięcia błędnego koła należałoby zastosować dowód formalny. Ale tego nie możemy wykonać z uwagi na drugie twierdzenie Gödla. Ktoś mógłby nam zarzucić, że taka konstatacja jest wynikiem nieuprawnionego utożsamienia człowieka z jakimś systemem formalnym, bo tylko taki podlega ograniczeniom

[29] Tamże.
[30] Zob. S. K r a j e w s k i: *Twierdzenie Gödla i jego interpretacje filozoficzne...*, s. 112.
[31] Inną okolicznością, osłabiającą tezę o niesprzeczności człowieka, jest budowanie logiki parakonsystentnej, która dopuszcza pojawienie się sprzeczności w systemie. Zob. więcej o takiej logice w: A. P i e t r y g a: *Status zasady sprzeczności w świetle logiki współczesnej*. Kraków 2004.
[32] J. R. L u c a s: *Umysły...*, s. 113.

godlöwskim. Niewykonalność takiego dowodu możemy stwierdzić jednak na innej drodze. Jeżeli wykażemy, że jakiś system (maszyna) odpowiadający części naszych mocy matematycznych nie może przeprowadzić takiego dowodu, będziemy mogli stwierdzić, że tym bardziej obszerniejszy system, a więc i człowiek nie może. System formalny zaś nie może wykazać swojej własnej niesprzeczności nie popadając przy tym w sprzeczność, co gwarantuje drugie twierdzenie Gödla.

Tak więc na drodze formalnej nie możemy stwierdzić swojej niesprzeczności; natomiast na drodze nieformalnej, jeżeli możemy, to jedynie w sposób, który nieuchronnie łączy się z jednoczesnym wykazaniem naszej wyższości nad maszyną niezależnie od procedury wygödlowania. Niemniej jednak należy zgodzić się z Lucasem, że aktem ze wszech miar racjonalnym jest uznanie, że jesteśmy niesprzeczni. Nie można jedynie zgodzić się, że zostało to udowodnione. Ale nawet jeżeli przyjmiemy, że jesteśmy niesprzeczni, to wcale nie musi to nam zagwarantować sukcesu w wygödlowaniu maszyny. Czy faktycznie procedura wygödlowania maszyny niesprzecznej się powiodła?

VI. Błędne koło po raz trzeci i kłopot z dialektyką Lucasa

Istota argumentu Lucasa polega na – wydawałoby się – niezwykłej, zastrzeżonej tylko dla ludzi zdolności widzenia prawdziwości zdania Gödla. Lucas ogłasza, na podobieństwo znajdującego się w ekstatycznym uniesieniu mistyka, że „[...] my stojąc na zewnątrz systemu potrafimy zobaczyć jego prawdziwość”[33]. Nasuwa się naturalnie pytanie czy jest to jakaś szczególna prawda, do ujrzenia której potrzeba wzmożenia wszystkich dostępnych człowiekowi mocy. Skoro nawet maszyna, która przewyższa człowieka szybkością i wydajnością jest ślepa na tę dostrzegalną przez nas prawdę, to poczucie tajemniczości jeszcze bardziej się nam udziela. Z jakim rodzajem prawdziwości mamy do czynienia?

Standardowa procedura, której wynikiem jest dostrzeżenie prawdziwości zdania Gödla polega na tym, że najpierw dowodzimy twierdzenia Gödla, które stwierdza niedowodliwość pewnej formuły, a następnie odwołujemy się do samego zdania Gödla. Czyli raz znajdujemy się na poziomie metateoretycznym i wykazujemy właściwość pewnego zdania, a potem na poziomie arytmetycznym, na którym badamy czy relacje jakie zachodzą tam między liczbami (a konkretnie numerami godlöwskimi) odpowiadają relacji zawartej w twierdzeniu z metapoziomu. I widzimy, że to co mówimy o pewnej formule (o jej niedowodliwości) potwierdza się w rzeczywistości, gdyż

[33] J. R. L u c a s: *Umysły...*, s. 107.

sama ta formuła stwierdza swoją niedowodliwość. Czy do wykazania prawdziwości zdania Gödla konieczne jest jednak wychodzenie poza system? Operację dowodzenia zdania Gödla można wykonać zwykłymi środkami logicznymi w ramach samej arytmetyki. Chodzi tutaj o wykazanie, że zachodzi równoważność: istnieje liczba, która jest numerem m pewnego zdania G wtedy i tylko wtedy, gdy nie istnieje liczba, która jest numerem dowodu zdania o numerze, który ma nazwę „m", czyli G(m) wtw nie Dow(„m",n), gdzie, m to numer G, Dow to relacja dowodu, „m" liczebnik, a n jest dowolną liczbą stanowiącą numer dowodu[34]. Dlaczego jednak wychodzimy na poziom metateorii? Ponieważ tylko stamtąd możemy wnosić, że system jest niesprzeczny, a to jest warunkiem prawdziwości zdania Gödla, a w rezultacie samej możliwości wygödlowania maszyny. Prawdziwość zdania Gödla dla teorii T nie jest nadzwyczajną prawdziwością daną dzięki szczególnemu wglądowi, ale jest wynikiem przyjęcia założenia, że teoria jest niesprzeczna. Dlatego jasne się staje dlaczego maszyna nie może zastosować tych prostych w gruncie rzeczy środków dowodowych dla podanej wyżej równoważności, z której wynika, że istnieje w teorii T formuła niedowodliwa. Mówiąc w dużym uproszczeniu, dowód tej równoważności jest dowodem nie wprost i polega na uznaniu, że formuła Gödla jest twierdzeniem teorii, a więc jest dowodliwa, co prowadzi nas do sprzeczności; ale my wiemy, że ta uzyskana sprzeczność jest wyrazem prawdziwości formuły Gödla, gdyż teoria jest niesprzeczna. Skąd jednak wiemy, że teoria jest niesprzeczna?

Hilary Putnam zauważył, że do wygödlowania maszyny nie wystarcza wiedza pochodząca z pierwszego twierdzenia Gödla[35]. Nie wystarczy więc, że wiemy, że jeżeli teoria jest niesprzeczna to zdanie Gödla jest prawdziwe, trzeba jeszcze wiedzieć, że teoria jest niesprzeczna, by móc stwierdzić, że zdanie Gödla jest prawdziwe. To w zasadzie banalne spostrzeżenie wykorzystujące prostą regułę *modus ponens* zadało poważny cios rozumowaniu Lucasa. Jak wiemy, nie można znaleźć dowodu niesprzeczności niesprzecznej teorii w sposób zaproponowany przez Hilberta, czyli za pomocą środków dowodowych w ramach tej teorii (w sposób absolutny). Jest jednak jeszcze inny sposób, który można scharakteryzować jako „relatywny" dowód niesprzeczności teorii. Polega na adekwatnej interpretacji interesującej nas teorii w ramach innej teorii czy innego modelu uważanego za niesprzeczny. Niesprzeczność teorii T „przełożonej" na język teorii N jest więc uwarunkowana niesprzecznością teorii N. Gołym okiem widać, że taka metoda obarczona jest wadą regresu w nieskończoność, który możemy przerwać jedynie naszą intuicją prawdziwości

[34] Zob. więcej S. K r a j e w s k i: *Twierdzenie Gödla i jego interpretacje filozoficzne...*, s. 118; zob. także A. O l s z e w s k i: *Teza Churcha a Twierdzenie Gödla...*, s. 1-6.
[35] Zob. S. K r a j e w s k i: *Twierdzenie Gödla i jego interpretacje filozoficzne...*, s. 119.

aksjomatów określonej teorii. Przykład teorii mnogości pokazuje jednak, że to co osadzone w zdroworozsądkowym poczuciu nie zawsze musi być niesprzeczne. Nie znaczy to oczywiście, że należy zwątpić w aksjomaty arytmetyczne (np. następnik liczby nie jest tej liczbie równy), trzeba jednak pamiętać, że nie możemy mieć absolutnej pewności co do niesprzeczności elementarnej arytmetyki, a co za tym idzie nie możemy wykluczyć jej sprzeczności. Poza teoretyczną zawodnością dowodu niesprzeczności zrelatywizowanego do jakiegoś modelu, trzeba zwrócić uwagę także na praktyczne trudności. Nie jest bowiem powiedziane, że każdą nową teorię uda nam się zdefiniować i zinterpretować w innym modelu.

W mniemaniu Lucasa przedstawione trudności i ograniczenia nie odnoszą się do jego argumentu i toku rozumowania, a to za sprawą dialektycznego charakteru jego procedury. Zdaje się twierdzić, że jego argument nie tyle dowodzi ogólnej tezy, że umysł nie jest maszyną, co raczej dowodzi, że każda konkretna teza mechanicystyczna może zostać w skuteczny sposób odparta. Toteż jest on raczej wywodem czy dyskusją między dwiema osobami, a nie „[...] ciągiem dowodowym”[36] skonstruowanym przez jednego człowieka; inaczej mówiąc „[...] jest schematem obalania”[37] jakiejkolwiek poszczególnej wersji mechanicyzmu. Należy więc określić twierdzenie Lucasa nie jako tezę, ale jako kontrtezę, która jest używana każdorazowo, gdy ktoś poddaje w wątpliwość, że człowiek nie jest maszyną. Traci charakter „autonomicznego” dowodu unieważniającego wszystkie potencjalne tezy mechanicystyczne. Staje się orężem do dławienia każdego poszczególnego zamachu na ludzką godność i wolność[38]. Wszystko więc zależy od tego czy mechanicysta uczyni pierwszy krok i przedłoży swoją tezę do sprawdzenia[39]. Uczynienie pierwszego kroku polegałoby na wskazaniu przez niego maszyny, którą uważa za model równoważny umysłowi. Wtedy w odpowiedzi zadajemy pytanie, czy ta maszyna (czyli odpowiadająca jej teoria formalna) może dowieść zdania Gödla dla siebie, czyli przedłożyć to zdanie w wyróżniony sposób. Jeżeli mechanicysta odpowiada, że tak, to jego maszyna jest sprzeczna; jeżeli zaś odpowiada, że nie, to jest niesprzeczna, ale wówczas zdanie G dla jej teorii jest prawdziwe, a więc maszynę można wygödlować.

Można więc sądzić, że według Lucasa dialektyczność jego procedury czyni zarzut Putnama chybionym[40], dlatego, że zwalnia z obowiązku poznania niesprzeczności dowolnej teorii, a wymaga

---

[36] J. R. L u c a s: *Human and Machine Logic: A Rejoinder*. http://users.ox.ac.uk (dostęp dnia: 15.01.2010), s. 1.
[37] Tamże.
[38] Lucas na końcu swojego artykułu zdradza, że traktuje mechanicyzm jako jedno ze źródeł odmawiania człowiekowi moralnej podmiotowości i negacji istnienia wolności człowieka. Zob. J. R. L u c a s: *Umysły...*, s. 116-117.
[39] Zob. J. R. L u c a s: *Human...*, s. 1.
[40] Zob. S. K r a j e w s k i: *Twierdzenie Gödla i jego interpretacje filozoficzne...*, s. 128.

takiej znajomości jedynie w odniesieniu do konkretnego systemu przedstawianego przez mechanicystę. Stanowisko to nie wytrzymuje jednak próby krytyki, gdyż posiada spotykany wcześniej w wywodach Lucasa mankament. Prowadzi mianowicie do błędnego koła. Uznajemy bowiem, że mechanicysta potrafi wskazać taką maszynę, która nie dowodzi zdania Gödla; czyli potrafi rozstrzygnąć, która maszyna (teoria) jest niesprzeczna. Widzieliśmy wyżej, że stwierdzenie niesprzeczności należy raczej do domeny intuicji czy dobrze ugruntowanego domysłu. Ponadto zbiór niesprzecznych maszyn odpowiadających maszynom Turinga nie jest zbiorem rekurencyjnym[41], a więc nie ma mechanicznej procedury dającej zawsze poprawną odpowiedź czy dana maszyna jest czy nie jest niesprzeczna. Nasze żądanie wobec mechanicysty, aby wykazał się wiedzą na temat niesprzeczności maszyny (bo do tego sprowadza się wiedza czy maszyna dowodzi czy nie) ma więc albo charakter zaporowy (ale wtedy dialektyka Lucasa umiera w zarodku) albo jest uznaniem, że mechanicysta ma nadzwyczajne, niemechaniczne zdolności oparte na intuicji. Wtedy jednak popadamy w błędne koło, gdyż zakładamy coś czego chcieliśmy dowieść, czyli w zasadzie niemechaniczność naszego adwersarza. Wniosek jest taki, że dialektyczny charakter argumentu nie tylko nie unieważnia zarzutu Putnama, ale wręcz uwypukla problem wiedzy o niesprzeczności maszyny.

Wobec dialektyczności argumentu Lucasa można wysunąć jeszcze inny zarzut. Chodzi mianowicie o algorytmiczny charakter procedury. Dobrą ilustracją problemu może być zaproponowana przez S. Krajewskiego analogia z podawaniem większej liczby od podanej przez adwersarza: daj mi liczbę, a ja podam większą[42]. Do tego przecież można sprowadzić dialektyczność argumentu Lucasa: pokaż mi maszynę (a precyzyjniej: daj mi jej kod), a ja ją wygödluje. W jednym i w drugim przypadku mamy do czynienia z mechanicznością postępowania. Wystarczy poznać kod maszyny, czyli jej specyfikację, na którą składa się zbiór instrukcji, by sformułować zdanie Gödla, którego maszyna, o ile jest niesprzeczna, nie przedłoży w sposób wyróżniony[43]. Ale szkopuł w tym, że poznanie specyfikacji maszyny, potrzebne do przedłożenia dla niej formuły Gödla i do jej wygödlowania, może być także udziałem innej maszyny. Wówczas

[41] Zob. tamże, s. 129-130.
[42] Tamże, s. 130.
[43] Niektórzy zwracają uwagę na to, że nie ma algorytmicznej metody, która mówi jak zastosować zdanie Gödla do wszystkich możliwych rodzajów systemów formalnych i wyciągają stąd wniosek, że istnieją obiektywne i nieprzekraczalne dla człowieka granice wygödlowania maszyny. Ten zarzut można jednak odeprzeć odwołując się właśnie do dialektycznego charakteru procedury. Ani Lucas ani mechanicysta nie muszą mieć efektywnej procedury ustalania kodu dla wszystkich maszyn, żeby przedstawić kod konkretnej maszyny traktowanej jako model potencjalnie równoważny ludzkiemu umysłowi. Dialektyczna procedura ma jeszcze inną godną wyróżnienia zaletę. Neutralizuje wszelkie wysiłki mechanicysty polegające na „odgödlowywaniu" maszyny, czyli dodawaniu do aksjomatów kolejnej maszyny formuły zastosowanej wobec wcześniejszej. Cokolwiek wymyśli mechanicysta i jak wiele formuł godlöwskich dopisze do teorii odpowiadającej maszynie, będzie można i dla niej sformułować zdanie Gödla.

nasza przewaga nad nią (tak jak moja przewaga nad kimś z kim bawię się w podawanie większej liczby) brałaby się tylko z tego, że jesteśmy innym obiektami, co nie znaczy, że nie jesteśmy maszynami.

Antymechanicysta w rodzaju Lucasa wpadł więc w potrzask, przez siebie mimowolnie zastawiony: swoją argumentację uczynił tak efektywną, że mechanizowalną, a więc dającą się symulować na maszynie. Oczywiście to nie dowodzi, że jesteśmy maszyną. Bo nawet jeżeli przyznajemy, że maszyna jest w stanie naśladować procedurę Lucasa to nie znaczy, że jest jemu równoważna. Z tego, że maszyna może symulować dowolny fragment naszego umysłu nie wynika logicznie, że maszyna może symulować każdy fragment naszego umysłu[44]. Ale skoro ta procedura jest filarem rozumowania antymechanicystycznego, to zwrócenie uwagi na jej mechaniczny charakter, osłabia siłą rzeczy jej wydźwięk w walce z mechanicyzmem. Obroną Lucasa przeciw zarzutowi o algorytmiczność argumentu, a w konsekwencji przeciw sugestii, że skoro jest to mechaniczna procedura, to może ją także w zasadzie wykonać maszyna, jest wskazanie, że coś zawsze pozostaje niesformalizowane. Taki wniosek filozoficzny możemy wyprowadzić z twierdzeń Gödla. Ale wtedy możemy zasadnie zapytać po co to całe zamieszanie z wygödlowaniem maszyny? Skoro już na wstępie możemy stwierdzić, że zawsze musimy odwołać się do argumentów i dowodów nieformalnych, intuicyjnych czy niemechanicznych. Wydaje się więc, że Lucas popada w błędne koło po raz kolejny.

VII. Odrzucenie argumentu Lucasa

Z rozważań dotychczasowych wynika, że Lucasa argument obarczony jest wieloma wadami. Tymi mniejszej wagi, które można skorygować bez uszczerbku dla istoty argumentu: ograniczenie się tylko do maszyn Turinga i pozostawienie w sferze fantazji możliwości, że istnieją bądź będą istniały inne maszyny oraz błąd ekwiwokacji, który jednak da się obejść. I większego ciężaru gatunkowego: wykonanie procedury Lucasa albo prowadzi do błędnego koła i przypisania mechanicyście tego, co chcemy dowieść albo staje się możliwe dla maszyny, a w konsekwencji nasza przewaga nad nią topnieje drastycznie. Trudno więc oprzeć się wrażeniu, że Lucas drobi w miejscu. Jest zdolny jedynie pokazać, że maszyny (odpowiadające im systemy formalne) są niezupełne i wymykają się im pewne prawdy arytmetyczne. Nie mówi zatem nic więcej niż

[44] Brak logicznego wynikania w tym przypadku dobrze obrazuje przykład z liczbami: dla każdej liczby można podać liczbę większą, ale nie można podać liczby większej od wszystkich liczb.

w punkcie wyjścia, a więc nic ponad to, co wynika z treści pierwszego twierdzenia Gödla.

Zobaczmy jednak na koniec czy nic nam nie umknęło. Poddajmy krótkiemu sprawdzianowi wywód Lucasa i przyjmijmy, że wyklucza możliwość, że jesteśmy maszynami (jeżeli argument Lucasa jest poprawny, to nie jesteśmy maszynami). Jeżeli założymy, że jesteśmy maszynami, to powinniśmy uzyskać sprzeczność, czyli natrafić na takie elementy bezsprzecznie prawidłowego rozumowania Lucasa, które zaprzeczają konsekwencjom wyprowadzonym z tezy, że jesteśmy maszynami. Maszynami możemy być sprzecznymi albo niesprzecznymi. Zacznijmy od pierwszej możliwości i załóżmy, że jesteśmy nie tylko maszynami, ale na dodatek tymi pośledniejszego gatunku.

Kiedy maszyna jest sprzeczna? Gdy przedkłada zdanie Gödla jako prawdziwe. Czy nasza przewaga nad maszynami niesprzecznymi nie polega właśnie na tym, że potrafimy uznać prawdziwość zdania Gödla? Czy nie jest to wyraz daleko posuniętego podobieństwa do maszyn sprzecznych? Nie. Różnica jest taka, że my widzimy prawdziwość bez popadania w sprzeczność, co jest możliwe tylko spoza systemu, gdyż tylko stąd widać niesprzeczność systemu. Dowód „wewnątrzsystemowy” prawdziwości zdania Gödla jest niemożliwy bez założenia niesprzeczności, gdyż prowadzi nieuchronnie do sprzeczności. A maszyna nie może wyjść poza siebie, czyli poza swój system. Cała nasza przewaga sprowadza się więc do wiedzy o niesprzeczności systemu. Ale skąd pewność, że system jest niesprzeczny? Pewności nie ma. Więc można dopuścić możliwość, że zdanie Gödla jest fałszywe, bo jest ono prawdziwe tylko przy założeniu, że system jest niesprzeczny. A skoro zdanie Gödla może być fałszywe, to systemy, które widzą jego prawdziwość, mogą być sprzeczne, a więc nasza zdolność do widzenia prawdziwości, może brać się z naszej sprzeczności.

Popadamy w sprzeczność pod warunkiem, że Lucas dowiódł, że nie jesteśmy sprzeczni. A wiemy, że nie uporał się w skuteczny sposób z taką możliwością – podał psychologiczno-społeczne argumenty za jej odrzuceniem. Ale to za mało, by mieć pewność – ukrytej głęboko sprzeczności nie sposób wykluczyć, a skoro tak to nasze dostrzeganie prawdziwości zdania Gödla może być jej wyrazem, jak w przypadku maszyn sprzecznych. Taki wariant nie jest zupełnie pesymistyczny: jeżeli jesteśmy sprzeczni, twierdzenie Gödla nas nie ogranicza, czyli „hulaj duszo, piekła nie ma”, chyba, że genialnym matematykiem jest sam diabeł, który wbrew Kartezjuszowi wprowadza nas w błąd.

Wywód Lucasa nie może wykluczyć, że jesteśmy sprzecznymi maszynami, ale może mógłby wykluczyć, że jesteśmy maszynami niesprzecznymi? Jeżeli znajdziemy choć jeden przypadek, w którym maszyna jest niesprzeczna, jest człowiekowi równoważna i nie prowadzi to do zupełności jej systemu, będziemy mogli uznać, że implikacja: jeżeli niezupełność to nie jesteśmy maszynami niesprzecznymi, nie zachodzi, a więc pośrednio będziemy mogli uznać, że wywód Lucasa nie dowodzi, że nie jesteśmy niesprzecznymi maszynami. Wykorzystam tu hipotezę sformułowaną przez samego K. Gödla[45].

Rozróżnijmy matematykę obiektywną od subiektywnej. Ta pierwsza jest ogółem zdań prawdziwych, zaś druga jest ogółem zdań dowodliwych, takich które człowiek może dowolnymi metodami dowieść. Pierwsza matematyka nie może być ujęta przez żaden rekurencyjny system aksjomatów, co wiemy dzięki twierdzeniu Gödla. Nie może więc być przedłożona przez maszynę Turinga. Pytanie zasadnicze dotyczy więc matematyki subiektywnej. Czy maszyna może wyprodukować wszystkie twierdzenia dowodliwe? Nie można tego wykluczyć, ponieważ ogół zdań dowodliwych jest ujmowalny w systemie rekurencyjnym. Znaczyłoby to ni mniej ni więcej, tylko że maszyna mogłaby wygenerować dokładnie te zdania, które może udowodnić człowiek; inaczej mówiąc, procedura generowałaby wszystkie zdania, które człowiek może uznać za prawdziwe. Ale... człowiek nie byłby w stanie stwierdzić z całą pewnością, że tak właśnie jest! Byłoby tak dlatego, że taka maszyna byłaby mu równoważna w zakresie mocy matematycznych, a więc nie mógłby poza nią (poza siebie) wyjść i stwierdzić, że jest niesprzeczna. Mógłby jej niesprzeczność wykazać, jedynie w ramach systemu, ale to jest niemożliwe za sprawą drugiego twierdzenia Gödla. Rzecz jasna nie pomogłaby nawet pełna znajomość kodu maszyny, czyli wszystkich jej aksjomatów, gdyż stwierdzenie ich prawdziwości, oznacza stwierdzenie, że nie można z nich wyprowadzić zdań sprzecznych, a to cofa nas do poprzedniego zdania.

Co może teraz Lucas powiedzieć? Że pomijamy fakt, iż ludzie jednak dostrzegają prawdziwość zdania Gödla, a tego faktu nie można pogodzić z tym co zostało powiedziane wyżej, gdyż w ramach matematyki subiektywnej, zdanie Gödla musiałoby zostać odrzucone jako prowadzące do sprzeczności. To zastrzeżenie można jednak harmonijnie pogodzić z powyższą wizją. Dostrzegamy prawdziwość zdania Gödla, ponieważ nie znamy swojego kodu, pełnego opisu swojego funkcjonowania. Innymi słowy, znając tylko część systemu aksjomatycznego, możemy się wobec niego zdystansować i założyć jego niesprzeczność, dzięki czemu dostrzegamy prawdziwość zdania Gödla. Natomiast nie można byłoby pogodzić ze sobą znajomości swojego kodu i dostrzegania prawdziwości zdania Gödla. To jest wykluczone przez pierwsze twierdzenie Gödla. Czy Lucas dowiódł, że mamy absolutną wiedzę na swój temat? Dla odpowiedzi na to pytanie nie

[45] Zob. więcej S. K r a j e w s k i: *Twierdzenie Gödla i jego interpretacje filozoficzne...*, s. 161-166.

ma nawet potrzeby zaglądania do jego tekstów.

Wniosek jaki mógłby nasunąć się z powyższego wywodu mógłby stwierdzać, że nie trzeba znać twierdzeń Gödla (szerzej twierdzeń metalogicznych) by dowodzić czy wykazywać wyższości umysłu nad maszyną. Wniosek ten nie jest jednak do końca uprawniony, gdyż twierdzenia Gödla mogą wzmocnić tezę o niemechanicznej naturze ludzkiego umysłu, jeżeli zostaną połączone z pewnymi dodatkowymi założeniami[46]. Natomiast w pojedynkę dowieść tej tezy nie mogą. To co nie ulega wątpliwości, to twierdzenie, że ich znajomość jest niezbędna żeby to zrozumieć i żeby mieć narzędzia do odparcia argumentów w stylu Lucasa.

[46] Tamże.