https://doi.org/10.18778/1689-4286.09.06

HYBRIS nr 9 (2009)
ISSN: 1689-4286

John R. Lucas

Umysły, maszyny i Gödel: retrospekcja [*]

Na wstępie muszę się wytłumaczyć. Mój pierwotny tekst, "Umysły, maszyny i Gödel", napisany został w odpowiedzi na opublikowany w *Mind* tekst Turinga z 1950 roku i miał za zadanie pokazać, że umysły nie są maszynami Turinga. Dlaczego więc nie sformułowałem swojego argumentu, odwołując się do twierdzenia Turinga, które jest dość proste do udowodnienia i bezpośrednio stosowalne do maszyn Turinga, zamiast do twierdzenia Gödla, które jest niezwykle trudne do udowodnienia i nie stosuje się do maszyn w sposób tak naturalny i oczywisty? Powodem było to, że twierdzenie Gödla dało mi coś więcej: postawiło pytania o prawdę, które niewątpliwie wpisane są w naturę człowieka, zaś twierdzenie Turinga – nie. Nie tylko pokazuje ono [twierdzenie Gödla – MZ], że poprawnie zbudowana formuła jest niedowodliwa-w-systemie, lecz także, że jest prawdziwa. Pokazuje coś na temat rozumowania – to mianowicie, że nie jest w zupełności związane regułami, tak więc my, którzy jesteśmy racjonalni, możemy wyjść poza reguły każdego poszczególnego systemu logicznego i skonstruować poprawnie zbudowaną formułę gödlowską nie będącą jedynie ciągiem symboli, lecz zdaniem, które jest prawdziwe. Twierdzenie Turinga można łatwo zastosować do komputera, co do którego ktoś utrzymuje, że reprezentuje ludzki umysł, lecz nie jest tak oczywiste, że to, czego komputer nie jest w stanie zrobić, umysł będzie w stanie. Jest jednak całkowicie oczywiste, że posiadamy pojęcie prawdy. Nawet jeśli, jak zostało to przedstawione w poprzednim tekście, nie jest to *summum bonum*, jest to *bonum* i jedna z charakterystycznych cech, po których poznaje się umysł. Reprezentacja ludzkiego umysłu, która nie mogłaby wziąć pod rozwagę prawdy, byłaby z natury nieprzekonująca. Choć z twierdzenia Turinga, wynika ten sam negatywny wniosek, co z twierdzenia Gödla, tj. że nawet wyidealizowane komputery pewnych rzeczy zrobić nie

[*] Tekst odczytany przez Johna Randolpha Lucasa na Turing Conference w Brighton 6 kwietnia 1990 r.; opublikowany w P. Millican, A. Clark (eds.), *Machines and Thought*, Oxford 1996 (przypisy oznaczoen gwiazdką pochodzą od tłumacza).

potrafią, to jednak nie wynika z niego wniosek pozytywny, że my, o ile jesteśmy racjonalnymi podmiotami, możemy zrobić tę właśnie rzecz, której komputer nie potrafi. Mimo to, czasami zastanawiałem się, czy nie mógłbym skonstruować podobnego argumentu opartego na twierdzeniu Turinga i w tym celu rozważałem ideę maszyny von Neumanna. Maszyna von Neumanna to czarna skrzynka, wewnątrz której schowany jest John von Neumann. Choć jednak zasadnym było przyznać na indukcyjnych podstawach maszynie von Neumanna moc rozwiązywania wszelkich problemów w skończonym czasie – takim mniej więcej, jaki potrzebny jest, by dostać się pociągiem z Nowego Jorku do Chicago – to jednak nie miała ona tej wymowy, co dowód Gödla jego pierwszego twierdzenia o niezupełności. Pozostawiam zatem uczestnikom tej konferencji dalsze przemyślenia na temat tego, w jaki sposób twierdzenie Turinga odnosi się do mechanicyzmu oraz czy maszyna Turinga mogłaby przekonująco reprezentować umysł, i wracam do argumentu, który rzeczywiście wysunąłem.

Dowodziłem, że twierdzenie Gödla pozwoliło nam opracować schemat obalania rozmaitych mechanicystycznych teorii umysłu, które mogą być wysuwane. Twierdzenie Gödla jest wyszukaną formą paradosku Kreteńczyka, przedstawionego przez Epimenidesa. Gödel pokazał, w jaki sposób możemy znaleźć reprezentację każdej rozsądnej teorii matematycznej w obrębie niej samej. Podczas gdy oryginalny paradoks Kreteńczyka, “To zdanie jest nieprawdziwe”, może zostać odrzucony dlatego, że jest brutalnie samozwrotny, a my nie wiemy, co to za zdanie, o którym twierdzi się, że jest fałszywe, dopóki nie zostanie wypowiedziane, a nie możemy go wypowiedzieć, dopóki nie wiemy, czym jest to, o czym twierdzi się, że jest fałszywe, Gödel oddala ten zarzut. Aby jednak to uczynić, musi nie tylko znaleźć w obrębie swojej teorii matematycznej reprezentację pewnych środków, za pomocą których odniesiemy się do zdania, lecz także pewnych środków, za pomocą których matematycznie wyrazimy to, o czym mówimy. Nie możemy tego zrobić przy użyciu prostego “prawdziwe” lub “nieprawdziwe”. Gdybyśmy bowiem mogli, prowadziłoby to bezpośrednio do sprzeczności. Tym, co Gödel mógł uczynić, było wyrażenie w obrębie swojego systemu

pojęcia bycia dowodliwym- i – co za tym idzie – niedowodliwym-w-tym-systemie. Wygenerował solidną, poprawnie zbudowaną formułę, której treść mogłaby zostać zinterpretowana jako: “Ta poprawnie-zbudowana formuła jest niedowodliwa-w-tym-systemie”. Z tego wynika, że musi być ona zarówno niedowodliwa-w-systemie, jak i prawdziwa. Gdyby bowiem była dowodliwa i pod warunkiem że system byłby solidny – taki, w którym jedynie poprawnie zbudowane formuły, wyrażające zdania prawdziwe, mogą być dowiedzione, wówczas byłaby prawdziwa. To zatem, co głosi, czyli że jest niedowodliwa-w-systemie, obowiązywałoby. Byłaby więc niedowodliwa-w-systemie. Nie może więc być dowodliwa-w-systemie. Lecz jeśli jest niedowodliwa-w-systemie, to rzeczy mają się tak, jak głosi, że się mają, a zatem jest prawdziwa. Jest więc prawdziwa, lecz niedowodliwa-w-systemie. Wyglądało na to, że twierdzenie Gödla było nie tylko zaskakującym wynikiem w matematyce, lecz miało również związek z teoriami umysłu, w szczególności z mechanicyzmem, który, jak wskazał przed dwoma dniami profesor Clark Glymour[**], jest tak naturalnym założeniem naszych czasów, jak do końca ubiegłego [tj. XIX – MZ] stulecia był nim klasyczny materializm w postaci wyrażonej przez Tyndale'a. Mechanicyzm utrzymuje, że działania umysłu mogą być w całości rozumiane jako działania ściśle określonego skończonego systemu, wykonującego operacje zgodnie ze ściśle określonymi deterministycznymi prawami. Entuzjaści Sztucznej Inteligencji często zaliczają się do mechanicystów i mają skłonność, by twierdzić, że pewnego dnia będą w stanie symulować wszelkie formy inteligentnego zachowania za pomocą wystarczająco skomplikowanego komputera wyposażonego w wystarczająco zaawansowane oprogramowanie. Operacje każdego takiego komputera mogłyby jednak być reprezentowane w języku pewnego formalnego rachunku logicznego ze ściśle określoną, skończoną (choć niewyobrażalnie wielką) liczbą możliwych poprawnie zbudowanych formuł i ściśle określoną, skończoną (choć prawdopodobnie mniejszą) liczbą aksjomatów i reguł inferencji. Formuła gödlowska takiego systemu byłaby tą, której komputer, wraz ze swoim oprogramowaniem, nie byłby w stanie dowieść. My jednak moglibyśmy. Zatem twierdzenie, że z zasady

[**] Lucas ma na myśli wystąpienie Clarka Glymoura, profesora uniwersytetu Carnegie-Mellon, zatytułowane *Computability, conceptual revolutions and the logic of Discovery*.

możliwe jest, by komputer symulował wszystkie nasze zachowania, załamuje się w tym jednym, kluczowym punkcie.

Argument, który wysunąłem, jest dwupoziomowy. Nie oferuję prostego, nokautującego dowodu na to, że umysły są z natury lepsze niż maszyny, lecz schemat konstrukcji kontrdowodu dla każdej brzmiącej wiarygodnie mechanicystycznej tezy, którą można by przedstawić. Sam kontrdowód zależy od konkretnej mechanicystycznej tezy, którą się utrzymuje, i nie ma wykazywać, że umysł jest w całej rozciągłości lepszy niż jego domniemana mechanicystyczna reprezentacja, a jedynie to, iż jest lepszy pod jednym względem – a przez to inny. To wystarcza, by obalić tę konkretną mechanicystyczną tezę. Oczywiście, wzięty z osobna, argument ten pozostawia wszystkie inne [tezy – MZ] nieobalone, a mechanicyście – swobodę, by wysunąć tezę alternatywną, do której skonstruowany przeze mnie kontrargument nie będzie bezpośrednio stosowalny. Twierdzę jednak, że możliwe będzie takie jego dopasowanie, by dawał odpór nowej tezie. Uchwyciwszy raz sens argumentu Gödla, umysł może go właściwie adaptować, by odpierać każdą tezę po kolei, a przecież wszystkie one oznajmiają, że umysł jest w swej istocie jakąś formą maszyny Turinga. Zasadniczo zatem, dwie części mojego argumentu to: po pierwsze, mocny argument negatywny, adresowany do mechanicysty, który wysuwa konkretne twierdzenie i wykazujący mu, za pomocą środków, które musi uznać za poprawne, że jego twierdzenie jest nie do obrony, i po drugie: uogólniony argument pozytywny, adresowany do ludzi inteligentnych – tak biernych obserwatorów, jak i samych mechanicystów opowiadających się za poszczególnymi wersjami mechanicyzmu – w celu uzmysłowienia im, że pewien typ argumentu, skrojony według powyższego wzorca, zawsze okaże się radzić sobie z każdą kolejną wersją mechanicyzmu, którą tylko można wymyślić.

Odczytałem swój tekst [“Umysły, maszyny i Gödel”] na posiedzeniu Oksfordzkiego Towarzystwa Filozoficznego w 1959 roku, następnie opublikowałem go w *Philosophy*[1],

1 J. Lucas, *Minds, Machines and Gödel,* “Philosophy”, vol. XXXVI (1961), ss. 112-127; przedruk w: K. M.

później zaś przedstawiłem swój argument bardziej szczegółowo w *The Freedom of the Will*[2]. Byłem zań wielokrotnie atakowany. Choć argumentowałem, jak miałem nadzieję, ze skromnością i – w pewnym stopniu – wstrzemięźliwością, wiele polemik pozbawionych było zarówno uprzejmości, jak i rozwagi. Musiałem trafić w czuły punkt. To oczywiście nie dowodzi, że miałem rację. W zasadzie muszę przyznać, że jest całkiem prawdopodobne, że w jakiejś części jej nie miałem i że inni będą w stanie wyłuszczyć swe argumenty jaśniej, a przez to bardziej przekonująco, niż ja to uczyniłem. Jestem jednak coraz silniej przekonany, że nie byłem pozbawiony racji w zupełności, a to ze względu na bardzo duże rozbieżności, jakie pojawiły się pośród moich krytyków odnośnie do tego, w którym dokładnie miejscu moje argumenty zawodzą. Każdy wskazuje na inny punkt, uznając, że te, które wzbudzają zastrzeżenia pozostałych krytyków, są w gruncie rzeczy w porządku i mając nadzieję, że to właśnie jego strzał okaże się śmiertelny. Żaden się nie okazał, o ile mi wiadomo. Próbowałem odpowiadać na każdy zarzut uczciwie i wyczerpująco, lecz ciało stawało się coraz słabsze. Często dochodziłem do wniosku, że dany krytyk nie krytykuje żadnego argumentu, który naprawdę wysunąłem, lecz taki, który chciałby, abym wysunął, nawet jeśli w rzeczywistości usilnie starałem się go pomijać. W ostatnich latach nie broniłem się już tak zaciekle i wiele artykułów w zupełności mi umknęło[3]. Być może pojawiły się jakieś nowe, rozstrzygające zarzuty, które w zupełności przeoczyłem. Te jednak, z którymi do tej pory się zetknąłem, dalekie były od statusu rozstrzygających.

Rozważyć każdy z zarzutów po kolei byłoby zadaniem nazbyt rozwlekłym, by podejmować je tutaj. Wskażę więc pięć powtarzających się motywów. Niektóre z zarzutów kwestionują idealizację przemyconą w sposobie, w jaki zorganizowałem pojedynek pomiędzy umysłem i maszyną; niektóre podnoszą kwestię modalności i skończoności; niektóre poruszają zagadnienie pozaskończonej arytmetyki; niektóre stawiają pytanie o stopień, w jakim racjonalne wnioskowania powinny być

Sayre, F. J. Crosson (red.), *The Modeling of Mind*, Indianapolis 1963, ss. 255-271 oraz A. R. Anderson (red.), *Minds and Machines*, Engelwood Cliffs 1954, ss. 43-59.

2 Tenże, *Freedom of the Will*, Oxford 1970.

3 Na końcu podaję listę najważniejszych krytyk, na które się natknąłem.

formalizowalne; niektóre zaś mówią o niesprzeczności.

Wielu filozofów kwestionuje idealizację przemyconą w gödlowskim argumencie. Kontekst rozciągnięty jest pomiędzy "umysłem" a "maszyną", lecz jest to wyidealizowany umysł i wyidealizowana maszyna. Rzeczywiste umysły zamknięte są w śmiertelnych ciałach, rzeczywiste maszyny często działają wadliwie lub się zużywają. Jako że prawdziwe maszyny nie są maszynami Turinga, ponieważ nie posiadają nieskończonej taśmy – to znaczy nieskończonej pamięci – można utrzymywać, że automatycznie nie mogą podlegać gödlowskim ograniczeniom. Twierdzenie Gödla nie stosuje się jednak tylko do arytmetyki Peano, z jej infinitystycznym postulatem rekurencyjnych rozumowań, lecz także do słabszej arytmetyki Q Robinsona, która jest nieskończona jedynie potencjalnie, nie aktualnie, i ledwie wykracza poza skalę prawdopodobnego rozwoju komputerów. W każdym razie, ograniczenia skończoności raczej zmniejszają niż zwiększają prawdopodobieństwo, że jakiś komputer będzie adekwatną reprezentacją umysłu. Rzeczywiste umysły zamknięte są w śmiertelnych ciałach. W krótkim okresie naszego życia nie jesteśmy w stanie osiągnąć nazbyt wiele i możemy mieć zbyt mało czasu lub inteligencji, by ustalić naszą formułę gödlowską. Hanson wskazuje, że mogłoby istnieć twierdzenie z zakresu elementarnej teorii liczb, którego nie byłbym w stanie udowodnić, gdyż dowód byłby zbyt długi lub złożony, bym mógł go przedłożyć[4]. Dowolna maszyna, która reprezentowałaby umysł, byłaby niezmiernie skomplikowana, a wyliczenie jej formuły gödlowskiej mogłoby leżeć poza możliwościami każdego ludzkiego matematyka[5]. Ale przecież można by mu pomóc. Inni matematycy mogliby przyjść w sukurs, przeczuwając, że również mają interes w tym, by skonfundować mechanicznego Goliata[6]. Prawda zdania gödlowskiego

4 W. Hanson, *Mechanism and Gödel Theorems*, "British Journal for the Philosophy of Science", vol. XXII (1971), s. 12; por. D. R. Hofstadter, *Gödel, Escher, Bach*, New York 1979, s. 475.

5 R. Rucker, *Gödel's Theorem: The Paradox at the heart of modern*, "Popular Computing", 2 (1985), s. 168.

6 Tę uwagę zawdzięczam Michaelowi A. E. Dummettowi, który poczynił ją na spotkaniu Oksfordzkiego Towarzystwa Filozoficznego 30 października 1959. Podobna uwaga jest *implicite* zawarta w: H. Wang, *From Mathematics to Philosophy*, London 1974, s. 316.

zinterpretowanego w zwyczajnej nieformalnej arytmetyce, jest prawdą matematyczną, która, nawet jeśli wskazana przez innych matematyków, nie zależy od ich świadectwa w taki sposób, w jaki zależą odeń przypadkowe stwierdzenia. Zatem nawet wspomagany wskazówkami innych matematyków, umysł oznajmiający prawdziwość zdania gödlowskiego byłby rzeczywistą podstawą dla rozróżnienia go od maszyny.

Niektórzy krytycy argumentu gödlowskiego – Dennett, Hofstadter i Kirk – narzekają, że jestem niewystarczająco wrażliwy na zaawansowanie współczesnej technologii komputerowej i że [w moje rozumowanie – MZ] wkradła się fatalna w skutkach dwuznaczność, nie pozwalająca odróżnić podstawowego poziomu operacji danej maszyny i poziomu wejścia i wyjścia, który to ma reprezentować umysł. We współczesnym żargonie: kodu maszyny i języka programowania, jak np. PROLOG. Choć jednak istnieje różnica poziomów, to nie unieważnia to argumentu. Kompilator jest całkowicie deterministyczny. Dowolny ciąg operacji określony w kodzie maszyny, może być jednoznacznie określony w języku programowania i *vice versa*. Stąd też całkiem słusznym jest scharakteryzować zdolności maszyny mechanicysty w terminach języka wyższego poziomu. By mogła ona zacząć być reprezentacją umysłu, musi być w stanie generować prostą arytmetykę. I wówczas, na tym poziomie, twierdzenie Gödla jest stosowalne. Ten sam kontrargument odnosi się do zarzutu Dennetta, iż porównanie ludzi i maszyn Turinga jest wysoce nieintuicyjne, ponieważ niewiele mamy okazji, by spacerować w kółko i wypowiadać mętne prawdy zwykłej nieformalnej arytmetyki. Niewielu z nas jest w stanie oznajmić zdanie gödlowskie, jeszcze mniejsza ilość chce to zrobić. "Ludzie nie siedzą bezczynnie, wypowiadając twierdzenia przy użyciu uniwersalnego słownika, lecz mówią rzeczy poważnie i żartobliwie, popełniają lapsusy, mówią kilkoma językami, sygnalizują zgodę poprzez skinienie głową lub inne niewerbalne działanie i – co w tym kontekście najbardziej kłopotliwe – głoszą wszelkiego rodzaju nonsensy i sprzeczności, zarówno rozmyślnie, jak i niechcący."[7] Oczywiście, pod tym względem ludzie są zupełnie niepodobni do maszyn i wielu filozofów odrzuca twierdzenia mechanicyzmu wyłącznie na tej podstawie.

7 D. C. Dennett, recenzja *The Freedom of the Will*, "Journal of Philosophy", vol. LXIX, 17 (1972), s. 530.

Mechanicyści odpowiadają jednak, że nazbyt pochopnie. Człowiek – mówią – jest bardzo skomplikowaną maszyną, skomplikowaną na tyle, by przedkładać na wyjściu wszystkie te niepodobne do maszynowych dane. Możemy uznać ich twierdzenie za wysoce nieintuicyjne, lecz nie powinniśmy z góry go odrzucać. Dlatego też traktuję poważnie, choć jedynie po to, by je obalić, twierdzenie, że można by zbudować maszynę, która reprezentowałaby ludzkie zachowanie. Gdyby tak było, musiałaby, pośród innych rzeczy, reprezentować zachowanie ludzkieg umysłu. Niektórzy ludzie – większość ludzi – są zdolni rozpoznać wiele podstawowych prawd arytmetyki i, w szczególności jeśli się ich poprosi (co interpretować można jako konkretną daną na wejściu), potrafią oznajmiać je jako prawdy. Choć "opis człowieka jako pewnego rodzaju maszyny dowodzącej twierdzeń"[8] byłby mniej niż wyczerpujący, to jednak byłby kluczowym elementem opisu maszyny, jeśli ta rzeczywiście miałaby być reprezentacją człowieka. Musiałaby być zdolna do tego, by zawrzeć na wyjściu to, co mogłoby być wzięte za stwierdzenia podstawowych prawd arytmetyki i do uznania za właściwe tych inferencji, które są dopuszczalne przez logikę pierwszego rzędu. To stanowi minimum. Oczywiście mogłaby być zdolna do robienia większej ilości rzeczy – w swej pamięci mogłaby przechowywać żarty do użytku w poobiednich pogawędkach albo osobiste wspomnienia do użytku przy innych okazjach – dopóki jednak na wyjściu – w odpowiedzi na pytania lub inne dane na wejściu – przedkładać będzie zbiór stwierdzeń zawierających elementarną teorię liczb, pozostanie marną reprezentacją niektórych ludzkich umysłów. Jeśli nie potrafi zdać matematyki na poziomie podstawowym, czy naprawdę skłonni będziemy wierzyć mechanicyście, kiedy stwierdzi, że jest ona reprezentacją absolwenta wyższej uczelni?

Rzeczywiste umysły są skończone w tym, co faktycznie osiągają. Wang i Boyer dostrzegają trudności związane z przypisywaniem nieskończonych możliwości umysłom, jeśli zestawić je z rzeczywistą skończonością ludzkiego życia. Boyer przyjmuje perspektywę *post mortem* i dostrzega, że wszystko, co rzeczywiście pojawiło się na wyjściu u Lucasa, Astaire'a i kogokolwiek innego, może być *ex post facto*

8 Tamże, s. 527.

reprezentowane przez maszynę[9]. Rzeczywiste osiągnięcia śmiertelnych ludzi są skończone i z tego względu symulowalne. Po mojej śmierci możliwe byłoby takie zaprogramowanie komputera z odpowiednimi możliwościami graficznymi, by pokazywał na ekranie film biograficzny o całym moim życiu. Po mojej śmierci jednak nie sztuką byłoby mnie przechytrzyć. Spór rozbija się o to, czy komputer może być kopią żyjącego mnie, gdy jeszcze nie zrobiłem tego wszystkiego, co mam do zrobienia i zrobić mogę jeszcze wiele rozmaitych rzeczy. To raczej o kwestię możliwości, nie zaś tego, co faktycznie jest, rozbija się spór. Wang zgadza się z tym i przyznaje, że mamy skłonność do mówienia, iż jest logicznie możliwe, by mieć umysł zdolny rozpoznawać każde prawdziwe zdanie z zakresu teorii liczb lub rozwiązać każdy zbiór problemów nierozwiązywalnych dla maszyny Turinga, ale życie jest krótkie[10]. Przy skończonej długości życia jedynie skończona liczba zdań może zostać rozpoznana, jedynie skończony zbiór problemów może zostać rozwiązany. A maszynę można zaprogramować, by to wykonała. Oczywiście, uważamy, że człowiek mógłby dużo więcej, lecz trudno uchwycić ten sens nieskończonych możliwości. To prawda. Trudno jest uchwycić sens nieskończonych możliwości. Jest to jednak kluczowy element naszego pojęcia umysłu i modalnie "płaskie" wyjaśnienie umysłu, uwzględniające jedynie to, czego dokonał, jest tak nieprzekonujące, jak wyjaśnienie przyczyny, które bierze wzgląd jedynie na stałe współwystępowanie, a nie na to, co by było, gdyby okoliczności były inne. Aby uchwycić ten sens możliwości, argument swój sformułowałem w kategoriach pojedynku, co czyni go otwartym na wszelkie próby odparcia, jakie tylko przeciwnik zażyczy sobie podjąć. W dwustronnych czy "dialektycznych" argumentach często udaje się zawrzeć pojęcia wymykające się eksplikacjom czysto monologicznym: dlatego też definicja typu epsilon-delta wielkości infinitezymalnych jest najbardziej komunikatywna, a bardziej ogólnie, każda alternacja kwantyfikatorów, jak w zasadach elementarnej arytmetyki zaproponowanych przez profesora Clarka Glymoura dla ostatecznej unifikacji teorii prawdy.

9 D. L. Boyer, *J. R. Lucas, Kurt Gödel and Fred Astaire*, "The Philosophical Quarterly", vol. XXXIII, 131 (1983), ss. 147-159.

10 H. Wang, dz. cyt., s. 315.

Choć pewien stopień idealizacji wydaje się dopuszczalny w rozważaniach nad umysłem nieskrępowanym moralnością i nad maszyną Turinga z nieskończoną taśmą, pozostają wątpliwości dotyczące tego, jak daleko wolno nam w tej nieskończoności pobłądzić. Pozaskończona arytmetyka leży u podstaw zarzutów Gooda i Hofstadtera. Problem bierze się ze sposobu, w jaki pojedynek pomiędzy umysłem i maszyną został zorganizowany. Celem pojedynku nie jest wykazanie, że umysł jest lepszy od maszyny, a jedynie, że jest inny niż ona; a to umysł osiąga poprzez gödlowanie maszyny. Bardzo naturalną odpowiedzią mechanicysty jest doączenie zdania gödlowskiego do maszyny, lecz oczywiście otrzymujemy w ten sposób inną maszynę z innym, jej własnym, zdaniem gödlowskim, którego ona nie może przedłożyć jako prawdziwego, umysł zaś może. Wówczas mechanicysta usiłuje dodać operator gödlujący, który w efekcie daje całą policzalną nieskończoność zdań gödlowskich. To jednak znów może zostać przebite przez umysł, który konstruuje zdanie gödlowskie owej nowej maszyny zawierającej operator gödlujący i wygödlowuje wszystko. Zasadniczo jest to ruch od ω – nieskończonego ciągu zdań gödlowskich generowanych przez operator gödlujący – do $\omega+1$ – kolejnej pozaskończonej liczby porządkowej. I tak to się toczy. Co pewien czas mechanicysta traci cierpliwość i dołącza do swojej maszyny następny operator, przeznaczony do tego, by za jednym zamachem wygenerować wszystkie zdania gödlowskie, którymi mentalista go przebija; to w efekcie daje nową graniczną liczbę porządkową. Choć jednak liczby porządkowe tego rodzaju nie mają poprzedników, to jednak mają następniki, jak wszystkie inne [liczby porządkowe – MZ], więc umysł może je wygödlować, generując zdanie gödlowskie dla nowej wersji maszyny i widząc jego prawdziwość, czego maszyna zrobić nie potrafi. Hofstadter sądzi, że – w świetle twierdzenia Churcha i Kleene'ego o formalnych definicjach pozaskończonych liczb porządkowych[11] – mentalista ma problem. Pokazali oni, że nie możemy zaprogramować maszyny tak, by nadała nazwy wszystkim liczbom porządkowym. Co pewien czas, gdy rozważymy wszystkie dotychczas nazwane liczby porządkowe i chcemy zawrzeć je w jednym zbiorze, którego możemy użyć do zdefiniowania nowego

11 D. R. Hofstadter, dz. cyt., s. 475.

rodzaju liczb porządkowych, przekraczającego wszystkie wcześniejsze, niezbędny jest jakiś nowy, pomysłowy krok. Hofstadter sądzi, że w świetle twierdzenia Churcha-Kleene'ego, umysł może opaść z sił i nie podołać wymyślaniu nowych liczb porządkowych jak to było wymagane i w ten sposób stracić ostatnią szansę na ustanowienie różnicy pomiędzy umysłem a jakąś maszyną. Takie podejście jest jednak błędne z dwóch powodów. Po pierwsze, pozostawia pytanie bez odpowiedzi, po drugie zaś, opacznie pojmuje istotę pojedynku.

Hofstadter zakłada, że umysł podlega tym samym ograniczeniom, co maszyna i że ponieważ nie istnieje mechaniczny sposób na nazwanie wszystkich liczb porządkowych, to umysł również nie jest w stanie tego zrobić. Dokładnie to jest jednak przedmiotem sporu. Sam Gödel odrzucił mechanicyzm w kontekście naszych zdolności do wymyślania nowych definicji dla pozaskończonych liczb porządkowych (i coraz mocniejszych aksjomatów teorii mnogości), a Wang się do tego skłania[12]. Przy tej okazji zasadnym jest zaznaczyć, że sam Turing był w tej sprawie tego samego zdania, co Gödel. Doszedł on “do logiki liczb porządowych, traktując ją jako drogę 'ucieczki' przed twierdzeniem Gödla o niezupełności”[13], lecz zorientował się, że “choć w czasach przedgödlowskich niektórzy sądzili, że możliwe będzie rozwinięcie tego programu do takiego stopnia, że... potrzeba intuicji zostanie całkowicie wyeliminowana”, to jednak w następstwie twierdzeń Gödla o niezupełności trzeba zamiast tego zwrócić się ku “niekonstruktywnym” systemom logiki, w których “nie wszystkie kroki dowodu są mechaniczne – niektóre są intuicyjne”. Turing przychyla się do tego, że kroki, w których rozpoznajemy formuły jako oznaczające liczby porządkowe, są intuicyjne i mówi dalej, że powinniśmy w miarę jasno pokazać, kiedy krok robi użytek z intuicji, a kiedy jest czysto formalny i że nacisk położony na intuicję powinien być minimalny[14]. Otwarcie,

12 H. Wang, dz. cyt., ss. 324-326.

13 S. Feferman, *Turing in the Land of O(z)*, [w:] R. Herken (ed.), *The Universal Turing Machine*, Oxford 1988, s. 121.

14 A. M. Turing, *Systems of logic based on ordinals*, “Proceedings of The London Mathematical Society”, vol. 2, 45 (1939), ss. 161-228; przedruk w: M. Davis (red.), *The Undecidable*, New York 1965; cytat w: S. Feferman, dz. cyt., s. 129.

jak Gödel, przyznaje, że zdolność umysłu do rozpoznawania nowych liczb porządkowych przewyższa zdolność każdego formalnego algorytmu, choć nie wyciąga tego, co Gödel, wniosku. Może być tak, że zdolność umysłu do rozpoznawania nowych liczb porządkowych jest sprawą, o którą powinna toczyć się walka. To właśnie utrzymywał Good[15] – choć sporom o notację dla liczb porządkowych brakuje ostrza gödlowskiego argumentu. Bez względu jednak na zasługi na innych polach bitwy, jasnym jest, że są one obszarami spornymi w tym samym konflikcie i niekwestionowane panowanie nad jednym nie może służyć za podstawę roszczeń do panowania nad innym.

W każdym razie Hofstadter opacznie pojmuje istotę pojedynku. Cały trud spoczywa na mechanicyście usiłującym zaprojektować maszynę, która nie mogłaby zostać wygödlowana. To mechanicysta jest tym, który ucieka się do ograniczania liczb porządkowych i który może mieć problemy z wymyśleniem nowych notacji dla nich. Umysł potrzebuje jedynie przejść do kolejnej [liczby porządkowej – MZ], co zawsze jest łatwym, bezproblemowym krokiem, i wygödlować to, mechanicysta akurat ostatnio zaoferował. Argument Hofstadtera, jak to często bywa, świadczy przeciw stanowisku, którego Hofstadter dowodzi i ukazuje słabość maszyn – nie ma powodów, by zakładać, że jest ona również udziałem umysłów, a w istocie problemu leży trudność dla tych, którzy usiłują wymknąć się z sideł argumentu gödlowskiego, nie zaś dla tych, którzy je zastawiają.

U podłoża argumentu Hofstadtera leży retoryczne pytanie, które zadaje wielu mechanicystów: “Skąd Lucas wie, że umysł może zrobić to, tamto lub coś innego?” Nie wystarczy – utrzymują – bym wyraził na ten temat opinię lub po prostu to stwierdził. Muszę to udowodnić. A jeśli to udowodnię, to ponieważ kroki mojego dowodu mogą zostać wprowadzone do maszyny, to i ona jest w stanie to uczynić. Good wykłada ten argument wprost:

15 I. J. Good, *Gödel's Theorem is a Red Herring*, “British Journal for the Philosophy of Science”, vol. XIX, 4 (1969), ss. 357-358.

> Tym, co musi udowodnić, jest to, że zawsze jest w stanie iść o krok dalej. Nie wystarczy być o tym przekonanym, gdyż przekonanie jest kwestią prawdopodobieństwa, a maszyny Turinga raczej nie posiadają zdolności wydawania sądów prawdopodobnych. Żaden taki dowód nie jest jednak możliwy, jako że – gdyby był dany – mógłby zostać użyty do zaprojektowania maszyny, która zawsze byłaby w stanie iść o krok dalej.

Tę samą uwagę czyni Webb w swej konsekwentnej i dogłębnej krytyce argumentu gödlowskiego:

> Jedynie dlatego, że Gödel wskazuje efektywny sposób konstrukcji zdania gödlowskiego, Lucas może czuć się pewnym, że potrafi znaleźć pięte achillesową każdej maszyny. Jeśli jednak Lucas jest w stanie efektywnie poradzić sobie z każdą maszyną, to musi istnieć maszyna, która również to zrobi[16]. [To] "jest podstawowy problem, wobec którego staje anty-mechanicyzm: gdy tylko konstrukcje użyte w argumentach na jego rzecz staną się wystarczająco efektywne, by dawały pewność (a każda efektywna dla człowieka procedura obliczeniowa może być symulowana przez maszynę Turinga), oznaczać to będzie, że maszyny mogą je symulować. W szczególności będzie to oznaczać, że nasze działanie, polegające na zastosowaniu argumentu Gödla do dowolnych maszyn w celu wykazania, że nie możemy być modelowani przez maszynę, w istocie może być modelowane przez maszynę. Odtąd każdy taki [tj. zakładający różnicę pomiędzy umysłem a maszyną – MZ] wniosek musi upaść, gdyż w przeciwnym wypadku musielibyśmy stwierdzić, że maszyna nie może być modelowana przez żadną maszynę! Krótko mówiąc, argumenty anty-mechanicystyczne muszą być albo nieefektywne, albo niezdolne do wykazania, że argumentujący nie jest maszyną.[17]

Jądrem tego argumentu jest założenie, że każdy nieformalny argument musi poddawać się formalizacji, albo – w przeciwnym wypadku – będzie nieważny. Takie założenie podkopuje rozróżnienie, które poczyniłem między dwoma sensami argumentu gödlowskiego: między argumentem negatywnym nawiązującym do dokładnej specyfikacji, którą miałaby spełnić zaprogramowana do tego maszyna i, z drugiej

16 J. C. Webb, *Mechanism, Mentalism and Metamathematics. An Essay on Finitism*, Dordrecht 1980, s. 230.

17 Tamże, s. 232, fragmenty pisane kursywą.

strony, pewnym stylem argumentowania, podobnym do oryginalnego argumentu Gödla jako źródła inspiracji, lecz nie do końca precyzyjnie określonym i z tego względu niemożliwym do wprowadzenia do maszyny, lecz możliwym do pojęcia i zastosowania przez inteligentny umysł. Prawdą jest, że nie możemy zaskorupiałemu mechanicyście dać dowodu na to, że potrafimy iść o krok dalej. Możemy jednak dojść do dobrze uzasadnionego przekonania, że to potrafimy, które da nam oraz niegdysiejszemu mechanicyście, o ile jest rozsądny i mało zaskorupiały, dobry powód, by odrzucić mechanicyzm.

Na stwierdzenie mentalisty, że zrozumiał, jak może zrobić coś, co nie może zostać opisane w języku mechanicznego programu, mechanicysta odpowiada “Gadanie!” i nie wierzy mu dopóki ten nie skonstruuje programu, który pokaże, jak by to zrobił. Wygląda to jak sprzeczka pomiędzy realistą i fenomenalistą. Realista twierdzi, że istnieją rzeczy nie obserwowane przez nikogo – fenomenalista domaga się empirycznego dowodu. Jeśli ten nie zostanie mu dany, pozostanie sceptyczny wobec twierdzeń realisty. Jeśli zaś zostanie – rzecz owa nie będzie nieobserwowana. W ten sam sposób mechanicysta pozostaje sceptyczny wobec twierdzeń mentalisty dopóki ten nie dokładnie nie opisze sposobu, w jaki chce zrobić to, czego maszyna nie potrafi. Jeśli dokładny opis sposobu nie zostanie mu dany, pozostanie sceptyczny. Jeśli zostanie – posłuży [ów sposób – MZ] za podstawę do takiego zaprogramowania maszyny, by również to zrobiła. Stanowisko mechanicysty, podobnie jak fenomenalisty, jest wprawdzie nienaruszalne, ale nieprzekonujące. Nie mogę udowodnić mechanicyście, że cokolwiek może być zrobione inaczej, niż zrobiłaby to maszyna, ponieważ do takiego stopnia ograniczył to, co akceptuje jako dowód, że tylko czynności “wykonalne przez maszyny” uznane będą za wykonalne w ogóle. Nie wszyscy mechanicyści są jednak tak ograniczeni. Wielu mechanicystów i wielu mentalistów to racjonalne podmioty, które rozważają, czy w świetle nowoczesnej nauki i cybernetyki mechanicyzm jest, czy nie jest prawdą. Nie zamknęli swych umysłów przez takie przedefiniowanie pojęcia dowodu, po którym jedynie mechanicystyczne wnioski będą do utrzymania. Dostrzegają u samych siebie, że coś zrozumieli, nawet jeśli nie można napisać

programu, by i maszyna to coś zrozumiała. Pomocną [dla zrozumienia problemu – MZ] jest analogia do argumentu Sorites. Argumentując przeciw finityście, który nie uznaje zasady indukcji matematycznej, mogę dostrzec z metapoziomu, że jeśli przyjął F(0) oraz ∀x (F(x) → F(x+1)), wówczas mogę stwierdzić, nie popadając w sprzeczność, ∀x F(x). Mogę być tego całkiem pewien, mimo że nie mam na to finitystycznego dowodu. Wszystko, co mogę zrobić, znajdując się *vis à vis* finitysty, to wykazać, że gdyby odrzucił moje twierdzenie dla jakiegokolwiek konkretnego przypadku, mógłbym udowodnić, że się myli. To prawda – [inny – MZ] finitysta również mógłby mu wykazać, że się myli. Ja jednak dokonałem uogólnienia w sposób, na który finitysta nie może sobie pozwolić, zatem choć każdy argument obalający jest skończony, to jednak twierdzenie jest nieskończone. W podobny sposób każdy argument gödlowski jest efektywny i nawet mechanicystę przekona, że się myli, ale jego uogólnienie od poszczególnego taktycznego obalenia do strategicznego twierdzenia nie musi być efektywne w tym samym sensie, mimo że umysł zupełnie racjonalnie może je uznać.

Pomimo to, pozostaje wrażenie paradoksalności. Idea całkowicie intuicyjnego, nie poddawalnego formalizacji argumentu wzbudza podejrzenia – jeśli coś może przekonywać, może zostać zakomunikowane, a jeśli może zostać zakomunikowane, może również zostać sformułowane i wyrażone w języku formalnym. Pozwolę sobie zatem podkreślić, iż nie twierdzę, że mój, lub jakikolwiek inny, argument kategorycznie nie poddaje się formalizacji. Dowolny argument może zostać sformalizowany, jak to żółw dowiódł Achillesowi, lecz formalny aksjomat lub reguła inferencji, które uczyni swą podstawą, nie będą bardziej przekonujące niż pierwotny, niesformalizowany argument. Nie twierdzę, że argument gödlowski nie może zostać sformalizowany, ale jakąkolwiek zastosujemy formalizację, pozostaną inne argumenty, w sposób oczywisty poprawne, lecz nie objęte przez tę formalizację. Nie tylko – ponownie, jak to żółw dowiódł Achillesowi – zawsze musimy być gotowi, by bez zbędnych ceregieli rozpoznać jakieś reguły inferencji jako stosowalne i same inferencje jako poprawne, lecz powinniśmy także, o ile jesteśmy racjonalni, być w stanie rozszerzyć zbiór przyjętych przez nas inferencji poza wszelkie wyznaczone uprzednio granice. To nie wyklucza

późniejszej ich formalizacji – jedynie nasze założenie, że każda formalizacja jest inferencyjnie zupełna.

Zawsze jednak możemy formalizować. W szczególności, możemy sformalizować argument, którego Gödel użył, by udowodnić, że formuła gödlowska jest niedowodliwa-w-systemie, a mimo to prawdziwa. Na pierwszy rzut oka wygląda to na paradoks. Argument Gödla ma pokazywać, że zdanie gödlowskie jest niedowodliwe, lecz prawdziwe. Jeśli jednak pokazuje, że zdanie gödlowskie jest prawdziwe, z pewnością zostało to w nim dowiedzione, zatem, koniec końców, jest ono dowodliwe. Paradoks zostaje w tym przypadku rozwiązany przez rozróżnienie dowodliwości-w-formalnym-systemie od nieformalnej dowodliwości, której dostarczyło nam rozumowanie Gödla. Rozumowanie to może być jednak sformalizowane. Możemy prześledzić argument Gödla krok po kroku i sformalizować go. Jeśli tak uczynimy, zorientujemy się, że niezbędnym założeniem dla jego argumentu mówiącego, iż zdanie gödlowskie jest niedowodliwe, jest to, że system formalny powinien być niesprzeczny. W przeciwnym razie każde zdanie byłoby dowodliwe, a zatem zdanie gödlowskie, zamiast być niedowodliwe i dlatego prawdziwe, mogłoby być dowodliwe i fałszywe. Tym zatem, co otrzymamy, jeśli sformalizujemy nieformalną argumentację Gödla, nie będzie formalny dowód w obrębie elementarnej teorii liczb (w skrócie – ETL) na to, że zdanie gödlowskie G jest prawdziwe, lecz formalny dowód w obrębie Elementarnej Teorii Liczb:

$$\vdash \text{Cons (ETL)} \rightarrow G,$$

gdzie Cons (ETL) jest zdaniem oznajmiającym niesprzeczność elementarnej teorii liczb. Tylko jeśli w obrębie elementarnej teorii liczb posiadamy dowód dający:

$$\vdash \text{Cons (ETL)},$$

możemy otrzymać, poprzez zastosowanie reguły modus ponens:

$\vdash G$.

Ponieważ wiemy, że:

$\neg \vdash G$ [tzn. G nie jest wywodliwe],

otrzymujemy także:

$\neg \vdash Cons\ (ETL)$ [tzn. Cons (ETL) nie jest wywodliwe].

Oto drugie twierdzenie Gödla. Wielu krytyków odwoływało się do niego, by obalić argument gödlowski. Tylko jeśli formalny system maszyny jest niesprzeczny, a my jesteśmy w stanie stwierdzić jego niesprzeczność, wówczas mamy prawo utrzymywać, że zdanie gödlowskie jest prawdziwe. Nie mamy jednak na to [tj. niesprzeczność maszyny – MZ] świadectw. Wszystko, co wiemy, to że maszyna, z którą mamy do czynienia, może być sprzeczna, a nawet jeśli jest niesprzeczna, nie jesteśmy uprawnieni by to stwierdzić. A wobec braku takiego uprawnienia, jedyne, co udało nam się udowodnić, to:

$\vdash Cons\ (ETL) \rightarrow G$,

a to może zrobić również maszyna.

Powyższe krytyki opierają się na dwóch głównych zarzutach: niesprzeczność systemu maszyny jest założona w argumencie gödlowskim, a nie zawsze może być ustalona w drodze standardowej procedury decyzyjnej. Pytanie “Jakim prawem umysł zakłada, że maszyna jest niesprzeczna?” jest wobec tego trafne. Jednak kroki podjęte przez mechanicystów, by odmówić umysłowi tej wiedzy, są nieprzekonujące. Paul Benacerraf sugeruje, że mechanicysta może ujść argumentowi gödlowskiemu, jeśli nie będzie

przyglądał się swojemu twierdzeniu [tj. że umysł jest maszyną – MZ] zbyt szczegółowo[18]. Mechanicysta oferuje "Czarną Skrzynkę" bez określania jej programu i odmawia podania dalszych szczegółów poza stwierdzeniem, że czarna skrzynka reprezentuje umysł. Takie stanowisko jest jednak bezsensowne i nie do obrony. Bezsensowne, ponieważ nie ma żadnej mechanicystycznej zawartości, dopóki nie zostanie podana jakaś specyfikacja – jeśli pokazuje mi się czarną skrzynkę, lecz zarazem "mówi, by nie zaglądać do środka", dlaczego mam sądzić, że zawiera ona maszynę, a nie, powiedzmy, małego czarnego człowieczka? Stanowisko mechanicysty jest także nie do obrony: mimo że nie zgodził się on, by określić, co to za maszyna, o której twierdzi, że reprezentuje umysł, to jasnym jest, że argument gödlowski działa dla każdej niesprzecznej maszyny, a maszyna sprzeczna byłaby z kolei reprezentacją niewiarygodną. Metoda trzymania kart bardzo blisko piersi w celu zaprzeczenia przesłankom, które są korzystne dla umysłu, jest w istocie przyznaniem się do porażki.

Putnam utrzymuje, że dokonuje się nieuprawnionego przejścia od prawdziwej przesłanki:

Widzę, że (Cons (ETL) → G)

do fałszywego wniosku:

Cons (ETL) → widzę, że (G)[19].

To jest drugi krok potrzebny do odróżnienia umysłu od maszyny, jako że tym, co pokazuje twierdzenie Gödla, jest:

Cons (ETL) → ETL-maszyna nie widzi, że (G),

18 P. Benacerraf, *God, The Devil and Gödel*, "The Monist", vol. LI, 1 (1967), ss. 9-32.

19 H. Putnam, *Minds and Machines*, w: S. Hook (ed.), *Dimensions of Mind: A Symposium*, New York 1960, ss. 148-180; przedruk w: H. Putnam, *Mind, Language and Reality*, Cambridge 1975, ss. 362-385.

ja jednak, według Putnama, jestem uprawniony jedynie do zrobienia pierwszego. Zarzut Putnama spala na panewce z powodu dialektycznej natury arguentu gödlowskiego. Umysł nie chodzi w kółko i nie wypowiada twierdzeń z nadzieją na podstawienie nogi maszynom, które mogą znajdować się w pobliżu. Takim jest raczej twierdzenie, które zupełnie poważnie utrzymuje mechanicysta, że umysł może być repezentowany przez maszynę. Nim zaczniemy tracić czas na twierdzenie mechanicysty, rozsądnym jest zadać mu parę pytań odnośnie do jego maszyny, by przekonać się, czy jego zupełnie poważnie utrzymywane twierdzenie ma jakieś poważne przesłanki. Rozsądnym jest spytać go nie tylko o specyfikację maszyny, lecz także o to, czy jest niesprzeczna. Jeśli nie jest niesprzeczna, twierdzenie nawet nie oderwie się od ziemi. Jeśli zaświadczy się, że jest niesprzeczna, da to wówczas umysłowi niezbędne przesłanki [by wszcząć procedurę gödlowania – MZ]. Niesprzeczność maszyny nie jest ustalona za pomocą matematycznych zdolności umysłu, lecz słowem mechanicysty. Mechanicysta twierdzi, że jego maszyna jest niesprzeczna. Jeśli tak – nie może ona udowodnić swojego zdania gödlowskiego, którego prawdziwość umysł jednak widzi. Jeśli nie – sprawa zostaje oddalona tak czy inaczej.

Wang przyznaje, że rozsądnym jest twierdzić, iż tylko niesprzeczne maszyny są poważnymi kandydatami do tego, by reprezentować umysł, lecz później zastrzega, że wymaganie to jest nazbyt rygorystyczne, by mechanicysta mógł mu sprostać, ponieważ nie ma procedury decyzyjnej, która w każdym przypadku orzekłaby, czy system formalny, wystarczająco mocny, by zawierać elementarną teorię liczb, jest niesprzeczny, czy nie[20]. Jednak fakt, że nie ma procedury decyzyjnej oznacza jedynie, że nie zawsze możemy orzec, a nie że nigdy nie możemy. Częstokroć możemy orzec, że system formalny nie jest niesprzeczny – np. gdy dowodzi jako twierdzenie:

$$\vdash p \wedge \neg p$$

20 H. Wang, dz. cyt., s. 317.

lub

$\vdash 0 = 1.$

Możemy również być w stanie orzec, że system jest niesprzeczny. Posiadamy finitarne dowody niesprzeczności dla rachunku zdań i rachunku predykatów pierwszego rzędu, a także odwołujący się do indukcji pozaskończonej dowód Genzena dla elementarnej teorii liczb. Nie prosimy zatem mechanicysty o niemożliwe, gdy wymagamy od niego, by dokonał jakiejś wstępnej selekcji, nim przedstawi kandydatów na wiarygodne reprezentacje umysłu. Jeśli swą niesprzecznością nie zadowolą egzaminatora – mechanicysty – na etapie eliminacji, nie będą uprawnione do wzięcia udziału w finałach, te zaś, które się tam zakwalifikują, mogą być pewne, że poniosą porażkę, gdyż nie będą w stanie oznajmić swojego zdania gödlowskiego.

Dwuetapowe postępowanie egzaminacyjne jest w stanie oddzielić sprzeczne plewy, które poległy już w kwalifikacjach, od niesprzecznego ziarna, które poległo w finałach i w ten sposób umożliwia nam podjęcie każdego wyzwania, nawet jeśli rzuci je sprzeczna maszyna, bez udawania, że posiadamy nadludzkie moce. Choć wszystkie maszyny są uprawnione do tego, by przystąpić do egzaminu na reprezentację umysłu, stosunkowo niewiele maszyn to wiarygodni kandydaci do reprezentowania umysłu i nie ma potrzeby brać kandydata poważnie tylko dlatego, że jest maszyną. Jeśli twierdzenie mechanicysty ma być brane poważnie, wymagana jest jakaś rekomendacja, a przynajmniej poręczenie niesprzeczności byłoby niezbędne. Wang skarży się, że to oznacza oczekiwanie od niego [mechanicysty – MZ] nadludzkich mocy, a ja, w nawiązaniu do książki Benacerrafa "God, The Devil and Gödel", podchwyciłem jego sugestię, że mechanicysta nie musiałby być tylko człowiekiem, lecz mógłby stać się samym Księciem Ciemności, do którego można by, spodziewając się odpowiedzi, zaadresować pytanie, czy dana maszyna jest sprzeczna, czy nie[21]. Zamiast zadawać

21 P. Benacerraf, dz. cyt., ss. 22-23; J. R. Lucas, *Satan Stultified*, "The Monist", vol. LII, 1 (1968), ss. 152-

górnolotne pytania o umysł, możemy raczej zadać mechanicyście jedno pytanie o to, czy maszyna, którą zaproponował jako reprezentację umysłu, oznajmiłaby zdanie gödlowskie swojego systemu. Jeśli mechanicysta powie, że jego maszyna oznajmi zdanie gödlowskie, wówczas umysł będzie wiedział, iż jest sprzeczna i oznajmi wszystko, zupełnie nie jak umysł, który w charakterystyczny sposób jest selektywny na swym intelektualnym wyjściu. Jeśli mechanicysta powie, że jego maszyna nie wygłosi zdania gödlowskiego, wówczas umysł będzie wiedział, że jest niesprzeczna, jako że istnieje co najmniej jedno zdanie, którego maszyna nie może udowodnić w swoim systemie. A wiedząc to, umysł będzie wiedział, że zdanie gödlowskie maszyny jest prawdziwe i w ten oto sposób będzie się różnił od maszyny na swoim intelektualnym wyjściu. Jeśli zaś mechanicysta jest tylko człowiekiem i, co więcej, nie wie, jakiej odpowiedzi udzieli maszyna na pytanie gödlowskie, oznacza to, że nie odrobił swojej pracy domowej właściwie i powinien pójść i spróbować się dowiedzieć, zanim oczekiwałby od nas, że potraktujemy go poważnie.

Zadając pytanie raczej mechanicyście niż maszynie robimy użytek z faktu, że mówimy o pewnej zasadzie, nie zaś praktyce. Mechanicysta nie przedstawia rzeczywistych maszyn, które faktycznie reprezentują intelektualne wyjście jakiejś istoty ludzkiej, lecz twierdzi, że z zasady taka maszyna może istnieć. Zachęca nas do wykonania intelektualnego skoku, dokonując ekstrapolacji różnych teorii naukowych i prześlizgując się po wielu trudnościach. Ma do tego prawo. Jeśli jednak już to zrobi, nie ma prawa do tego, by być nieskorym do ujawnienia intelektualnych zdolności swej z-zasady maszyny lub tego, by odmawiać odpowiedzi na kłopotliwe pytania. Eksperyment myślowy, skoro już podjęty, musi być dobrze przemyślany. A gdy już zostanie dobrze przemyślany, [mechanicysta – MZ] znajdzie się w ślepej uliczce. Albo maszyna potrafi udowodnić w swoim systemie zdanie gödlowskie, albo nie potrafi. Jeśli potrafi, jest sprzeczna i nie stanowi odpowiednika umysłu. Jeśli nie potrafi, jest niesprzeczna, a wówczas umysł potrafi stwierdzić prawdziwość zdania gödlowskiego. W obu wypadkach maszyna nie odpowiada umysłowi, a teza mechanicysty upada.

153.

Wielu myślicieli zdecydowało się zabrnąć w ślepą uliczkę sprzeczności. Jesteśmy maszynami – mówią – ale bardzo ograniczonymi, omylnymi i sprzecznymi. W świetle wielu naszych sprzeczności, zmian zdania i braku logiki, nie mamy świadectw, by zakładać, że umysł jest niesprzeczny i z tego względu nie mamy również podstaw, by za sprzeczność wykluczyć maszynę z grona kandydatów na reprezentację umysłu. Hofstadter sądzi, że w zupełności możliwe byłoby zbudować sztuczną inteligencję, w której rozumowania zdaniowe pojawiłyby się raczej jako konsekwencje niż coś uprzednio zaprogramowanego. "I nie ma szczególnego powodu, by zakładać, że ścisły rachunek zdań, ze swoimi sztywnymi regułami i raczej głupią definicją niesprzeczności, którą one za sobą pociągają, wyłoniłby się z takiego programu."[22]

Żaden z tych argumentów nie znajduje sposobu, by sprzeczną maszynę uczynić wiarygodną reprezentacją umysłu. Prawdą jest, że słowo "niesprzeczny" użyte jest w różnych znaczeniach i twierdzenie, że umysł jest niesprzeczny pociąga za sobą raczej inne znaczenie niesprzeczności i dochodzi się do niego za pomocą innego rodzaju argumentów niż w przypadku, w którym twierdzi się, że maszyna jest niesprzeczna. Jeśli to wystaczy, by zarysować różnicę pomiędzy umysłami i maszynami, wszystko pięknie. Wielu mechanicystów nie da się jednak tak szybko przekonać i będzie utrzymywać, że maszyna może zostać zaprogramowana, tak jak to, w pewnej mierze, zakłada Hofstadter, by naśladować ludzkie zachowanie. W takim wypadku mówimy raczej o maszynowej niesprzeczności, nie zaś o ludzkiej. Każda maszyna, jeśli ma zacząć reprezentować wyjście umysłu, musi być zdolna operować na symbolach, które mogą być wiarygodnie interpretowane jako negacja, koniunkcja, implikacja itd. i w ten sposób musi podlegać zasadom którejś odmiany rachunku zdań. Dopóki coś przypominającego rachunek zdań z jakimś porównywalnym warunkiem niesprzeczności nie wyłoni się z programu maszyny, nie będzie ona wiarygodną reprezentacją umysłu, bez względu na to, jak dobrym jest okazem sztucznej

22 Hofstadter, dz. cyt., s. 578; por. C. S. Chihara, *On Alleged Refutations of Mechanism using Gödel's Incompleteness Results*, "Journal of Philosophy", vol. LXIX, 19 (1972), s. 526.

inteligencji. Oczywiście każdą wiarygodną reprezentację umysłu musiałoby cechować zachowanie, którego przykład dał Wang, polegające na nieustannym kontrolowaniu, czy otrzymano sprzeczność i próbach rewizji swoich podstawowych aksjomatów, jeśli do tego doszło. Musiałoby się to jednak odbywać zgodnie z pewnymi regułami. Musiałby istnieć program udzielający dokładnych instrukcji, w jaki sposób owa kontrola miałaby zostać podjęta i w jakiej kolejności aksjomaty miałyby być rewidowane. Niektóre aksjomaty musiałyby być dość odporne na rewizję. Choć pewni myśliciele gotowi są wyobrazić sobie rachunek logiczny, w którym nie obowiązują podstawowe inferencje rachunku zdań (np. od p□q do p) albo odrzucone zostały aksjomaty elementarnej teorii liczb, każda maszyna, która uciekłaby się do takiego fortelu, by uniknąć sprzeczności, straciłaby także wszelką wiarygodność jako reprezentacja umysłu. Choć czasem sobie zaprzeczamy i zmieniamy zdanie, pewne elementy naszej struktury pojęciowej pozostają bardzo stabilne i odporne na rewizję. Oczywiście nie jest to całkowita odporność. Można dopuścić kartezjańską możliwość rewizji pojęciowej bez narażania się, jak zakłada Hutton[23], na sprzeczność, gdy stwierdza się, że posiada się wiedzę o swojej własnej niesprzeczności. Stwierdzić, że coś się wie, to nie tyle stwierdzić własną nieomylność, co posiadać odpowiednie uzasadnienie dla tego, co zostało stwierdzone. W przeciwnym razie wiedza o prawdach przygodnych byłaby niemożliwa. Choć nie można powiedzieć: “Wiem to, choć być może się mylę”, jest w zupełności dopuszczalne powiedzieć: “Wiem to, choć możliwe do pomyślenia jest, że się mylę”. Tak długo, jak ktoś posiada dobre powody, może z pełną odpowiedzialnością wystawić gwarancję – w postaci stwierdzenia – na to, że wie, nawet jeśli my potrafimy wyobrazić sobie okoliczności, w których to, co twierdzi, okałoby się fałszem i musiałoby zostać odwołane. Tak jest również w przypadku naszych roszczeń do wiedzy na temat podstawowych elementów naszej struktury pojęciowej, takich jak zasady rozumowania zawarte w rachunku zdań lub prawdy zwykłej nieformalnej arytmetyki. Posiadamy dostateczny, więcej niż dostateczny, powód, by stwierdzić naszą własną niesprzeczność i prawdziwość, a stąd także niesprzeczność nieformalnej arytmetyki i w ten sposób możemy zasadnie stwierdzić, że

23 A. Hutton, *This Gödel is Killing Me*, “Philosophia”, vol. VI, 1 (1976), ss. 135-144.

wiemy i że każda maszynowa reprezentacja umysłu musi przedstawiać to, co na wyjściu, wyrażone przez formalny (skoro jest maszyną) system, który jest niesprzeczny i zawiera elementarną teorię liczb (skoro ma reprezentować umysł). Pozostaje jednak kartezjańska możliwość, że się mylimy i że musimy to teraz przedyskutować. Niektórzy mechanicyści przyznawali, że niesprzeczna maszyna mogłaby być wygödlowana przez umysł, lecz utrzymywali, że maszynowa reprezentacja umysłu jest maszyną sprzeczną, lecz której sprzeczność jest tak głęboko ukryta, że upłynęłoby dużo czasu, zanim wyszłaby na światło dzienne. Z tego względu uniknęłaby szybkiego odstrzału z powodu swojego braku selektywności. Choć co do zasady mogłaby być zmuszona do stwierdzenia wszystkiego, w praktyce byłaby selektywna, stwierdzając pewne rzeczy, a innym zaprzeczając. Tylko w dalszej perspektywie zestarzałaby się – lub złagodniała, jak to grzecznie określamy – a wówczas "siadła" i przestała zaprzeczać czemukolwiek. A w dalszej perspektywie i my umrzemy – zazwyczaj przed zapadnięciem na starczą demencję. Taka sugestia współgra ze sposobem rozumowania, który był zauważalny w myśli zachodniej od osiemnastego stulecia. Rozum, jak się utrzymuje, uwikłany jest w pewne antynomie i przez swą własną dialektyczną strukturę daje pożywkę wewnętrznym sprzecznościom, których nie ma jak wzajemnie pogodzić, co w efekcie sprawia, że cały gmach musi lec w gruzach. Jeśli umysł naprawdę jest sprzeczną maszyną, wówczas ci filozofowie nawiązujący do tradycji heglowskiej, którzy mówili o autodestruktywności rozumu, są tymi, u których sprzeczność ujawniła się względnie szybko. Są tymi, którzy pojęli nieodłączną sprzeczność rozumu i którzy, negując negację, porzucili nadzieję na racjolany dyskurs i doprowadzając umysł na skraj wytrzymałości, oferują jedynie rzucenie się w objęcia rozpaczy.

Przeciwko temu stanowisku argument gödlowski nic nie zdziała. Raczej inne argumenty i inne zachowania są potrzebne jako antidotum przeciwko nihilizmowi. Długo uważano, że materializm prowadzi do nihilizmu – i argument gödlowski czyni tę redukcję wyraźną. Bo jest to redukcja. Dla mechanicyzmu powinno to być racjonalne stanowisko. Opiera swe argumenty na postępach nauki, podstawowych założeniach naukowego myślenia i na aktualnych osiągnięciach badań naukowych. Choć inni ludzie

mogą zostać doprowadzeni do nihilizmu przez uczucia strachu lub inne ujawnienia nicości, mechanicysta musi przedstawić argumenty lub wycofać swoje poparcie dla wszystkiego [co do tej pory utrzymywał – MZ]. W świetle tego wszystkiego, nie jesteśmy maszynami. Można by przytoczyć argumenty na rzecz tego, że pozory mylą i w rzeczywistości jesteśmy maszynami, lecz argumenty zakładają racjonalność a jeśli, dzięki argumentowi gödlowskiemu, jedyną możliwą do obrony formą mechanicyzmu jest [ta, która utrzymuje – MZ], że jeststeśmy maszynami sprzecznymi, a więc, wszystkie umysły są ostatecznie sprzeczne, wówczas sam mechanicyzm skazany jest na nieracjonalność argumentów i żaden racjonalny argument na jego rzecz nie może zostać utrzymany.

Przełożył Michał Zawidzki