Zagadnienia Filozoficzne
w Nauce
XLVII (2010), 70–88

Wojciech P. Grygiel
Wydział Filozoficzny, Uniwersytet Papieski Jana Pawła II w Krakowie
Centrum Kopernika Badań Interdyscyplinarnych w Krakowie

JAK UNIESPRZECZNIĆ SPRZECZNOŚĆ UMYSŁU?

WPROWADZENIE

W opinii większości filozofów od czasów Arystotelesa, nasza racjonalność związana jest ściśle z logiką klasyczną, warunkowaną trzema fundamentalnymi zasadami: prawem niesprzeczności ($\neg(p \wedge \neg p)$), prawem wyłączonego środka ($p \wedge \neg p$) oraz prawem tożsamości ($p \leftrightarrow p$). Najbardziej podstawową z nich wydaje się prawo niesprzeczności, o którego oczywistości przekonuje zdrowy rozsądek. Mówiąc o racjonalności, która jest atrybutem jednostki, nie sposób pominąć kwestii umysłu. W tradycyjnym ujęciu reguły logiki klasycznej warunkują funkcjonowanie naszego umysłu nie dopuszczając pojawiania się w nim sądów sprzecznych. Stanowisko, zgodnie z którym jedną z podstawowych cech umysłu jest niesprzeczność, łączy się zwykle z przekonaniem o niealgorytmiczności umysłu, a więc stwierdzeniem, że nie jest on programem komputerowym[1]. Mimo tego, że nasze rozumowania są nieraz błędne, nasza pamięć szwankuje i nie jest trudno wykazać komuś głoszenie sprzecznych poglądów; powszechnie przyjmuje się, że umysł jest najdoskonalszym produktem ewolucji biologicznej. Co

[1]J.R. Lucas, *Minds, Machines and Gödel*, „Philosophy", vol. XXXVI, 1961, 112-127. Dostępny w języku polskim: J.R. Lucas, *Umysły, Maszyny i Gödel*, tłum. M. Zawidzki, „Hybris — internetowy magazyn filozoficzny", nr 8 (2009), dostęp online [15.05.2010]: <http://www.filozof.uni.lodz.pl/ hybris/ pdf/ h09/6.%20Lukas2%20[7498].pdf>.

więcej, uważa się, że posiadanie samoświadomości i wysokorozwiniętych zdolności kognitywnych odróżnia nas od reszty ożywionej sfery uniwersum. Czy jednak nie można podać sensownych argumentów za tym, że działanie mózgu, warunkujące działanie umysłu, polega na przetwarzaniu informacji, a więc na obliczeniach, a sam umysł jest systemem sprzecznym?

W formie bardziej technicznej powyższe zagadnienia powracają w kontekście aplikacji twierdzenia Gödla o niezupełności systemów formalnych opartych na arytmetyce do filozofii umysłu i kognitywistyki. Twierdzenia Gödla wykorzystywane są często w dyskusjach nad algorytmicznością umysłu i sztuczną inteligencją, jako argument za stanowiskiem, zgodnie z którym umysł jest niealogrytmiczny i nie może być adekwatnie symulowany przy pomocy komputera. Taki sposób aplikacji twierdzeń limitacyjnych nazywać będziemy w skrócie „argumentem gödlowskim". Autorami najbardziej znanych prac w tej kwestii są John Lucas[2] oraz Roger Penrose[3].

Celem niniejszego opracowania jest wskazanie, że na obecnym poziomie badań nad mózgiem i umysłem paradygmat niesprzeczności umysłu nie jest jedynym uprawnionym. Zagadnienie to osadzone zostanie w kontekście badań nad algorytmicznością procesów mentalnych oraz ograniczeniami, jakie twierdzenia limitacyjne rzekomo nakładają na sztuczną inteligencję. Obecnie duża część debaty, jaka toczy się w ramach filozofii umysłu, koncentruje się wokół podziału różnych teorii umysłu na algorytmiczne i niealgorytmiczne. Entuzjastami podejścia algorytmicznego są przede wszystkim przedstawiciele *computer science*, ze szczególnym uwzględnieniem badaczy zajmujących się sztucznymi sieciami neuronowymi, a także inspirujący się wynikami tej dziedziny filozofowie[4]. Z drugiej strony zaś zwolennikami koncepcji, zgodnie z którą umysł jest niealgorytmiczny, są neurobiolodzy i filozofowie, akcentujący wagę biochemicznego podłoża działania

[2]Zob. tamże.

[3]Zob. R. Penrose, *Cienie umysłu. Poszukiwanie naukowej teorii świadomości*, tłum. P. Amsterdamski, Zysk i S-ka, Poznań 2000.

[4]Zob. np. P.S. Churchland, T.J. Sejnowski, *The Computational Brain*, MIT Press, Cambridge — London 1996.

mózgu, które ich zdaniem nie może być adekwatnie symulowane przy pomocy metod obliczeniowych[5].

Celem niniejszej pracy jest wskazanie, że kwestią bardziej fundamentalną dla filozofii umysłu i kognitywistyki może być pytanie o niesprzeczność umysłu. W tym wypadku dychotomią bardziej fundamentalną niż algorytmiczny / niealgorytmiczny jest sprzeczny / niesprzeczny. Biorąc pod uwagę powyższe dychotomie można *a priori* rozważać cztery zasadnicze sytuacje: [i] umysł algorytmiczny niesprzeczny, [ii] umysł algorytmiczny sprzeczny, [iii] umysł niealgorytmiczny niesprzeczny oraz [iv] umysł niealgorytmiczny sprzeczny. Naturalnym środowiskiem dla koncepcji sprzeczności umysłu jest paradygmat komputacjonistyczny, w którym pojęcie sprzecznego systemu formalnego równoważne jest błędnemu algorytmowi[6]. Przedstawione zostaną racje, zgodnie z którymi nie można obecnie wykluczyć, że umysł równoważny jest sprzecznej maszynie Turinga. Przywołane będą argumenty osłabiające konsekwencje aplikacji twierdzeń Gödla do filozofii umysłu, jak i przykłady sprzeczności umysłu. Postawiona zostanie hipoteza, zgodnie z którą „globalna" sprzeczność umysłu pojawiać może się na poziomie integracji „lokalnych" modułów obliczeniowych, których istnienie postuluje psychologia ewolucyjna[7].

Zaznaczyć należy również, że niniejsza praca ma charakter wysoce hipotetyczny. Zgodnie z sugestią Michała Hellera, zaczerpniętą od Karla Poppera, poglądy filozoficzne traktować należy podobnie, jak zdania formułowane w ramach nauk empirycznych: nie jako *dogmaty* i ostateczne wyjaśnienia, ale jako poddające się rewizji *hipotezy*. Jeśli w naukach empirycznych za kryterium naukowości uznajemy *falsyfikowalność*, tak w filozofii, kryterium sensowności powinna być możli-

[5]Zob. np. J.R. Searle, *Umysł na nowo odkryty*, tłum. T. Baszniak, PIW, Warszawa 1999.

[6]Zob. tamże, s. 127 n.

[7]Zob. S.M. Downes, *Evolutionary Psychology*, [w:] *The Stanford Encyclopedia of Philosophy*, red. E.N. Zalta, dostęp online [29.06.2010]: <http://plato.stanford.edu/entries/ evolutionary-psychology/>.

wość krytycznej *dyskusji* nad daną hipotezą[8]. Wydaje się, że mimo spekulatywności, przedstawiane w niniejszej pracy tezy są *dyskutowalne*.

MODEL NIESPRZECZNEGO UMYSŁU: LUCAS, PENROSE

Aby przybliżyć problematykę sprzeczności umysłu, warto rozpocząć od przedstawienia modelu, który zakłada niealgorytmiczność umysłu. Dobrym przykładem jest kwantowy model Rogera Penrose'a, gdyż zawiera zarówno stronę fizyczno-biologiczną, jak i logiczno-matematyczną[9]. Warto podkreślić, że Penrose rozważa problematykę mózgu i umysłu w kontekście poszukiwań teorii fizycznej, która połączy mechanikę kwantową z ogólną teorią względności. Fundamentem jego rozważań jest „globalna" ontologia, która postuluje istnienie trzech światów: świata matematyki, świata fizyki oraz świata umysłu[10]. Zadaniem nowej, uogólnionej teorii fizycznej ma być scalenie tych światów w ramach jednego formalizmu. Podstawą dla świata fizycznego jest świat bytów matematycznych. Byty te rozumiane są w sposób platoński, tj. jako pozaprzestrzenne, pozaczasowe i istniejące poza umysłem matematyka. Świat fizyczny, a konkretnie jego struktury związane z budową układu nerwowego są z kolei podstawą dla zaistnienia zjawisk mentalnych. Penrose jest fizykalistą, gdyż sądzi, że zjawiska mentalne redukowalne są do procesów fizycznych. Światy tworzą hierarchiczną całość, a najbogatszym z nich jest świat umysłu. Ontologia Penrose'a przedstawiana jest graficznie w formie trójkąta. Jego domknięcie stanowi odniesienie umysłu do uniwersum platońskich obiektów matematycznych. Możliwe jest to dzięki *wglądowi matematycznemu* związanemu z *rozumieniem*. Wgląd ten nazywany jest

[8]Zob. M. Heller, *Przeciw fundacjonizmowi*, [w:] tenże, *Filozofia i Wszechświat*, Znak, Kraków 2006, s. 95.

[9]Koncepcja umysłu Penrose'a została omówiona szczegółowo w: W.P. Grygiel, M. Hohol, *Rogera Penrose'a kwantowanie umysłu*, „Filozofia nauki", XVII, nr 3(67), 2009, ss. 5–31.

[10]Zob. R. Penrose, *Droga do rzeczywistości. Wyczerpujący przewodnik po prawach rządzących Wszechświatem*, tłum. J. Przysława, Prószyński i S-ka, Warszawa 2006, s. 7–21.

również *intuicją matematyczną*[11]. Podstawową cechą intuicji matematycznej jest jej niealgorytmiczność, która związana jest z niemożliwością adekwatnego symulowania przy pomocy komputera. Penrose jest przeciwnikiem obliczeniowych modeli umysłu, związanych ze stanowiskiem silnej sztucznej inteligencji.

W tym kontekście wymienia on cztery stanowiska związane z możliwością naukowego wyjaśnienia świadomości przy pomocy obliczeń. Zgodnie z pierwszym z nich (*A*) umysł jest *tożsamy* z odpowiednim algorytmem, który wykonywany może być przez dowolny komputer. Zaznaczyć należy, że Penrose pojęcia takie jak *obliczenia* oraz *algorytm* traktuje jako synonimy. Stanowisko (*A*) określane jest jako *silna sztuczna inteligencja*. Stanowisko (*B*) głosi, że działanie mózgu można adekwatnie symulować przy pomocy obliczeń, jednak symulacji taka nigdy nie osiągnie świadomości. Pogląd ten nazywany jest *słabą sztuczną inteligencją*. Stanowisko (*C*), którego zwolennikiem jest sam Roger Penrose mówi, że nie można obliczeniowo symulować nie tylko zjawisk mentalnych, ale także działania mózgu. Ich naukowe wyjaśnienie jest możliwe, jednak z uwagi na niealgorytmiczność konieczne jest stworzenie nowej fizyki. Zgodnie z ostatnim ze stanowisk (*D*) umysł na zawsze pozostanie tajemnicą dla nauki. Nie można go wyjaśnić ani przy pomocy obliczeń, ani w żaden inny naukowy sposób[12]. Warto zaznaczyć, że stanowiska (*A*) i (*B*) nie dopuszczają zdaniem Penrose'a, że algorytm jest błędny (równoważny sprzecznemu systemowi formalnemu).

Jak zostało już powiedziane, Penrose próbuje dowieść niealgorytmiczności umysłu w dwóch wymiarach: logiczno-matematycznym oraz fizyczno-biologicznym. Warto rozważyć obecnie kwestie składające się na pierwszy z nich, gdyż to właśnie w kontekście ich krytyki pojawia się hipoteza sprzeczności umysłu[13]. Roger Penrose zaczyna swoją argumentację od zdroworozsądkowego przeświadczenia, że *ro-*

[11] Zob. R. Penrose, *Cienie umysłu...*, dz. cyt., s. 511.

[12] Zob. tamże, s. 31 n.

[13] Całościowa struktura argumentacji Penrose'a za niealgorytmicznością umysłu uwzględniająca również blok fizyczno-biologiczny przedstawiona została w: W.P. Grygiel, M. Hohol, *Rogera Penrose'a kwantowanie umysłu...*, art. cyt., s. 10–11.

zumienie nie polega na wykonywaniu algorytmu. Szczególnym przypadkiem, który staje się przedmiotem jego refleksji jest doświadczenie uprawiania matematyki. W związku z uznawanym przez niego stanowiskiem platonizmu matematycznego, Penrose uważa, że rozwiązywanie problemów matematycznych możliwe jest dzięki szczególnemu *wglądowi* w świat wiecznych bytów matematycznych. Wgląd ten związany jest z *rozumieniem*, które jego zdaniem nie może być utożsamione z żadnym algorytmem. Pojęcie algorytmu można wyrazić w języku niesformalizowanym następująco:

> Istotą *algorytmu* jest to, że rozwiązanie problemu (np., jaki jest iloraz dwóch danych liczb) polega na mechanicznym wykonywaniu czynności dyktowanych przez kolejne instrukcje przekształcania określonych obiektów, w szczególności symboli, gwarantując poprawne wykonanie zadania w skończonej liczbie kroków; instrukcje te biorą pod uwagę jedynie fizyczne cechy symboli, jak kształt czy położenie, abstrahują natomiast od ich znaczenia czy od wywołanych przez nie myśli[14].

Penrose podaje wiele przykładów niealgorytmiczności w matematyce, takich jak problem słowa, zagadnienie pokrycia płaszczyzny Euklidesa różnokształtnymi płytkami czy też problem rozwiązań równań diofantycznych[15]. W kwestiach poza matematycznych odwołuje się on natomiast do przykładu gry w szachy. Choć w istocie szachy są grą obliczalną, przykłady prostych błędów, jakie maszyny popełniają w rozgrywkach szachowych są, zdaniem Penrose'a, argumentem na rzecz niealgorytmiczności umysłu i przewagi człowieka nad komputerem[16]. Intuicyjnie zrozumiałe pojęcie algorytmu definiowane jest funkcjonalnie przy pomocy maszyny Turinga lub rachunków równoważnych, co zdaniem Penrose'a jest argumentem na rzecz tezy Churcha[17]. Procesami niealgorytmicznymi nazwać można więc takie, których symulacji nie można przeprowadzić w oparciu o maszynę Turinga.

---

[14]W. Marciszewski, *Sztuczna inteligencja*, Znak, Kraków 1998, s. 64

[15]Zob. R. Penrose, *Nowy umysł cesarza. O komputerach umyśle i prawach fizyki*, tłum. P. Amsterdamski, PWN, Warszawa 2000, s. 132, s. 152–161.

[16]Zob. R. Penrose, *Cienie umysłu...*, dz. cyt., s. 71 n.

[17]Zob. Tamże, s. 37.

Penrose wprowadza następnie argument za niealgorytmicznością umysłu, oparty na zmodyfikowanej wersji twierdzenia Gödla o niezupełności systemów formalnych, zawierających arytmetykę. Bardzo podobny sposób argumentacji logicznej za niealgorytmicznością umysłu przedstawiony został w 1961 roku przez Johna Lucasa[18]. Różnica ujawnia się w motywacjach obydwu uczonych. Penrose wychodzi od racji naukowych, natomiast Lucas od kwestii światopoglądowych. Ponadto Lucas ogranicza się do analizy pojęciowej, charakterystycznej dla filozofii analitycznej, natomiast Penrose formalizuje niektóre elementy wywodu oraz buduje molekularny model mózgu. W warstwie logicznej, gdzie mowa jest głównie o twierdzeniu Gödla, rozumowania obydwu uczonych traktować należy jako równoważne.

Dwoma ważnymi cechami systemu formalnego powinna być zupełność i niesprzeczność. W sprzecznym systemie formalnym, na mocy prawa logicznego Dunsa Szkota $p \wedge \neg p \rightarrow q$ dowieść można czegokolwiek (*ex contradictione quodlibet* (dalej: ECQ)). W 1931 roku Kurt Gödel ogłosił dwa słynne twierdzenia. Zgodnie z pierwszym z nich (TG1), w każdym systemie formalnym (aksjomatycznym) przynajmniej tak bogatym, że zbudować można w nim arytmetykę liczb naturalnych, istnieją poprawnie zbudowane w języku tego systemu zdania, których nie można wyprowadzić z aksjomatów tego systemu przy pomocy odpowiednich reguł inferencji. Natomiast zgodnie z drugim twierdzeniem Gödla (TG2) nie jest możliwe sformułowanie dowodu niesprzeczności systemu formalnego w oparciu o środki tego systemu. Dowód niesprzeczności wykorzystywać musi zawsze środki z metasystemu[19]. Argument gödlowski Penrose'a za niealgorytmicznością umysłu oparty jest na powyższych twierdzeniach. Jego celem jest wykazanie niealgorytmiczności rozumowań matematycznych, która ekstrapolowana jest następnie na inne zjawiska mentalne. Możliwe jest to dzięki stwierdzeniu przez Penrose'a, że *rozumienie* nie jest możliwe

[18]Zob. J.R. Lucas, *Minds, Machines and Gödel*, „Philosophy", vol. XXXVI, 1961, ss. 112-127. Dostępny w języku polskim: J.R. Lucas, *Umysły, Maszyny i Gödel*, tłum. M. Zawidzki, „Hybris — internetowy magazyn filozoficzny", nr 8 (2009), dostęp online [15.05.2010]:

[19]Zob. np. S. Krajewski, *Twierdzenie Gödla i jego interpretacje filozoficzne. Od mechanicyzmu do postmodernizmu*, IFiS PAN, Warszawa 2003, s. 63–67.

bez *inteligencji*, z kolei zaś inteligencja bez *świadomości*[20]. Jak zostało powiedziane, koncepcja Penrose'a nawiązuje do pomysłu Johna Lucasa z 1961 roku. Wnioski obydwu uczonych zobrazować można następującym cytatem:

> Twierdzenie Gödla musi stosować się do maszyn cybernetycznych, ponieważ w istocie maszyny tkwi, że jest ona konkretną realizacją systemu formalnego. Z tego wynika, że jeśli daną mamy jakąkolwiek maszynę, która jest niesprzeczna i zdolna generować prostą arytmetykę, istnieje formuła, której [...] nie jest w stanie przedłożyć jako prawdziwej, tj. formuła, która jest niedowodliwa-w-systemie, lecz której prawdziwość my widzimy. Z tego zaś wynika, że żadna maszyna nie może być zupełnym lub adekwatnym modelem umysłu; że umysły są istotowo różne od maszyn[21].

Zdaniem Penrose'a prawdziwość formuł niedowodliwych dla maszyny Turinga (Lucas używa nazwy „maszyna cybernetyczna") jest dla nas dostępna, gdyż umysł, w przeciwieństwie do maszyny posługuje się *zasadami refleksji* nad znaczeniem aksjomatów i reguł dowodzenia[22]. Intuicja matematyczna, związana z rozumieniem, jest cechą specyficznie ludzką. Penrose uważa, że odkrywanie prawdy matematycznej związane jest z niealgorytmiczną intuicją, która pozwala wkroczyć umysłowi w platoński świat matematyki. Oryginalne twierdzenie Gödla odnosi się do systemów formalnych, natomiast gödlowski argument Penrose'a operuje pojęciem algorytmu[23]. Zarówno Lucas, jak i Penrose traktują pojęcia *systemu formalnego* oraz *algorytmu* jako równoważne[24] zakładając w ten sposób prawdziwość tezy Churcha. Niesprzeczny system formalny utożsamiany będzie w takim wypadku z poprawnie działającym algorytmem, natomiast system sprzeczny, z algorytmem błędnym. Warto zauważyć, że tak jak dzieje się to w bardziej

[20] Zob. R. Penrose, *Cienie umysłu...*, dz. cyt., s. 60 n.

[21] J.R. Lucas, *Umysły, Maszyny i Gödel*, dz. cyt., s. 97–98.

[22] Zob. R. Penrose, *Nowy umysł cesarza...*, dz. cyt., s. 132.

[23] Dowód dla gödlowskiego argumentu Penrose'a znajduje się w: R. Penrose, *Cienie umysłu...*, dz. cyt., s. 101-106.

[24] Dowód równoważności pojęć systemu formalnego i algorytmu znaleźć można w: tamże, s. 127 n.

złożonych programach komputerowych o strukturze modularnej, niedoskonałość algorytmu prowadzi do generowania przez program błędnych wyników tylko w określonych wypadkach[25]. Błędny algorytm może być więc jak najbardziej funkcjonalny. Rozumowania Lucasa i Penrose'a opierają się na założeniu niesprzeczności umysłu. Podobnie, jak w kwestii systemów formalnych, tak i dla umysłu, twierdzenie Gödla ma sens tylko, gdy umysł jest niesprzeczny. Przeświadczenie o niesprzeczności umysłu powodowane jest natomiast głównie racjami zdroworozsądkowymi. Jak pisze Lucas:

> Nie tylko możemy uczciwie stwierdzić, że wiemy, iż jesteśmy niesprzeczni, pomimo błędów, które popełniamy, ale wręcz musimy w każdym wypadku zakładać, że jesteśmy, jeśli jakakolwiek myśl w ogóle ma być możliwa. Co więcej, cechujemy się selektywnością [...]. I wreszcie, możemy, w pewnym sensie, zdecydować się na bycie niesprzecznymi; w tym mianowicie, że możemy postanowić nie tolerować sprzeczności w naszym myśleniu i mówieniu oraz eliminować je, gdy tylko się pojawią, poprzez wyparcie się i odwołanie jednego z członów sprzeczności[26].

MODEL NIESPRZECZNEGO UMYSŁU: ZARZUTY

Niealgorytmiczny model umysłu Penrose'a, jak i wcześniejsza propozycja Lucasa, spotykały się z silną krytyką ze strony logików, filozofów i kognitywistów. Najwięcej argumentów krytycznych wysuniętych zostało pod adresem zastosowania twierdzenia Gödla w dowodzeniu niealgorytmiczności umysłu. Przykładowo, logicy tacy jak Willard van Orman Quine czy Paul Benacerraf twierdzą, że sama procedura generowania zdania gödlowskiego, czyli niedowodliwego-w-systemie w istocie jest również algorytmiczna[27]. Stwierdzenie to poddaje w wątpliwość sensowność uznawania niealgorytmicznej intuicji matematycz-

[25] Na temat problematyki obliczalności oraz modułów obliczeniowych w kontekście *neuroscience* zob. P.S. Churchland, T.J. Sejnowski, *The Computational Brain*, dz. cyt.

[26] J.R. Lucas, *Umysły, Maszyny i Gödel*, dz. cyt., s. 113.

[27] Zob. P. Benacerraf, *God, the Devil and Gödel*, „The Monist", 51, 1967, ss. 9–32, także dostęp online [29.06.2010]: <http://www2.units.it/ ~etica/ 2003_1/ 3_monographica.htm>.

nej. Dodatkowo, w przypadku Penrose'a, z silną krytyką spotkała się jego fizyczno-biologiczna część argumentacji[28]. Dla celów niniejszej pracy pożyteczne będzie rozważenie pewnych zarzutów wobec argumentu gödlowskiego, które opierają się na:

1. stwierdzeniu, że umysł jest sprzeczny i w związku z tym ograniczenia wynikające z twierdzenia Gödla nie obowiązują (podejście reprezentowane przez Alana Turinga i Hilary'ego Putnama, i Patricię Churchland);

2. wykazaniu trudności w dowodzeniu niesprzeczności umysłu (problem zauważany był już przez Kurta Gödla)[29].

Jak zostało już wyżej powiedziane, gdyby system formalny był sprzeczny, na mocy prawa ECQ można by dowieść w nim dowolnego zdania. Nie podpadałby on jednak pod ograniczenia nakładane przez twierdzenie Gödla. Zwolennikami koncepcji, zgodnie z którą umysł jest sprzecznym systemem formalnym, co równoważne jest z błędnym algorytmem, są filozofowie i naukowcy, tacy jak Alan Turing, Hilary Putnam, Rick Grush oraz Patricia Churchland[30]. Ich zdaniem nasze zdolności kognitywne generowane są przez czynniki algorytmiczne, co pozawala na uniknięcie ograniczeń wynikających z twierdzenia Gödla. Przyjęcie takiego stanowiska pozwala ich zdaniem na zachowanie obliczeniowej teorii umysłu (silnej sztucznej inteligencji). Wydaje się, że argument gödlowski jest najsilniejszą przesłanką logiczną za odrzuceniem obliczeniowej teorii umysłu. Podważenie tego typu argumentacji przez przyjęcie, że umysł związany jest ze sprzecznym systemem for-

[28] Większość głosów krytycznych wobec koncepcji Penrose'a rozważona została w: W.P. Grygiel, M. Hohol, *Rogera Penrose'a kwantowanie umysłu*, dz. cyt.

[29] Zob. S. Krajewski, *Twierdzenie Gödla i jego interpretacje filozoficzne*, dz. cyt., s. 86-87.

[30] W przypadku Turinga zob. np.: R. Penrose, *Cienie umysłu...*, dz. cyt., s. 169 n.; ustna opinia wygłoszona przez Putnama wspominana jest w: J.R. Lucas, *Umysły, Maszyny i Gödel*, dz. cyt., s. 108; natomiast jeśli chodzi o Grusha i Churchland zob.: R. Grush, P.S. Churchland, *Gaps in Penrose's Toilings*, [w:] red. P.M. Churchland, P.S. Churchland, *On the contrary. Critical essays 1987-1997*, MIT, Boston 1998, s. 227.

malnym, może wydawać się trudne do przyjęcia ze zdroworozsądkowego punktu widzenia, jednak odbywa się bez szkód dla obliczeniowej teorii umysłu. Sztuczna inteligencja jest bowiem programem badawczym, nastawionym głównie na aplikacje praktyczne. Penrose i Lucas nie zgadzają się z takim podejściem. Nie podają oni jednak żadnych przekonujących argumentów za niesprzecznością umysłów. Co więcej, jak zostanie poniżej pokazane, podanie takich argumentów wydaje się być niemożliwe.

Jak wiadomo z TG2, nie można dowieść niesprzeczności dostatecznie bogatego systemu formalnego bez korzystania ze środków z metasystemu, który rozumiany jest jako system bogatszy. Gdyby zaaplikować to twierdzenie bezpośrednio do modelowania umysłu, otrzymalibyśmy twierdzenie zgodnie, z którym nie możemy dowieść niesprzeczności naszego umysłu, gdyż nie dysponujemy żadnym metasystemem zawierającym bogatsze środki. Stanisław Krajewski proponuje następujące rozumowanie wykorzystujące tezę Churcha. Gdyby istniała możliwość podania ścisłego dowodu dla niesprzeczności umysłu, dowód taki mógłby być sformalizowany, a następnie przeprowadzony przy pomocy maszyny Turinga. Maszyna ta mogłaby symulować część zdolności matematycznych, jakie posiada człowiek i dowodzić własnej niesprzeczności. Zgodnie z twierdzeniem Gödla, jej algorytm byłby jednak błędny. Tym bardziej sprzeczny byłby, więc algorytm równoważny wszystkim zdolnościom matematycznym i w ogóle całemu umysłowi. Gdy założymy możliwość dowodzenia niesprzeczności umysłu, dowodzimy więc zarazem jego sprzeczności. A zatem, nawet jeśli w istocie jesteśmy niesprzeczni, nie ma możliwości by tego dowieść[31].

Innym problemem jest sama możliwość wyrażenia niesprzeczności umysłu. Według Krajewskiego *a priori* można to zrobić na dwa sposoby: przez zdroworozsądkowe stwierdzenie wyrażone w języku niesformalizowanym lub na sposób formalny. Jeśli chodzi o przekonanie zdroworozsądkowe, trudno jest wyobrazić sobie dowód, który byłby przekonujący dla wszystkich lub przynajmniej dla większości lu-

[31] Zob. S. Krajewski, *Twierdzenie Gödla i jego interpretacje filozoficzne*, dz. cyt., s. 113 n.

dzi zainteresowanych tematyką. Poza tym trzeba by założyć istnienie nieformalnych dowodów, a to skutkowałoby automatycznie błędnym kołem w rozumowaniu: *a priori* umysłowi zostałyby przyznane niealgorytmiczne zdolności. Przy dowodzie formalnym konieczne byłoby założenie, że odpowiedni system formalny równoważny jest zdolnościom dowodowym, jakie posiada ludzki umysł. W takim wypadku działa z kolei prezentowany wyżej argument na temat niemożliwości formalnego dowiedzenia niesprzeczności umysłu[32]. Argument ten pokazuje, że nie można w sposób przekonujący wykazać, że umysł nie jest równoważny sprzecznemu systemowi formalnemu, czyli błędnemu algorytmowi. Warto wspomnieć również, że sam Kurt Gödel, który był dualistą i spirytualistą (w typologii Penrose'a zaliczyć należy go niewątpliwie do stanowiska *D*), uważał, że jego twierdzenia bez przyjęcia dodatkowych założeń matematyczno-filozoficznych nie implikują niealgorytmiczności umysłu[33].

SPRZECZNOŚCI UMYSŁU

Potoczne doświadczenie wskazuje, że ludzkie przekonania są nieraz sprzeczne. Sami zresztą często odkrywamy sprzeczności w naszych wypowiedziach i przekonaniach. Aby zrozumieć łatwiej problem sprzeczności umysłu warto przytoczyć obecnie zjawisko samooszukiwania się (*self deception*). Zwykle oszukujemy kogoś innego, niż my sami. W takim wypadku będąc sami przekonani o prawdziwości danego sądu, który ma być przedmiotem oszustwa, staramy się przekonać swojego przeciwnika do przyjęcia negacji tego sądu. Problem pojawia się w przypadku zjawiska samooszukiwania, czyli gdy się oszukujący i oszukiwany są tą samą osobą. Osoba ta twierdzi bowiem zarazem, że $p \wedge \neg p$, a zatem zgodnie z prawem logicznym ECQ ($p \wedge \neg p \rightarrow q$) z jej przekonań wynika zdanie dowolne. Z paradoksem tym poradzić można sobie przyjmując, że proces samooszukiwania przebiega nie-

[32] Zob. tamże, s. 114.
[33] Zob. tamże, s. 166-168.

świadomie[34]. W takim wypadku nie można mówić raczej o sądach w sensie logicznym. Niewątpliwie można mówić jednak o przetwarzaniu sprzecznych informacji. Proces ten przebiega faktycznie poza świadomością, jednak zachodzi na poziomie układu nerwowego. Wielu ewolucjonistów uważa, że zjawisko samooszukiwania się jest adaptacją w ewolucyjnym „wyścigu zbrojeń". Celem w tym wypadku jest jak najlepsze ukrywanie własnych oszustw i jak najskuteczniejsze oszukiwanie innych. Tłumaczyłoby to automatyzm i nieświadomość procesu samooszukiwania się[35]. Zaznaczyć należy, że posiadanie sprzecznych przekonań nie jest na pewno tym samym, co sprzeczność całego umysłu, jednak podobnie jak w przypadku systemów formalnych, pomiędzy sprzecznościami zachodzić musi istotny związek. O sprzeczności całego umysłu moglibyśmy mówić z pewnością dopiero, gdybyśmy dysponowali ostateczną teorią działania mózgu i umysłu. Na podstawie wytworów umysłu możemy stawiać jednak pewne hipotezy, co do jego funkcjonowania.

Obszarami umysłu, w których pojawiają się sprzeczności są domeny różnych pól aktywności poznawczej człowieka, takich jak nauka i religia. Z sytuacją sprzeczności między nauką i religią poradzić można sobie odrzucając któryś z sądów lub tolerując sprzeczność w pewien sposób. Jeśli chodzi o drugą z możliwości Michał Heller pisze, że:

> W skrajnych (ale to nie znaczy rzadkich) przypadkach postawa taka prowadzi do tzw. teorii dwu prawd, czyli do przekonania, iż jest rzeczą rozsądną uznawać dwa sprzeczne ze sobą zdania (lub układy zadań), pod warunkiem, że każde z nich należy do „innej dziedziny", np. do zbioru prawd religijnych i twierdzeń nauki[36].

Teoria dwu prawd, wyrażająca akceptację sprzeczności „prawd nauki" i „prawd wiary" kilka razy powracała w dziejach teologii oraz

[34]J.R. Searle, *Umysł, język, społeczeństwo*, tłum. D. Cieśla, CiS, Warszawa 1999, s. 116.

[35]Zob. M. Hohol, *Zjawisko kłamstwa w perspektywie nauk ewolucyjnych i neurokognitywnych*, „Semina Scientiarum", nr 8, 2009, s. 104 n.

[36]M. Heller, *Sens życia i Sens Wszechświata. Studia z teologii współczesnej*, Biblos, Tarnów: 2008, s. 86 n.

badań nad relacją nauka-wiara[37]. W XIII wieku Europa za pośrednictwem Arabów odkryła pisma Arystotelesa. Odkrycie to wywołało kryzys w chrześcijaństwie wspieranym tradycją platońską, rozwiniętą przez Ojców Kościoła oraz św. Augustyna z Hippony. Poglądy Arystotelesa oraz interpretatorów, takich jak Averroes w wielu miejscach były jawnie sprzeczne z religią chrześcijańską. Przykładowo, averroiści, tacy jak Siger z Brabantu oraz Boecjusz z Dacji przyjmowali naukę Averroesa, zgodnie z którą świat jest wieczny, a jednostkowe dusze są śmiertelne. Jednocześnie uznawali oni za prawdziwą naukę chrześcijańską na temat stworzenia świata oraz nieśmiertelności duszy[38]. Stosując się do doktryny dwu prawd przyjąć można by cokolwiek. Akceptując logikę klasyczną z zasadą ECQ uznać należy, że teoria dwu prawd jest nie do utrzymania, gdyż prowadzi do przepełnienia systemu, w tym wypadku poznawczego. Jak zauważa Heller, silną teorię dwu prawd przyjmować może tylko cynik, albo ktoś zupełnie niekonsekwentny w myśleniu[39]. Wydaje się, że podstawowe założenie teorii dwu prawd związane jest ze stwierdzeniem, że poszczególne sądy odnoszą się do dwóch rozłącznych obszarów semantycznych — religii i nauki, a zatem mówi się o nich w dwóch niesprowadzalnych do siebie domenach orzekania. W takim wypadku warto przytoczyć wypowiedź Józefa Marii Bocheńskiego:

> Religia jest zapisana językiem ludzkim, a więc musi podlegać prawom semantyki ludzkiej. To jest wielki błąd u teologów, którzy twierdzą, że skoro religia jest dana przez jakiś czynnik pozaświatowy, to nie stosują się do niej reguły semiotyki ludzkiej. A to nieprawda[40].

Zauważyć należy jednak, że faktycznie w przypadku naszych umysłów posiadanie sprzecznych (lub najczęściej pozornie sprzecznych)

[37]Problematyka związana z teorią dwu prawd omówiona została wyczerpująco w: B. Brożek, *The Double Truth Controversy. An Analytical Essay*, Copernicus Center Press, Kraków 2010.

[38]Zob. M. Heller, *Sens życia i Sens Wszechświata...*, dz. cyt., s. 88.

[39]Zob. tamże, s. 103.

[40]J.M. Bocheński, *Między Logiką a Wiarą. Z Józefem Marią Bocheńskim rozmawia Jan Parys*, Les Éditions Noir Sur Blanc, Warszawa 1998, s. 175.

przekonań w kwestii nauka-religia nie prowadzi wcale do przepełnienia systemu kognitywnego i generowania „czegokolwiek". Nasz umysł normatywnie nastawiony jest na unikanie sprzeczności, jednak gdy faktycznie występują, potrafi sobie z nimi radzić.

W kwestii sprzeczności, jakie pojawiają się w umyśle, wspomnieć warto po krótce na koniec również poglądy Grahama Priesta, jakie wyraził on w książce *Beyond the Limits of Thaught*[41]. Choć Priest jest logikiem (stworzył systemy logik parakonsystentnych), koncepcje formułowane w tej pracy często ocierają się o fenomenologię. Jak zauważa Robert Poczobut, Priest uznaje silne podejście dialektyczne, tj. uznaje, że sprzeczności rzeczywiście mogą realizować się w świecie, a także umyśle. Jego zdaniem treści umysłu z jednej strony wyznaczają sztywne, nieprzekraczalne ramy, z drugiej zaś, już w momencie wyznaczania tych ram, są przekraczane przez myśli[42]. Priest mówi wręcz o „prawie zachowania sprzeczności" w naszych umysłach. Gdy sprzeczność zostanie usunięta z jednego „miejsca", natychmiast pojawia się gdzieś indziej w schemacie konceptualnym[43].

KONKLUZJE

Głównym celem niniejszej pracy było zwrócenie uwagi, że pytanie o (nie)sprzeczność umysłu może być bardziej pierwotne niż pytanie o jego (nie)algorytmiczność. Jednym ze sposobów ominięcia ograniczeń nakładanych przez twierdzenie Gödla na sztuczną inteligencję jest uznanie, że umysł jest sprzecznym systemem formalnym. Rozwiązanie takie wydaje się podważać klasyczny paradygmat, zgodnie z którym umysł związany jest z logiką klasyczną ze szczególnym uwzględnieniem zasady niesprzeczności. Na gruncie współczesnej wiedzy z zakresu kognitywistyki, uznanie, że umysł jest sprzecznym systemem formalnym, wydaje się jednak równie uprawnione jak wysuwanie teorii niealgorytmicznych. Nie jest ono ponadto w żaden sposób szkodliwe

[41] G. Priest, *Beyond the Limits of Thaught*, Oxford University Press, Oxford — New York 2002.

[42] Zob. R. Poczobut, *Spór o zasadę niesprzeczności. Studium z zakresu filozoficznych podstaw logiki*, Towarzystwo Naukowe KUL, Lublin 2000, s. 387 n.

[43] Zob. tamże, s. 389.

dla praktyków sztucznej inteligencji. O wadze zagadnienia sprzeczności umysłu świadczą również „pierwszoosobowe" przykłady sprzeczności, jak opisywane powyżej zjawisko samooszukiwania się, teoria dwu prawd czy poglądy G. Priesta. Wytłumaczeniem tych problemów byłoby uznanie, że umysł dopuszcza sprzeczne dane, ponieważ sam jest układem sprzecznym.

Wydaje się, że największym problemem w naukach o umyśle jest brak teorii, która tłumaczyłaby wewnętrzne („pierwszoosobowe") stany mentalne na zewnętrzne („trzecioosobowe") parametry, czyniąc za dość wymogowi intersubiektywności, jaki stawiają nauki przyrodnicze. Choć na wskazane wyżej „pierwszoosobowe" przykłady, ze szczególnym podkreśleniem koncepcji Priesta, patrzeć należy z pewną dozą podejrzliwości, z powodu braku „twardych danych" nie mogą być one pominięte. Zwolennicy koncepcji sprzeczności umysłu, odpowiedzieć muszą na, jak się wydaje, dwa pytania. Pierwsze z nich dotyczy wyjaśnienia, dlaczego umysł jest sprzeczny, tzn. jakiego typu mechanizmy generują sprzeczność? Pytanie drugie koncentruje się natomiast wokół rewizji logiki. Skoro, aby opisać umysł konieczny jest system odrzucający ECQ, jaka logika w takim razie zastąpić powinna logikę klasyczną? Wydaje się, że odpowiedzi na pierwsze z pytań próbować można udzielić w oparciu o Modularną Teorię Umysłu funkcjonującą m.in. w ramach psychologii ewolucyjnej. W takim ujęciu poszczególne moduły składające się na umysł byłyby niesprzeczne, jednak ze względu na niezgodności pomiędzy nimi, umysł, jako całość byłby sprzeczny. Natomiast, jeśli chodzi o rachunek, który zastąpić mógłby logikę klasyczną, można by dokonać tego w oparciu o którąś z logik parakonsystentnych.

LITERATURA CYTOWANA

P. Benacerraf, *God, the Devil and Gödel*, „The Monist", 51, 1967, ss. 9–32, także dostęp online [29.06.2010]: <http://www2.units.it/~etica/ 2003_1/3_monographica.htm>.

J.M. Bocheński, *Między Logiką a Wiarą. Z Józefem Marią Bocheńskim rozmawia Jan Parys*, Les Éditions Noir Sur Blanc, Warszawa 1998.

B. Brożek, *The Double Truth Controversy. An Analytical Essay*, Copernicus Center Press, Kraków 2010.

P.M. Churchland, P.S. Churchland, *On the contrary. Critical essays 1987-1997*, MIT, Boston 1998.

P.S. Churchland, T.J. Sejnowski, *The Computational Brain*, MIT Press, Cambridge — London 1996.

S.M. Downes, *Evolutionary Psychology*, [w:] *The Stanford Encyclopedia of Philosophy*, red. E.N. Zalta, dostęp online [29.06.2010]: <http://plato.stanford.edu/ entries/ evolutionary-psychology/>.

W.P. Grygiel, M. Hohol, *Rogera Penrose'a kwantowanie umysłu*, „Filozofia nauki", XVII, nr 3(67), 2009, s. 5–31.

M. Heller, *Przeciw fundacjonizmowi*, [w:] tenże, *Filozofia i Wszechświat*, Znak, Kraków 2006, s. 82–101.

M. Heller, *Sens życia i Sens Wszechświata. Studia z teologii współczesnej*, Biblos, Tarnów 2008.

M. Hohol, *Zjawisko kłamstwa w perspektywie nauk ewolucyjnych i neurokognitywnych*, „Semina Scientiarum", nr 8, 2009, s. 91–109.

S. Krajewski, *Twierdzenie Gödla i jego interpretacje filozoficzne. Od mechanicyzmu do postmodernizmu*, Warszawa: IFiS PAN 2003.

J.R. Lucas, *Minds, Machines and Gödel*, „Philosophy", vol. XXXVI, 1961, ss. 112–127. Dostępny w języku polskim: J.R. Lucas, *Umysły, Maszyny i Gödel*, tłum. M. Zawidzki, „Hybris — internetowy magazyn filozoficzny", nr 8 (2009), dostęp online [15.05.2010]: <http://www.filozof.uni.lodz.pl/ hybris/ pdf/ h09/ 6.%20Lukas2%20[7498].pdf>.

W. Marciszewski, *Sztuczna inteligencja*, Znak, Kraków 1998.

D. McDermott, *[STAR] Penrose is Wrong. A Review of Shadows of the Mind by Roger Penrose*, „Psyche. An interdisciplinary

journal of research on consciousness", vol. 2, 1995, 9.10, dostęp online [15.05.2010]: <http://www.theassc.org/ files/ assc/ 2335.pdf>.

R. Penrose, *Cienie umysłu. Poszukiwanie naukowej teorii świadomości*, tłum. P. Amsterdamski, Zysk i S-ka, Poznań 2000.

R. Penrose, *Droga do rzeczywistości. Wyczerpujący przewodnik po prawach rządzących Wszechświatem*, tłum. J. Przysława, Prószyński i S-ka, Warszawa 2006.

R. Penrose, *Nowy umysł cesarza. O komputerach umyśle i prawach fizyki*, tłum. P. Amsterdamski, PWN, Warszawa 2000.

R. Poczobut, *Spór o zasadę niesprzeczności. Studium z zakresu filozoficznych podstaw logiki*, Towarzystwo Naukowe KUL, Lublin 2000.

G. Priest, *Beyond the Limits of Thaught*, Oxford University Press, Oxford — New York 2002.

J.R. Searle, *Umysł, język, społeczeństwo*, tłum. D. Cieśla, CiS, Warszawa 1999.

J.R. Searle, *Umysł na nowo odkryty*, tłum. T. Baszniak, PIW, Warszawa 1999.

SUMMARY

TOWARDS THE CONSISTENCY OF AN INCONSISTENT MIND

The common sense conviction that rationality is based on the classical logic requires major revision since the essential assumption of many standpoints in the cognitive science, concerning the non-contradictory character of mind, seems to be no longer tenable. Firstly, the non-algorithmic models of mind proposed by John Lucas and Roger Penrose are presented. In the context of these models, the importance of the Gödel incompleteness theorems for the philosophy of mind and artificial intelligence is debated. Secondly, several specific difficulties in applying the 'Gödelian arguments' in the modeling of mind are pointed out. As the main thesis of the article, it is stipulated that mind operates according to a wrong algorithm that is functionally equivalent to a contradictory formal system. The examples of the contradictory contents of mental states, evidenced in the phenomenon of self-deception and the me-

diaeval double truth theory in science, are discussed. Some consequences of the model of an inconsistent mind, based on the revision of the classical logic, are surveyed.