Józef Dębowski

Pułapki komputacjonizmu

Filozofia Nauki 12/1, 29-50

2004

Filozofia Nauki
Rok XII, 2004, Nr 1(45)

Józef Dębowski

Pułapki komputacjonizmu[1]

Wstęp I. W XXI wiek i trzecie tysiąclecie wkroczyliśmy pełni niepokojów, ale też niepozbawieni nadziei. Być może nie ma w tym nic dziwnego, być może tak było zawsze, zwłaszcza zaś wówczas, gdy ludzkość czy inna mniejsza wspólnota cywilizacyjna, przekraczała jakąś granicę — granicę choćby czysto konwencjonalną (wyznaczoną np. kalendarzem) czy zgoła wyimaginowaną, urojoną, magiczną, a więc niekoniecznie rzeczywistą. Dzisiaj jednak dobrze wiemy, i m.in. na tym polegałaby odmienność (osobliwość) naszej obecnej sytuacji w stosunku do wszystkich minionych tego typu, że ważnym źródłem wielu naszych lęków i nadziei (zarówno frustracji, jak i krzepiących uniesień) były i są rozmaite dwudziestowieczne mistyfikacje (legendy, mity): poznawcze, technologiczne, społeczne i cywilizacyjne. Nie będę ich katalogował, kategoryzował czy hierarchizował — być może jeszcze na to za wcześnie. Zatrzymam się tedy tylko przy jednej z nich, mianowicie przy komputacjonizmie. Naturalnie, nie można twierdzić (i ja wcale tak nie twierdzę), że w całości jest on czymś niespełnionym, złudnym albo oszukańczym. Bez wątpienia jednak dał mniej niż obiecywał i wcale nie to, co obiecywał — przynajmniej pod pewnym względem. Wszelako, z drugiej strony i pod innym względem, zarazem też dał więcej niż obiecywał, choć, ponownie, raczej nie jest to dokładnie to, co obiecywał.

Wstęp II. Niekiedy, choć raczej rzadko, zdarza się, że tuż przed wykładem przerażeni uświadamiamy sobie, iż długo i pracowicie przygotowywany tekst wykładu gdzieś się zawieruszył, został zagubiony, a nawet uległ bezpowrotnemu zniszczeniu. Jeśli rozpacz nas nie obezwładni, próbujemy wybrnąć z kłopotu w ten sposób, że na

[1] Niniejszy tekst stanowił podstawę referatu, który dnia 24.09.2002 r. wygłosiłem w czasie konferencji zorganizowanej przez Wydział Filozofii i Socjologii UMCS pod nazwą „Filozofia wobec XXI wieku". Ponieważ opublikowanie materiałów konferencyjnych stanęło pod znakiem zapytania, o wydrukowanie tego tekstu poprosiłem Redakcję „Filozofii Nauki".

gorąco, wielkim nakładem energii, mocno skoncentrowani próbujemy zrekonstruować opracowane uprzednio wyniki i wcześniej przygotowany tekst. Nie zawsze, ale czasami ponoć się to udaje. Ba, bywa, że właśnie w takich (nadzwyczajnych) okolicznościach powstają wykłady najwspanialsze, najbardziej dramatyczne i najbardziej odkrywcze — wykłady długo potem wspominane przez studentów i samych wykładowców. Niestety, mnie nic takiego się nie zdarza. Co to znaczy? To znaczy, że tekstu wykładu nigdy nie zapominam i nie gubię, i że wobec tego, podobnie jak wszystkie inne moje wykłady, także dzisiejszy referat będzie długi i nudny. Ale... Ale być może można temu jakoś zaradzić. Na przykład, mogę spróbować czytać co drugie zdanie, co trzecie słowo, co czwarty akapit albo też pomijać wszystkie wyrazy pięcioliterowe lub zaczynające się na „k", „ch" i „p". Bez trudu mogę też powierzyć spreparowanie takiego tekstu swojej obliczeniowej maszynie, czyli osobistemu komputerowi. Nie wątpię, że zadanie skrupulatnie wykona. Czy jednak równie skutecznie wykona też jeszcze jedno zadanie — zadanie całkiem podobne do poprzedniego, z tą tylko różnicą, że tym razem każę mu pomijać pewne pojęcia i myśli, np. słowa i zdania o treści erotycznej i/lub antysemickiej albo też wszystkie słowa i zdania wzbudzające tęsknotę za minionym latem, wywołujące odrazę, przywołujące zielone wzgórza Afryki i kojarzące się jednoznacznie ze smakiem dojrzałej czereśni? Mam wrażenie, że mój nastoletni syn znakomicie wywiąże się z tego zadania — przynajmniej zasadniczo. A co na to mój komputer?

1. By nie utonąć w szczegółach, niech naszym pytaniem wyjściowym będzie dobrze znane pytanie Alana Turinga sprzed przeszło półwiecza: „czy maszyny mogą myśleć?"[2] Wszak, rozprawiając o komputacjonizmie, nie można dzisiaj zacząć inaczej. Jak wiadomo, odpowiedź samego Turinga na to pytanie była odpowiedzią twierdzącą[3] — oczywiście, pod warunkiem, że każda taka maszyna pomyślnie przejdzie procedurę sprawdzającą, zwaną przez pomysłodawcę „grą w udawanie", czyli, jak dzisiaj mówimy, „test Turinga".[4] Przypomnę, iż w procedurze tej rozstrzygająca jest okoliczność, czy maszyna, postawiona dokładnie wobec tych samych zadań (pytań), wobec których równolegle postawiony jest konkretny człowiek, poradzi sobie z nimi równie skutecznie (lub porównywalnie), co ów człowiek — skutecznie do tego stopnia, że rozwiązania (odpowiedzi) maszyny i człowieka pozostaną praktycznie nieodróżnialne. Ważne przy tym jest także to, że gdy Turing pisał o maszynie zdolnej zastąpić człowieka w obmyślonej przez siebie „grze w udawanie", to miał na

---

[2] A. M. Turing, *Maszyna licząca a inteligencja*, przeł. M. Szczubiałka, [w:] *Filozofia umysłu*, wybrał i wstępem opatrzył B. Chwedeńczuk, Fundacja ALETHEIA — Wydawnictwo Spacja, Warszawa 1995, s. 271 i n. Por. też pierwodruk: A. M. Turing, *Computing Machinery and Intelligence*, „Mind", 1950, nr 236.

[3] „Należy oczekiwać, pisał Turing, że maszyny będą w końcu rywalizować z ludźmi we wszystkich czysto intelektualnych dziedzinach". A. M. Turing, *Maszyna...*, s. 299. Por. też: tamże, s. 281. Naturalnie, Turing był również przekonany, iż maszyny cyfrowe będą w stanie same się uczyć. Tamże, s. 293 i n.

[4] Tamże, s. 271-280.

myśli komputer cyfrowy, złożony z trzech zespołów: (1) zespołu rejestrów do przechowywania informacji oraz listy odpowiednich instrukcji (pamięć); (2) jednostki wykonującej poszczególne operacje, np. obliczenia; (3) zespołu sterującego wykonywaniem operacji, nadzorującego ich przebieg i kolejność, dzisiaj zwanego krótko programem.[5]

Pytanie postawione przez Turinga, nade wszystko zaś jednoznaczna i zdecydowana odpowiedź na nie (oraz jej uzasadnienie), wzbudziło w latach pięćdziesiątych falę wielkiego zainteresowania możliwościami „myślących maszyn", a także perspektywą ich zbudowania i zastosowania. To zapewne ów wzrost zainteresowania, efektywnie podsycany przez rozwój ówczesnej nauki i techniki (elektronika, elektrofizjologia, cybernetyka, informatyka), doprowadził w 1956 roku do znamiennego spotkania w Dartmouth College (New Hampshire), którego uczestnikami byli m.in.: John McCarthy (główny inicjator spotkania i twórca terminu „sztuczna inteligencja"), Marvin Minsky, Herbert Simon i Allen Newell. W tym samym roku, tym razem w Massachusetts Institute of Technology, odbyło się jeszcze jedno naukowe sympozjum, w którym, poza przedstawicielami nauk ścisłych (A. Newell, H. Simon), uczestniczyli także psychologowie (J. Brunner, G. Miller, M. Posner) i językoznawcy (N. Chomsky). Wymienione dwa spotkania z 1956 roku uznaje się dzisiaj za kluczowe dla zapoczątkowania systematycznych badań nad tzw. sztuczną inteligencją (*Artifical Intelligence*; w skrócie AI) oraz dla narodzin pierwszych interdyscyplinarnych programów badawczych z zakresu tzw. nauk kognitywnych (*Cognitive Science*) — kluczowe dla, jak krótko można powiedzieć, komputacjonizmu i kognitywistyki. Tematyka i atmosfera odbytej debaty sprawiły też, że mocno powiało optymizmem, gdy chodzi o możliwości konstruowania „myślących maszyn" — maszyn wykorzystywanych następnie do rozwiązywania problemów, które wymagają od człowieka znacznego wysiłku intelektualnego, umiejętności analitycznego myślenia, intelektualnej dyscypliny, precyzji, a zarazem posiadają dużą doniosłość praktyczną i poznawczą.

Przy tym, wielu badaczom niemal od początku wydawało się naturalne to, że — zgodnie z myślą przewodnią „testu Turinga" — skoro maszyna jest w stanie zastąpić człowieka tam, gdzie musi się on wykazać znaczną inteligencją, to nie ma **żadnych** podstaw po temu, by maszynie (i tylko dlatego, że jest maszyną) tej inteligencji odmawiać. W ten sposób, a więc niemal bezwiednie (choć zgodnie z pierwotną intencją Turinga), doszło do wytworzenia i upowszechnienia czysto **funkcjonalnego** pojęcia inteligencji i, generalnie, czysto funkcjonalnego pojęcia myślenia. Zgodnie z tym (funkcjonalnym) podejściem, „inteligentnym" albo „myślącym" jest każdy taki przedmiot albo system przedmiotów, który — całkiem niezależnie od tego, co go stanowi (człowiek, maszyna czy jakikolwiek inny przedmiot fizyczny) — jest w stanie **inteligentnie działać**, tj. gwarantować osiąganie efektów, które wymagają od człowieka podjęcia działań inteligentnych, czytaj: zwykle uważanych za inteligentne (=

[5] Tamże, s. 275 i n.

rozumne = angażujące myślenie).[6] Zauważmy od razu, a nawet spróbujmy to specjalnie podkreślić, że z czysto **funkcjonalnego** punktu widzenia o istocie inteligencji, a także „myślenia" czy „rozumności", bez reszty przesądzają określone wyniki działania. Nieistotne są zaś wszystkie pozostałe czynniki, a więc m.in. substrat tego działania (podłoże, na którym się ono dokonuje), jego tworzywo, ba, na dobra sprawę nieistotne są także czas i kierunek działania oraz jego struktura (mechanizm).

2. Skoro zatem, wykorzystując wyżej przywołane ujęcie inteligencji i myślenia w ogóle, ponad wszelką wątpliwość stwierdzono, że myśli nie tylko człowiek albo zwierzę, lecz także myślą (w wyżej ustalonym sensie) również maszyny, to po to, by nareszcie odpowiedzieć na pytanie o istotę czy naturę **ludzkiego** myślenia, wystarczy dokładnie przeanalizować operacje, jakie wykonują myślące maszyny. Perspektywa była niezwykle ponętna i spośród grona pierwszych entuzjastów sztucznej inteligencji (*Artificial Intelligence*, w skrócie: AI) nikt bodaj nie był w stanie jej się oprzeć. Wszak, jak sobie obiecywano, dzięki programom badawczym AI i praktycznym możliwościom konstruowania maszyn cyfrowych nareszcie została otwarta droga do zrozumienia, drobiazgowego opisania i wyczerpującego wyjaśnienia **ludzkiego** myślenia, **ludzkiej** aktywności umysłowej i, w ogólności, **ludzkiego umysłu.** A więc to, co przez tysiąclecia było niezrozumiałe, zagadkowe i niewyjaśnialne, za sprawą powstania maszyn cyfrowych i programów badawczych AI, nagle mogło się stać przeraźliwie proste, czytelne, zrozumiałe i niemal w pełni wyjaśnialne. Prócz wcześniej przyjętych założeń wystarczyło jeszcze tylko postawić kropkę nad „i" i wyraźnie stwierdzić, że również **ludzki mózg** — podobnie jak każdy inny układ fizyczny, w tym nawet wszechświat jako całość — działa dokładnie tak jak maszyna cyfrowa.

Analogia między ludzkim mózgiem i komputerem sama zresztą narzucała się już w momencie, gdy została przyjęta i szerzej uznana czysto funkcjonalna definicja inteligencji (= myślenia = rozumności), w szczególności, gdy szeroko i bez zastrzeżeń zaakceptowali ją również psychologowie, psychiatrzy i filozofowie umysłu.[7] W miarę jak postępował rozwój badań nad tymi maszynami i rosły techniczne możliwości ciągłego ich doskonalenia, analogię tę należało już tylko bliżej konkretyzować, uzupełniać i uściślać. Istotnie, od dawna czyni się to na wiele różnych sposobów. Przy tym, podstawowa idea tej analogii i różnych jej wersji pozostaje niezmiennie ta sama: **umysł jest dla mózgu tym, czym program dla komputera.** Idea ta jest też bodaj najważniejszą tezą teorii sztucznej inteligencji w jej **wersji silnej** (*strong AI*) — najważniejszą, choć nie jedyną specyficzną dla tej wersji. Niekiedy uważa się także, iż zarazem jest to teza wszelkiego **kognitywizmu**, a więc poglądu, który ukształtował

[6] To czysto funkcjonalne podejście do myślenia jest wyraźnie zaznaczone przez M. Minsky'ego już w 1962 roku. M. Minsky, *Na drodze do stworzenia sztucznej inteligencji*, [w:] *Maszyny matematyczne i myślenie*, red. E. Feigenbaum, J. Feldman, PWN, Warszawa 1972, s. 420.

[7] Jak wiadomo, szczególna rola w upowszechnianiu tej analogii, podobnie zresztą jak i tzw. funkcjonalnej koncepcji umysłu, przypada psychologowi-behawioryście Hilaremu Putnamowi. Do rozwijanej od wczesnych lat sześćdziesiątych XX wieku koncepcji Putnama (i jego pism) będę wielokrotnie nawiązywał na dalszych stronach.

się w najnowszej filozofii umysłu, psychologii, neurofizjologii i matematyce pod wpływem, z jednej strony, tzw. nauk kognitywnych, z drugiej zaś, teorii sztucznej inteligencji w jej wersji silnej.[8]

Z tego, co przed chwilą powiedziałem, wynika jednak, iż tzw. silna wersja teorii sztucznej inteligencji nie jest jej jedynym wariantem. Rzeczywiście. Idąc za odpowiednimi rozróżnieniami Johna R. Searle'a z roku 1980, prócz **silnej** AI należy jeszcze wyróżnić tzw. **słabą** albo „złagodzoną" („ostrożną") wersję teorii sztucznej inteligencji.[9] Z kolei nawiązując do innych znanych rozróżnień, rozróżnień zaproponowanych przez Rogera Penrose'a, istnieją cztery główne odmiany teorii sztucznej inteligencji. Penrose oznacza je literkami A, B, C i D, przy czym porządek alfabetyczny ma tu ilustrować stopień uskrajnienia tych teorii — od najwyższego po najniższy, a to ze względu na stopień algorytmizacji procesów umysłowych (*scil.* stosunek algorytmu do świadomości).[10] Ponieważ rozróżnienia te są dobrze znane, nie będę ich tu przywoływał i bliżej analizował.[11] W zamian zaś chciałem bliżej rozważyć skrajną wersję teorii sztucznej inteligencji i niektóre jej implikacje. Mówiąc krótko, teoria ta głosi, że „mózg jest rodzajem komputera, umysł zaś rodzajem komputerowego programu".[12] Zatem, by odpowiedzieć na pytanie, czym jest umysł, wystarczy poznać zasadę działania programów komputerowych. Spróbujmy zbadać tę możliwość i ewentualnie zdać sobie sprawę z jej ograniczeń.

---

[8] W sprawie, czym jest kognitywizm i jaki jest jego stosunek do teorii sztucznej inteligencji i nauk kognitywnych, por. J. R. Searle, *Umysł, mózg i nauka*, przeł. J. Bobryk, Wydawnictwo Naukowe PWN, Warszawa 1995, s. 38-51.

[9] Polskojęzycznemu rozróżnieniu na **silną** (albo mocną) i **słabą** (albo złagodzoną, ostrożną) wersję teorii sztucznej inteligencji w języku oryginału odpowiadają zwroty *Strong Artificial Intelligence* (w skrócie: SAI) i *Weak Artificial Intelligence* (w skrócie WAI). J. R. Searle, *Minds, Brains, and Programs*, „The Behavioral and Brain Sciences", 1980, vol. 3, s. 417. Por. też polski przekład: B. Chwedeńczuk, *Umysły, mózgi i programy*, w: *Filozofia umysłu...*, s. 301 i n. Warto może zauważyć, że przymiotniki „silna" i „słaba" raz odnoszone są do różnych teorii sztucznej inteligencji lub różnych programów badawczych nad sztuczną inteligencją, innym zaś razem wprost do sztucznej inteligencji. Wydaje się, iż, ściśle biorąc, poprawny jest jedynie pierwszy sposób ich użycia, natomiast drugi można uznać za rodzaj wygodnego skrótu (i chyba nic więcej, w innym bowiem razie doszłoby do ewidentnego pomieszania poziomów języka). W niniejszym referacie stosuję obie konwencje, pamiętając wszelako, że druga jest wyłącznie krótszym odpowiednikiem pierwszej (właściwej). Podobne pomieszanie poziomów języka (i supozycji) ma miejsce w używaniu zwrotu „sztuczna inteligencja" — *Artificial Intelligence* (w skrócie: *AI*). Za autorami anglojęzycznymi przyjmijmy jednak, że kontekst, w którym się one pojawiają jest wystarczająco przejrzysty i pozwala skutecznie uniknąć ewentualnych nieporozumień.

[10] R. Penrose, *Shadows of the Mind. A Search for Missing Science of Consciousness*, Oxford University Press, Oxford 1994, s. 12 (*sub.* 1.3.)

[11] Zwięzłe omówienie wymienionych rozróżnień Penrose'a można znaleźć w: W. Marciszewski, *Sztuczna inteligencja*, wyd. I, Społeczny Instytut Wydawniczy „Znak", Kraków 1998, s. 15-23.

[12] J. R. Searle, *Umysł...*, s. 25.

3. Jak wiadomo, istota i sposób działania komputera zostały dobrze opisane już przez Alana Turinga. Nawiązujemy do tego opisu zawsze wtedy, gdy dla jakiś celów charakteryzujemy działanie tzw. „maszyny Turinga" (zwykłej lub uniwersalnej) — działanie, które polega na wykonywaniu określonych **algorytmów**, oczywiście, dzisiaj powiemy raczej: programów komputerowych.[13] To po pierwsze. Po wtóre zaś, istotę operacji, jakie wykonuje każdy komputer cyfrowy, stanowi to, że wszystkie one (bezwyjątkowo) mają charakter czysto **formalny** — polegają na operowaniu **symbolami**, najczęściej symbolami liczbowymi, którym ostatecznie odpowiadają określone sekwencje zer i jedynek.

Jak sądzę, te dwie charakterystyki w zupełności już wystarczają, by dostrzec istotne ograniczenia analogii komputerowej — ograniczenia na tyle poważne, by móc ją zakwestionować, a może i odrzucić. W czym rzecz?

Istotę pierwszego ograniczenia widać chyba najlepiej w świetle *I twierdzenia Gödla o niezupełności*. Jak może warto przypomnieć, twierdzenie to głosi, że w każdym bogatszym i niesprzecznym systemie logicznym zawierają się zdania (zdania poprawnie zbudowane i dające się, podobnie jak i pozostałe zdania systemu, arytmetycznie zakodować), które za pomocą środków tego systemu nie mogą być ani udowodnione, ani obalone.[14] Doniosłość tego twierdzenia dla metamatematyki i filozofii nauki trudno jest dziś przecenić. Podobnie zresztą jak i doniosłość innych współbrzmiących z nim twierdzeń metamatematycznych — twierdzeń niekiedy zwanych dzisiaj **limitacyjnymi**.[15] Zapytajmy jednak, jaką wagę ma to twierdzenie dla sposobu działania komputerów i dla nadziei, jakie z działaniem komputerów wiążą przedstawiciele silnej AI.

Jak myślę, podstawowy sens *I twierdzenia Gödla o niezupełności* — wprawdzie w sposób dość swobodny, ale chyba bez poważniejszych zniekształceń — można wyrazić również w sposób następujący. Nie wszystko, co da się poprawnie i niesprzecznie pomyśleć albo wypowiedzieć (w języku konkretnego systemu), jest w pełni formalizowalne i algorytmizowalne. Znaczy to, że są problemy i rozstrzygnięcia, wobec których algorytm jest bezradny i na które, by się tak wyrazić, jest „głuchy" i „ślepy". Zresztą, gwoli sprawiedliwości odnotujmy tu, iż z wymienionej trudności dobrze zdawał sobie sprawę także A. M. Turing. Mówi o tym tzw. *twierdzenie Turinga*, które

---

[13] A. M. Turing, *Maszyna...*, s. 273-280.

[14] *Twierdzenia o niezupełności* (zasadniczo jest ich dwa) zostały przez K. Gödla sformułowane i w pewien szczególny sposób udowodnione w 1931 roku w artykule (rozprawie) pt. *Über formal unentscheidbare Sätze der „Principia Mathematica" und verwandter Systeme*, „Monatshefte für Mathematik und Physik", Bd. XXXVIII 1931, s. 173-198. Ich szczegółowe omówienie można znaleźć m.in. w: E. Nagel, J. R. Newman, *Twierdzenia Gödla*, przeł. B. Stanosz, PWN (Seria „Omega"), Warszawa 1966. Por. też J. Dębowski, *Świadomość, poznanie, naoczność poznania*, Wydawnictwo UMCS, Lublin 2001, s. 110-123.

[15] W sprawie twierdzeń limitacyjnych współczesnej metamatematyki oraz ich znaczenia dla filozofii nauki por. J. Woleński, *Metamatematyka a epistemologia*, Wydawnictwo Naukowe PWN, Warszawa 1993, zwł. s. 73-85.

w odniesieniu do maszyn cyfrowych (zarówno prostych, jak i uniwersalnych) głosi, iż maszyna nie może wykonać niektórych dobrze sprecyzowanych działań; niezależnie od tego, ile czasu damy jej do dyspozycji, będzie albo odpowiadała błędnie, albo też nie będzie odpowiadała wcale.[16] Sam Turing jednak, z powodów całkowicie dla mnie niezrozumiałych, wymienioną trudność zlekceważył, a przynajmniej istotnie umniejszył jej wagę.[17]

Tymczasem nie jest to trudność błaha. Jest to raczej trudność zasadnicza, albowiem jej przezwyciężenie w żadnej mierze nie jest zależne od czasu, stopnia zaawansowania programów komputerowych lub poziomu technicznego urządzeń. Przy czym rzecz nie polega na tym, że są to problemy nigdy przez nikogo w żadnej sytuacji i w żaden sposób nierozwiązalne (nierozwiązalne niejako z ich własnej istoty, a więc na mocy definicji), lecz raczej na tym, że są to problemy nierozwiązalne tylko wtedy, gdy działa się w pewien ściśle określony sposób — mianowicie w taki sposób, w jaki działa i działać może komputer, a więc na drodze wykonywania jakichś algorytmów (programów). Wszelako, jak się zdaje, możliwości działania (i myślenia) w inny sposób, niż oparty na algorytmach (programach), zwolennicy silnej wersji teorii sztucznej inteligencji w ogóle nie biorą w rachubę — nie biorą w rachubę zarówno w przypadku „myślącego człowieka", jak i w przypadku „myślących maszyn" (komputerów). Tak zresztą, jak i nie biorą tego pod uwagę również ci wszyscy, dla których jedyną istotną funkcją ludzkiego rozumu jest jedynie jego funkcja **porządkująca**.[18]

4. Wymieniony kłopot, choć uparcie ignorowany, istnieje w teorii sztucznej inteligencji niemal od jej poczęcia. Jest też kłopotem, który (przynajmniej spośród tych najbardziej zasadniczych) najwcześniej został dostrzeżony przez krytyków AI, mianowicie już w roku 1961.[19] Można podejrzewać, iż jego przemyślenie uwolniło falę dalszej i znacznie już śmielszej krytyki. Dobrym jej przykładem jest stanowisko Huberta L. Dreyfusa wyłożone po raz pierwszy w roku 1972 w książce pt. *What Computers Can't Do. The Limits of Artificial Intelligence.*[20] Hubert D. Dreyfus, jako bodaj pierwszy, próbował wykazać całkowitą bezpodstawność analogii komputerowej, podobnie jak i brak jakichkolwiek podstaw do ekstrapolacji prawidłowości fizycznych na sferę przeżyć psychicznych i jej wytworów. Odwołując się m.in. do analiz feno-

---

[16] A. M. Turing, *Maszyny*..., s. 283.

[17] Tamże, s. 283-284.

[18] Tezę o wyłącznie porządkującej funkcji rozumu (jako jedynej jego funkcji istotnej) znajdziemy dzisiaj m.in. w: Z. Cackowski, *Rozum między chaosem a „Dniem Siódmym" porządku*, „Kategorie Ludzkiego Doświadczenia" (2), Wydawnictwo UMCS, Lublin 1997. Jest to bodaj główna teza tej książki.

[19] Por. J. R. Lucas, *Minds, Machines and Gödel*, „Philosophy", 1961, vol. XXXVI No. 137, s. 112-127.

[20] Wymieniona książka H. L. Dreyfusa miała wiele wydań, ciągle przez autora udoskonalanych i wzbogacanych. Ostatnie jej wydanie, z roku 1992, ma także nieco zmieniony (pod-) tytuł. Por. H. L. Dreyfus, *What Computers Can't Do. A Critic of Artificial Reason*, Harper & Row Publishers, New York 1992.

menologicznych, wskazał, iż w obrębie aktywności umysłowej człowieka, prócz elementów formalizowanych i algorytmizowalnych, efektywnie współwystępują i działają elementy opierające się formalizacji, a nawet umykające jakimkolwiek prawom kojarzenia czy postaciowania. Wystarczy wszak zauważyć, że o trafnym rozwiązaniu rozmaitych problemów częstokroć przesądza spontaniczna intuicja, swobodne skojarzenie, nieprzewidywalny kontekst sytuacyjny, jakaś treść nieistotna lub uświadamiana peryferyjnie. Znaczenie wymienionych czynników trudno również przecenić w prowadzeniu rozmaitych gier, w przekładach językowych (zwłaszcza języków naturalnych) oraz w obrazowaniu i rozpoznawaniu obrazów. Łącząc metody fenomenologii z analizami pojęciowo-lingwistycznymi, H. L. Dreyfus rozwijał również tzw. **adwerbialną** teorię percepcji. Utrzymywał w niej, że treści naszych doznań zmysłowych często mają charakter prekategorialny i antepredykatywny, *scil.* są kategorialnie nieuformowane, przynajmniej w jakimś stopniu i przynajmniej te z obrzeży pola percepcji. Z tego między innymi powodu najbardziej odpowiednią formą językowego ich wyrażenia jest **forma przysłówkowa**. Dla wyrażenia treści perceptywnych wydaje się ona najbardziej właściwa także dlatego, że pozwala uniknąć pytania o przedmiotowe odniesienie tych treści, a tym samym umożliwia wyeliminowanie z języka pojęcia „zjawiska". Ogólnie można więc powiedzieć, iż forma przysłówkowa ma ten walor, że — w odróżnieniu od innych, zwłaszcza zaś propozycjonalnej (wyrażenia językowe *de dicto*) — nie zakłada ani nie implikuje żadnych tez metafizycznych.[21]

Oczywiście, punkt widzenia Huberta L. Dreyfusa — jego wszechstronny opis procesów umysłowych, jego krytyka różnych postaci redukcjonizmu i mechanicyzmu, wreszcie jego argumentacja na rzecz podejścia holistycznego i metodologicznie elastycznego — bynajmniej nie ostudziła zapału komputacjonistów. Wydaje się nawet, że wprost przeciwnie. W dyskusjach spowodowanych wystąpieniem Dreyfusa komputacjoniści, powołując się na sukces technologiczny (który rozgrzeszał z wszystkiego i legalizował wszystko), podkreślali pełną zasadność dokonanych uproszczeń i jeszcze dobitniej formułowali swoje poglądy. Niektórzy z nich (np. H. Simon, A. Newell lub J. McCarty) nie wahali się twierdzić, że maszyny cyfrowe myślą już w sensie dosłownym i, podobnie jak ludzie lub zwierzęta, mają swoje przekonania.[22] Ba, według J. McCarty'ego, przekonania ma już termostat. Bardzo często eksponowano też istotne przewagi (m.in. w perspektywie ewolucyjnej) procesów komputacyjnych maszyny nad procesami myślowymi człowieka i komputera nad ludzkim mózgiem. Na fali gwałtownie rosnącego entuzjazmu lat siedemdziesiątych, większość twórców i zwolenników silnej wersji AI nie mogła się oprzeć również temu, by nie określać bliższej lub dalszej perspektywy czasowej, w której już niechybnie dojdzie

[21] Jednak za inicjatora i głównego rzecznika adwerbialnej teorii percepcji uchodzi R. M. Chisholm. Zapoczątkował ją już w latach pięćdziesiątych ub. stulecia. Więcej informacji na temat tej teorii zob. w: R. M. Chisholm, *Perceiving. A Philosophical Study*, Cornell University Press, Ithaca (N.Y.) 1957. Por. też R. M. Chisholm, *Teoria poznania*, przeł. R. Ziemińska, Instytut Wydawniczy „Daimonion", Lublin 1994, s. 82-95.

[22] J. R. Searle, *Umysł...*, s. 26-27.

do stworzenia maszyny (lub przynajmniej programu) znacznie przewyższającej poziom ludzkiej inteligencji i ludzkich sprawności umysłowych.[23] Przy tym dalej konsekwentnie utrzymywano, iż istota wszelkiego rozumienia i wszelkiej inteligencji (tak ludzkiej, jak i komputerowej) zasadniczo polega na jednym — na zdolności manipulowania symbolami, a więc, nieco innymi słowy, na wykonywaniu operacji poddających się pełnej formalizacji.

4. To, że w żadnym razie tak nie jest i w żadnym razie tak być nie może — poniekąd wynika już z twierdzeń Gödla i ewentualnie pozostałych twierdzeń limitacyjnych współczesnej metamatematyki.[24] Jednak w sposób bezprecedensowo przejrzysty i bardziej bezpośredni zostało to okazane dopiero w roku 1980, a autorem tego przedsięwzięcia jest kalifornijski filozof John R. Searle. Chodzi oczywiście o argument zwany krótko „chińskim pokojem" — argument słynny dzisiaj w takim stopniu, że może nawet nie byłoby specjalnej przesady, gdyby nazwać go „klasycznym". Po raz pierwszy został on przedstawiony przez Searle'a w artykule pt. *Umysły, mózgi i programy*.[25] Potem, przez całą dziewiątą dekadę ubiegłego wieku, był przez J. R. Searle'a wielokrotnie modyfikowany i rozbudowywany — zapewne wskutek ożywionej dyskusji, jaką wówczas wywoływał.[26] Krótko i w wielkim uproszczeniu można go streścić w sposób następujący. Gdyby człowieka, który (jak np. J. R. Searle) nie rozumie ani słowa w języku chińskim, umieścić w zamkniętym pokoju, a następnie przez wąską szczelinę podrzucać mu poszczególne znaki alfabetu chińskiego wraz z anglojęzyczną instrukcją na temat tego, co ma z nimi robić, to w rezultacie skrupulatnego wykonania odpowiednich instrukcji człowiek ów byłby w stanie „opowiedzieć" w języku chińskim każdą dowolną historię — byłby w stanie „opowiadać" te historie w języku, z którego nie znał wcześniej żadnego znaku i w którym uprzednio nie rozumiał żadnego słowa.[27]

---

[23] Pokusie tej zresztą nie umiał się oprzeć już A. M. Turing, tak jak dzisiaj nie umie się jej oprzeć na przykład M. Minsky, który utrzymuje, że już najbliższa generacja komputerów będzie w takim stopniu inteligentna, iż winniśmy być szczęśliwi, jeśli pozostawią nas w domach w charakterze domowych zwierzątek. Por. J. R. Searle, *Umysł*..., s. 27.

[24] R. Penrose, *Nowy umysł cesarza. O komputerach, umyśle i prawach fizyki*, przeł. P. Amsterdamski, wyd. 2, Wydawnictwo Naukowe PWN, Warszawa 1996, s. 127 i n.

[25] J. R. Searle, *Minds, Brains, and Programs*, „The Behavioral and Brain Sciences", 1980, vol. 3, s. 417-424. Polskie wydanie pt. *Umysły, mózgi i programy* w przekładzie B. Chwedeńczuka por. w: *Filozofia umysłu*..., s. 301-324.

[26] Historię oraz szczegółową i wielostronną analizę wymienionego argumentu Searle'a (w licznych jego wariantach) por. w: J. Kloch, *Świadomość komputerów? Argument „Chińskiego Pokoju" w krytyce mocnej sztucznej inteligencji według Johna Searle'a*, Ośrodek Badań Interdyscyplinarnych PAT, Wydawnictwo BIBLOS, Tarnów 1996.

[27] Ta moja jednozdaniowa rekonstrukcja eksperymentu myślowego J. R. Searle'a jest oczywiście znacznym jego uproszczeniem. Próbuję w niej zachować wierność tylko podstawowej intencji Searle'a. Muszę jednak nadmienić, że w wersji oryginalnej (czy raczej wielu oryginalnych wersjach) był on obudowany szeregiem dodatkowych założeń i detali — założeń i detali często jednak zmienianych również przez samego autora. Por. na ten temat J. Kloch, *Świadomość*..., s. 19-29.

Powstaje oczywiście pytanie — i jest to pierwsze pytanie, jakie stawia Searle — czy na podstawie spreparowanego w ten sposób „tekstu chińskiego", można zasadnie utrzymywać, że ów lokator „chińskiego pokoju" istotnie rozumie język chiński, a przynajmniej rozumie historyjki, które w tym języku „opowiada"?[28] Jak wiadomo, odpowiedź Searle'a jest zdecydowanie negatywna. Nie, mieszkaniec „chińskiego pokoju", choć poprawnie wykonał wszystkie instrukcje i skonstruował tekst, który może wprawić w podziw niejednego sinologa (jest bowiem całkowicie nieodróżnialny od wypowiedzi rodowitych Chińczyków), dalej nie zna języka chińskiego i nie rozumie historyjek, które w tym języku „opowiada".[29]

W związku z przedstawionym eksperymentem myślowym, J. R. Searle stawia jeszcze jedno ważne pytanie. Jeśli lokatora „chińskiego pokoju" i zestaw przewidzianych instrukcją operacji uznać za dobry odpowiednik komputera i wykonywanego przez komputer programu, to powstaje pytanie, czy wykonywanie programu komputerowego, prócz tego, że jest we właściwym sensie rozumieniem, zarazem jest też dostatecznym **wyjaśnieniem rzeczywistej zdolności rozumienia**, np. rzeczywistego rozumienia przez ludzi opowieści, pytań, odpowiedzi, słów? Ponownie, także w tym przypadku, pada odpowiedź negatywna.[30] Wszak po to, powiada J. R. Searle, by ludzie rzeczywiście **rozumieli** opowiadane sobie historyjki (pytania i odpowiedzi), wykonywanie jakiegoś programu komputerowego nie jest ani niezbędne, ani (tym bardziej) wystarczające. Człowiek z „chińskiego pokoju" poprawnie wykonuje pewien program, ale niczego nie rozumie. Z kolei druga sytuacja — sytuacja, w której ktoś znakomicie rozumie jakiś język i opowiadane w nim historie (np. historie opowiadane we własnym języku ojczystym) — nie ma zgoła **nic wspólnego** z wykonywaniem komputerowego programu, czyli obliczeniami i, generalnie, operowaniem elementami, których cała istotna charakterystyka wyczerpuje się w charakterystyce czysto formalnej (syntaktycznej).[31]

5. Powróćmy jednak do pierwszego pytania, przedłożonej przez Searle'a odpowiedzi i jej uzasadnienia. Ta pierwsza kwestia, jej rozstrzygnięcie i jego uzasadnienie wydają się bowiem absolutnie kluczowe dla zrozumienia największej bodaj trudności (być może nawet można by ją nazwać trudnością krytyczną), w jaką wikła się myślenie konsekwentnie komputacjonistyczne. Jak wcześniej wspominałem i jak w zasadzie powszechnie wiadomo, wszyscy zwolennicy silnej wersji AI są przekonani, że komputer cyfrowy, podobnie jak i bohater eksperymentu myślowego Searle'a, wykonując pewien program (np. program Schanka lub program umożliwiający przekład języka angielskiego na język chiński), doskonale rozumie to, co robi: zachowuje się inteligentnie (np. dokonuje trafnych wyborów i eliminuje nietrafne), myśli (np. liczy, czegoś szuka lub coś porządkuje), posiada stany umysłowe (np. pyta, odpowiada,

---

[28] J. R. Searle, *Umysły...*, s. 302. Por. też tamże, s. 29-30.

[29] Tamże, s. 304.

[30] Tamże, s. 304-305.

[31] Tamże, s. 305 i n.

waha się, strofuje użytkowników) itd. Dla komputacjonistów wszystko to znaczy, że komputer, niczym człowiek lub zwierzę, posiada umysł. Searle tymczasem stanowczo utrzymuje, że komputer, wykonując nawet najbardziej finezyjne programy, niczego z nich nie rozumie, wcale nie myśli, nie posiada żadnych stanów umysłowych, a wobec tego nie posiada też umysłu. Wszystkie operacje wykonywane przez komputer są więc, jak się wyraża Searle, całkowicie »bezmyślne«. „**Rozumienie komputerów**, pisze, nie jest tylko (tak jak moje rozumienie języka niemieckiego) częściowe czy niezupełne; **jest zerowe**".[32] Dlaczego?

W uzasadnieniu można wyjść od idei komputera — i to każdego cyfrowego komputera, bez względu na stopień technologicznego zaawansowania. Jak od czasów Turinga doskonale wiemy, podstawową cechą tych urządzeń jest to, że wszystkie wykonywane przez nie operacje dają się scharakteryzować czysto formalnie, czyli za pomocą abstrakcyjnych symboli (np. odpowiednich sekwencji zer i jedynek), a więc dokładnie tak, jak ich działanie opisywał niegdyś Allan Turing. Znaczy to, że dla sprawnego działania tych maszyn zupełnie bez znaczenia pozostaje fakt, jakie treści są lub będą wiązane z symbolami, którymi manipulują. Dla kierunku, sposobu i rezultatu ich działania istotne natomiast są, jak powiedzą logicy, czysto **syntaktyczne** związki między tymi symbolami, a więc np. ich kształt, ilość, sposób uporządkowania, kolejność i własności tym podobne (patrz *Wstęp II*).

Jeśli tak sprawy się mają, to o operacjach wykonywanych przez komputer w trakcie realizacji jakiegoś programu nie wolno nam utrzymywać, że są to operacje umysłowe i że, w szczególności, są to czynności rozumienia albo myślenia. Każdy stan lub proces mentalny — bezwyjątkowo i niejako na mocy definicji — zawsze bowiem ma jakąś **treść**. Wskutek tego zawsze jest ku czemuś skierowany, do czegoś się odnosi, czegoś dotyczy. Od przeszło stulecia mówi się w tym kontekście, że każda czynność umysłowa — każda myśl i każda czynność świadoma — z istoty swej zawsze ma charakter **intencjonalny**.[33] Mówiąc negatywnie, można też powiedzieć, że nie istnieje żadna taka myśl, która nie byłaby myślą o czymś. Podobnie jak nie istnieją również spostrzeżenia, przypomnienia, wyobrażenia, pragnienia, oczekiwania czy niepokoje, które nie byłyby spostrzeganiem **czegoś**, przypominaniem **czegoś** *etc.*[34] Mówiąc jeszcze inaczej: w każdym umyśle, w każdym stanie umysłowym, w każdej umysłowej czynności i każdym umysłowym procesie zawiera się coś więcej niż zbiór albo ciąg samych tylko symboli — zawierają się także pewne treści, czyli **znaczenia** tych symboli. Tymczasem w przypadku komputerów i programów komputerowych — bez względu na to, który z poziomów strukturalnych takiego programu weźmiemy

---

[32] Tamże, s. 307. Podkreślenie moje — J. D.

[33] F. Brentano, *Psychologia z empirycznego punktu widzenia*, przeł. W. Galewicz, Wydawnictwo Naukowe PWN, Warszawa 1999, s. 126. W sprawie różnych koncepcji intencjonalności (i trudności, jakie implikują różne jej pojęcia) por. J. Dębowski, *Świadomość...*, s. 22-24 i 45-56.

[34] Szersze omówienie teorii intencjonalności Searle'a (oryginalnej, choć dość kontrowersyjnej) por. w: J. Dębowski, *Bezpośredniość poznania. Spory — dyskusje — wyniki*, Wydawnictwo UMCS, Lublin 2000, s. 174-186.

pod uwagę — sprawy wyglądają całkiem inaczej, a nawet, w pewnym istotnym sensie, wprost przeciwnie. Albowiem nie dość powiedzieć, że do sprawnego wykonania programu komputerowego **wystarczają** działania na beztreściowych symbolach. Nadto jeszcze trzeba podkreślić, iż **beztreściowość** owych symboli i ich ciągów stanowi **konieczny** warunek wykonywania tych programów.

Główny morał eksperymentu myślowego J. Searle'a, jakim jest „chiński pokój", można tedy wyrazić również w następujący sposób. Syntaktyka i związki czysto syntaktyczne, konieczne i wystarczające dla sprawnego działania komputerów cyfrowych, w aktywności umysłowej (człowieka, zwierząt czy innych istot rzeczywiście obdarzonych umysłem), nie są ani konieczne, ani wystarczające. Dlatego, według Searle'a, analogia komputerowa (analogia mózgu i komputera oraz umysłu i programu komputerowego) jest — wbrew temu, co twierdzą rzecznicy teorii sztucznej inteligencji w jej wersji silnej (J. McCarty, H. Simon, A. Newell, M. Minsky i pozostali) — **całkowicie** chybiona i **zupełnie** bezpodstawna.[35] Zauważmy, iż J. Searle odmawia tej analogii **jakiegokolwiek** uprawnienia. Nie jest skłonny pójść tutaj na **żadne** koncesje, a więc uznać ją np. z pewnymi zastrzeżeniami lub przy pewnych ograniczeniach. Stanowczo obstaje, że działanie komputerów, w przeciwieństwie do stanów i operacji umysłowych, jest całkowicie bezmyślne i że ich zdolność rozumienia jest zerowa.

6. W stanowisku Searle'a niczego istotnego nie zmieniły również długie i burzliwe dyskusje, jakie ów eksperyment myślowy wywołał. Jak się zdaje, dyskusje te spowodowały tylko tyle, że dzięki ciągłemu udoskonalaniu jego struktury logicznej (dzięki jej wzbogacaniu i ujaśnianiu) kolejne warianty stawały się coraz bardziej precyzyjne i bardziej przejrzyste, a poza tym nabierały charakteru bardziej uniwersalnego. Natomiast sedno sprawy nieprzerwanie pozostawało to samo. Żaden program komputerowy nigdy nie spowoduje, by komputer zaczął myśleć, ponieważ o działaniu programu przesądza (i to bez reszty: na każdym jego poziomie i każdym etapie) struktura czysto formalna i czysto syntaktyczne związki między symbolami. Natomiast każdy stan umysłowy lub każda umysłowa czynność, w odróżnieniu od operacji wykonywanych przez komputer lub lokatorów „chińskiego pokoju", zawsze zawiera pierwiastek semantyczny: jest nasycona pewną treścią i ma charakter intencjonalny. Powołując się na swoich bohaterów z „chińskiego pokoju", Searle konsekwentnie utrzymywał też, że sama syntaktyka nigdy nie wytworzy semantyki, bo do jej wytworzenia nie wystarcza.

Przy tym, jak podkreślał Searle, opisanej w ten sposób sytuacji w niczym nie odmienią ani technologiczny postęp, ani **systemy konekcyjne**. Zaludnienie „chińskiego pokoju" wieloma mieszkańcami i równoległe wykonywanie przez nich wielu operacji naraz, jeśli wszystkie pozostałe reguły pozostają bez zmian, niczego istotnie nowego nie wnosi i, w porównaniu z poprzednią sytuacją, w sposób istotny niczego nie zmienia. Natury wykonywanych przez komputer operacji nie odmienią więc także procesory działające równolegle, tj. procesory umożliwiające współbieżne działanie wielu

[35] J. R. Searle, *Umysł...*, s. 28 i n.

różnych programów i równoległe wykonywanie wielu różnych operacji na wielu różnych poziomach.[36] Podobnie jak nie odmienią tej natury inne czynniki tego typu, np. pojemność pamięci, wielkość programu i stopień jego złożoności, szybkość wykonywanych operacji, a nawet interakcje z otoczeniem za pośrednictwem robotów.

Krótko mówiąc, komputer pozostanie komputerem (a więc jedynie i tylko bezmyślną maszyną obliczeniową) tak długo, jak długo wykonywane przez niego operacje będą całkowicie zdeterminowane syntaktyką. Świadomość, myślenie i wszystkie pozostałe zjawiska umysłowe wymagają bowiem czegoś więcej niż sama syntaktyka. Dlatego, według Searle'a, nie ma żadnych podstaw do tego, by działanie programu w komputerze uznawać za najmizerniejszy choćby substytut działania umysłu w mózgu. Program komputerowy może jedynie (i co najwyżej) to działanie **symulować** (pozorować, udawać, modelować), podobnie jak może też symulować działanie czegokolwiek innego, lecz w żadnym razie nie może stanowić **duplikatu** umysłu. Nawet najskromniejszej czynności umysłowej nie może odtworzyć, skopiować czy powtórzyć, ponieważ wykonywanie tego typu czynności jest w ogóle poza jego zasięgiem. To tylko przy założeniach psychologii behawioralnej można sobie roić czy obiecywać, że skoro pewien system **zachowuje się** tak jakby coś rozumiał (myślał), to **jest** on systemem, który **posiada** zdolność rozumienia (myślenia) i który rzeczywiście coś rozumie (myśli). A tego typu roszczenia były w swoim czasie chlebem powszednim nie tylko pośród samych twórców teorii sztucznej inteligencji, lecz także występowały i szeroko promieniowały poza tym środowiskiem. Swymi wpływami objęły m.in. nauki kognitywne, psychologię i, generalnie, filozofię umysłu (gdziekolwiek by nie powstała).

Pierwszym bodaj, który już od początku lat sześćdziesiątych ub. wieku szeroko i energicznie upowszechniał ten punkt widzenia, był behawiorysta Hilary Putnam. To on pierwszy z entuzjazmem podchwycił analogię między ludzkim umysłem a maszyną cyfrową, a następnie, poszukując dla niej solidniejszego uzasadnienia teoretycznego, wytworzył pogląd zwany do dzisiaj albo **funkcjonalizmem**, albo wprost **komputacyjną** koncepcją umysłu.[37] Według funkcjonalizmu i komputacjonizmu, nie zachodzi żadna istotna różnica między pracą ludzkiego mózgu (zbudowanego z substancji organicznej) a pracą urządzeń elektronicznych, w szczególności komputerów cyfrowych (budowanych z tranzystorów, obwodów scalonych i substancji krzemowej). Albowiem w obu przypadkach — niezależnie od materiału, z którego zbudowana jest warstwa *hardwarowa* — zasada działania pozostaje ciągle i dokładnie ta sama. W obu przypadkach mamy do czynienia z „maszyną Turinga", a więc z procesami oblicze-

[36] Z J. R. Searle'm polemizowali w tej sprawie m.in. Jerry Fodor i Ernest LePore. Por. J. Fodor, E. LePore, *Czym jest zasada koneksji?*, „Przegląd Filozoficzny — Nowa Seria", R. V, 1996, nr 3, s. 119-128.

[37] H. Putnam, *Minds and Machines*, [w:] *Dimensions of Mind*, red. S. Hook, New York 1960, s. 148-179. Por. też H. Putnam, *Brains and Behavior*, [w:] *Analytical Philosophy*, red. R. Butler, Oxford 1965, s. 1-20 oraz H. Putnam, *The Mental Life of Some Machines*, [w:] *Intentionality, Minds and Perception*, red. H. Castañeda, Detroit 1967, s. 177-200.

niowymi i wykonywaniem algorytmów. Zatem, z punktu widzenia zarówno Putnamowskiego funkcjonalizmu, jak i teorii sztucznej inteligencji w jej silnej wersji, pewien konkretny stan ludzkiego umysłu jest identyczny z pewnym konkretnym stanem „maszyny Turinga", a więc stanem komputera cyfrowego.

Z kolei pierwszą krytyczną odpowiedzią na tego typu pogląd, zarazem odpowiedzią utrzymaną całkiem w stylu Putnama (jak wiadomo, uwielbiającego metaforykę komputerowo-scjentystyczną), był eksperyment myślowy Neda Blocka, zwany „wielkim mózgiem Chin".[38] Prawdopodobnie stanowił on dla Searle'a jedno z ważniejszych źródeł inspiracji podczas obmyślania eksperymentu z „chińskim pokojem", a przynajmniej tego wariantu „chińskiego pokoju", który nosi nazwę „chińskiej sali gimnastycznej".[39] W obu eksperymentach chodzi o precyzyjne odtworzenie pracy komputera cyfrowego. W obu eksperymentach pojawiają się Chińczycy, którzy ów komputer zastępują, wiernie naśladując działanie komputerowego programu. Operują więc pozbawionymi znaczeń symbolami (sygnałami). Przede wszystkim, jednak w obu eksperymentach pojawia się jedna i ta sama puenta: ich aktywność nie wystarcza do tego, by pojawiło się rozumienie, zakiełkowała myśl, rozbłysła świadomość lub wytworzony został jakikolwiek stan umysłowy.

Oczywiście, i dla Searle'a i dla Blocka znaczy to, że — wbrew funkcjonalizmowi Putnama i silnej wersji teorii sztucznej inteligencji — procesów komputacyjnych (niezależnie od stopnia ich złożoności) nie można utożsamiać z procesami umysłowymi (choćby najprostszymi). O pierwszych bez reszty decydują bowiem elementy czysto formalne, podczas gdy istotę drugich stanowi intencjonalność i semantyka. A zatem, nawet gdy komputer zachowuje się tak jakby zdolność myślenia posiadał, to — wbrew behawiorystom — wcale to nie znaczy, że on rzeczywiście tę zdolność posiada. O komputerze wykonującym mniej lub bardziej skomplikowane programy można przyjąć co najwyżej tyle, iż on tę zdolność myślenia symuluje. Natomiast myśleć w sensie właściwym i dosłownym, jak utrzymuje Searle, może tylko istota organiczna wyposażona w mózg.[40] Z tym, że — jak zaraz dodaje — nawet czynności mó-

[38] Analizę eksperymentu Blocka i jego związku z „chińską salą gimnastyczną" Searle'a por. w: J. Kloch, *Świadomość...*, s. 41-44.

[39] J. R. Searle, *Is the Brain's Mind a Computer Program?*, „Scientific American", January 1990, s. 26-31. Można powiedzieć, iż ostatnia wersja „chińskiego pokoju" w postaci „chińskiej sali gimnastycznej" była odpowiedzią Searle'a na proroctwa Churchlandów, według których systemy konekcjonistyczne spowodują istotny przełom w teorii i praktyce sztucznej inteligencji, m.in. przyspieszą proces uczenia się komputerów i sprawią, że ich „myślenie" stanie się równie twórcze i równie plastyczne jak ludzkie. Por. P. M. Churchland, *A Neurocomputational Perspective. The Nature of Mind and the Structure of Science*, The MIT Press, Cambridge (Mass.) 1989, zwł. s. 129-135 (*Rozdział 7*). Searle z kolei jest przekonany, że programy współbieżne i systemy konekcjonistyczne czy neuropodobne również nie są w stanie doprowadzić do wytworzenia semantyki, a wobec tego nie osiągają nic więcej ponad to, co jest osiągalne za pomocą programów sekwencyjnych (szeregowych). I taki właśnie sens miało zmodyfikowanie „chińskiego pokoju" w „chińską salę gimnastyczną".

[40] J. R. Searle, *Umysł...*, s. 35.

zgu, gdyby ograniczyć je do realizacji programu komputerowego, nie doprowadzą do wytworzenia umysłu.[41] Ponieważ jednak człowiek rzeczywiście myśli i posiada umysł, przeto jego mózg — wbrew temu, co twierdzi M. Minsky („mózg to komputer zbudowany z mięsa") — bez wątpienia jest czymś więcej niż cyfrowy komputer czy nawet system takich komputerów. Jeśli zatem komuś się marzy stworzenie artefaktu, który rzeczywiście miałby jakieś stany umysłowe (a nie tylko je symulował), to — bez względu na to, czym by to było pod każdym innym względem — musi on posiadać siłę przyczynowego oddziaływania, która jest porównywalna z siłą przyczynowego oddziaływania ludzkiego mózgu (jego możliwościami).[42]

7. Na podstawie przeprowadzonych analiz i dyskusji oraz poczynionych ustaleń J. R. Searle zdecydowanie odrzuca silny program sztucznej inteligencji. Co najwyżej gotów jest uznać tzw. słabą wersję teorii sztucznej inteligencji i towarzyszący jej program badawczy. Zgodnie ze słabą wersją AI, komputer i realizowane przez niego programy są w stanie jedynie symulować pracę mózgu. Natomiast w żadnym razie i w najskromniejszym nawet sensie nie są repliką (duplikatem) mózgu i wytwarzanych przez mózg procesów umysłowych.[43] Należy przy tym wyraźnie podkreślić, iż kiedy Searle mówi o symulacji procesów umysłowych przez programy komputerowe, to ma na myśli wyłącznie symulację czysto **funkcjonalną.** Zdecydowanie odrzuca zaś myśl (z powodów dokładnie tych samych, które przesądzają o odrzuceniu silnego programu AI), by w grę mogła tu wchodzić symulacja **strukturalna,** by zatem programy komputerowe mogły udanie symulować sam **mechanizm** przetwarzania informacji, nie zaś powodować tylko, że dane wyjściowe komputera będą identyczne lub zbliżone (podobne) do rezultatów osiąganych przez człowieka.[44] Wszak pod względem strukturalnym nie istnieje zgoła żadne podobieństwo pomiędzy maszyną i człowiekiem — zarówno w warstwie *hardware*'u, jak i w warstwie *software*'u. Fakt ten musiałby uznać nawet Hilary Putnam — musiałby go uznać, gdyby nie to, że względy doktrynalne (behawioryzm i pragmatyzm) zabraniają mu wypowiadania się o jakichkolwiek mechanizmach czy strukturach.

Jedna z największym pułapek intelektualnych (i największych mistyfikacji), do jakich prowadzi myślenie w stylu Putnama, a w ślad za nim komputacjonistów i kognitywistów, polega m.in. na tym, iż próbuje się wmawiać, że sama symulacja lub

---

[41] Tamże, s. 36.

[42] Tamże, s. 36-37.

[43] Tamże, s. 42-51.

[44] Stosunek czynności umysłowych do działania komputera dość często próbuje się odwzorować w terminach relacji: już to relacji homomorficznej, już to relacji izomorficznej. Przy takim ujęciu homomorfizm odpowiadałby symulowaniu (J. R. Searle), izomorfizm zaś byłby odpowiednikiem duplikowania (Putnam i silna AI). Powstaje jednak pytanie, czy różnicy między homomorfizmem a izomorfizmem zarazem odpowiadałaby też wskazana wyżej różnica między modelowaniem czysto funkcjonalnym a modelowaniem strukturalnym? Mam wątpliwości. Mam je także dlatego, że — jak powszechnie się uważa — izomorficzne odwzorowanie jakiegokolwiek układu biologicznego zasadniczo jest zadaniem niewykonalnym.

imitacja czegoś może uchodzić za autentyk, a więc albo wprost za to, co pozoruje lub imituje, albo przynajmniej za duplikat. Próbuje się wmawiać, że np. myślenie zachodzi tam, gdzie faktycznie go nie ma, że maszyna obliczeniowa rzeczywiście (w dosłownym tych słów znaczeniu) „przetwarza informacje", że komputer równie literalnie jak człowiek (gdy np. myśli lub działa) przestrzega reguł i równie skrupulatnie jak człowiek (a może i bardziej) się do nich stosuje. Zasługa Searle'a polega na tym, że pomysłowo i bezprecedensowo skutecznie wykazał bezpodstawność wymienionych uroszczeń — że je zdemaskował i nieomal sfalsyfikował.

8. Rzecz jasna, Searle'a krytyka silnej AI ani nie doprowadziła do zaniechania dotychczasowych kierunków poszukiwań, ani też nie spowodowała, że wszyscy dotychczasowi promotorzy silnej AI w istotny sposób zrewidowali swoje przekonania.[45] Krytyka Searle'a nie jest też krytyką, która wyczerpuje listę zastrzeżeń, jakie wzbudzała i wzbudza teoria sztucznej inteligencji. W dziesięć lat później Roger Penrose, należycie doceniając ciężar argumentacji Searle'a, w krytyce programów badawczych sztucznej inteligencji idzie jeszcze krok dalej.[46] W odróżnieniu od Searle'a, uważa bowiem, iż procesy umysłowe (intelektualne, świadomościowe) nie mogą być przez komputer nawet symulowane (nie mówiąc już o duplikowaniu). W tym kontekście wskazuje na poznanie matematyczne, które, jak dowodzi, nie poddaje się algorytmizacji i jest zasadniczo nieredukowalne do procesów komputacyjnych. Według Penrose'a, właściwość ta jednoznacznie wynika m.in. z twierdzeń Gödla, Churcha, a nawet Turinga. Świadczy o tym także, i byłby to argument pozytywny, istnienie matematyki nierekurencyjnej i matematycznego wglądu — wglądu, który umożliwia odkrywanie i penetrowanie świata przedmiotów matematycznych bez potrzeby uruchomienia mechanizmów komputacyjnych. Dzięki matematycznym wglądom, Penrose mówi w tym kontekście o tzw. „logicznej zasadzie refleksji", odsłania się przed nami sens podstawowych pojęć matematycznych oraz treść tych twierdzeń, wobec których jakiekolwiek procedury dowodowe są całkowicie nieskuteczne, a które mimo to są prawdziwe. Wymienionych faktów i okoliczności nie jest dziś w stanie ignorować żaden twórczy matematyk.[47]

W stosunku do argumentacji Searle'a, argumentacja Penrose'a przeciwko rozmaitym wariantom programu badawczego sztucznej inteligencji jest nie tylko bardziej

---

[45] Zapewne można to tłumaczyć rozmaicie, ale jest faktem, że w latach osiemdziesiątych ze stanowiska radykalnego funkcjonalizmu począł wycofywać się Hilary Putnam. Obecnie sądzi on, że ze względu na nieporównywalną z czymkolwiek plastyczność ludzkiego umysłu (w tym również rozumu matematycznego) funkcjonalizm jest nie do utrzymania. Znaczy to m.in. tyle, że właśnie ze względu na tę plastyczność „poziom intencjonalny" jest zasadniczo nieredukowalny ani do „poziomu fizycznego", ani do „poziomu obliczeniowego". Por. H. Putnam, *Representation and Reality*, The MIT Press, Cambridge (Mass.) 1988, s. XII-XV i 339-340. Jednak jako funkcjonalista pełną kapitulację Putnam ogłosił dopiero w latach dziewięćdziesiątych (por. artykuł z 1992 roku: *Why Functionalism Didn't Work?*). Natomiast przy swoim, a więc przy programie silnej AI, dalej obstają m.in. D. Hofstadter, M. Minsky, J. Fodor i Z. Pylyshyn.

[46] R. Penrose, *Nowy...*, s. 34-36.

[47] Tamże, s. 120-172 i 445-470.

kompetentna i wskutek tego bogatsza merytorycznie, lecz także, jak się zdaje, do krytyki myślenia komputacyjnego wnosi całkiem nową jakość. Penrose bowiem, w odróżnieniu od Searle'a, nie ogranicza się do zakwestionowania możliwości posiadania przez maszyny obliczeniowe stanów umysłowych porównywalnych z ludzkimi (teza tzw. funkcjonalizmu maszynowego, wielce irytująca Searle'a z powodu pomieszania duplikacji z symulacją), ale odmawia maszynom nawet zdolności symulowania zjawisk umysłowych. Świadomość — powiada Penrose — wymaga elementów nieobliczalnych, a jej istotę stanowi „widzenie", nie zaś komputacja.[48] I dlatego nie jest w stanie jej wytworzyć, a nawet mniej lub bardziej udolnie naśladować, żaden algorytm i żaden, niechby i najbardziej finezyjny, program komputerowy.[49]

9. Sformułowana przed chwilą myśl Penrose'a całkiem niedawno znalazła swoje dodatkowe potwierdzenie — potwierdzenie dość spektakularne. Oto bowiem Gregory J. Chaitin — amerykański matematyk, od lat sześćdziesiątych pracujący naukowo w IBM Thomas J. Watson Research Center w Yorktown Heights (N.Y.) — dokonał odkrycia, które, przynajmniej wedle jego własnej opinii, stanowi istotne rozwinięcie znanych konkluzji Gödla i Turinga.[50] Punktem wyjścia był dla Chaitina problem znany szeroko jako tzw. „dziesiąty problem Hilberta" lub tzw. *Entscheidungsproblem*. Problem ten krótko można wyrazić w sposób następujący: czy istnieje metoda, która umożliwia jednoznaczne rozstrzygnięcie, że dane równanie diofantyczne (*scil.* równanie algebraiczne o całkowitych współczynnikach) ma rozwiązanie w dziedzinie liczb całkowitych?[51] Jak wiadomo, problem ten żywo zajmował m.in. Alana Turinga, przez którego został przeformułowany w tzw. „problem zakończenia pracy" lub tzw. „problem stopu". W bezpośrednim związku z nim, poza twierdzeniami Gödla, pozostają także twierdzenia A. Churcha, E. Posta i innych (meta-) matematyków, którym udało się ostatecznie wykazać, że oba problemy — *Entscheidengsproblem* i tzw. „problem stopu" — są sobie równoważne i że dla obu w rachubę wchodzi jedynie rozstrzygnięcie negatywne.[52]

---

[48] Tamże, s. 456-458 i dalsze.

[49] Według Penrose'a, nie należy również oczekiwać, by jakiś istotny przełom mogły spowodować całkiem nowe technologie — technologie dziś jeszcze nieznane, a oparte np. na fizyce kwantowej. Albowiem taki „kwantowy komputer" i tak nie mógłby wykonywać operacji niealgorytmicznych, a wobec tego nie mógłby uprawnienie uchodzić za model mózgu. Tamże, s. 440-443.

[50] Por. G. J. Chatin, *Randomness and Complexity in Pure Mathematics*, „International Journal of Bifurcation and Chaos", 1994, vol. 4, s. 3-15, w tym *Abstrakt* (s. 3). Chciałem nadmienić, iż na osobę Chaitina oraz na jego badania i publikacje po raz pierwszy zwrócił mi uwagę mój student z Uniwersytetu Marii Curie-Skłodowskiej w Lublinie, student III roku filozofii, pan Piotr Czarnota. Chciałem mu w tym miejscu za to podziękować.

[51] Tamże, s. 4 i n. Por. też G. J. Chaitin, *Randomness in Arithmetic*, „Scientific American", 1988, No. 1 (259), s. 80-81 i n.

[52] W tym zakresie Gregory J. Chaitin specjalnie ceni sobie osiągnięcia J. P. Jonesa z Uniwersytetu w Calgary i I. W. Matijasewicza z Instytutu Matematycznego im. W. A. Stiekłowa w Leningradzie. G. J. Chaitin, *Randomness in Arithmetic*, s. 82.

Jak powiedziałem, pewnym szczególnym wariantem „dziesiątego problemu Hilberta" zajął się również G. J. Chaitin. Mianowicie, używając jako narzędzia pracy oprogramowania komputerowego (LISP), które specjalnie zostało napisane do celów matematycznych i które chodzi na IBM-ie RS/6000, skonstruował — jak pisze — „przewrotne (niezwykle skomplikowane) 200-stronicowe równanie algebraiczne z parametrem N i 17 tysiącami niewiadomych".[53] Następnie postawił pytanie: „Czy dla każdej całkowitej wartości liczbowej parametru N istnieje skończona czy też nieskończona ilość całkowitych liczbowych rozwiązań?"[54]

Odpowiedź wypadła zdumiewająco. Jeśli bowiem do równania podstawiać kolejne wartości liczbowe parametru N oraz w przypadku skończonej liczby rozwiązań przyjmować 0, w przypadku zaś nieskończonej 1, to jego rozwiązaniem będzie ciąg zer i jedynek, którego w żaden sposób nie można odróżnić od ciągu zestawiającego wyniki nieskończenie wielu rzutów monetą. Nieobliczalną liczbę rzeczywistą z przedziału między 0 i 1, odpowiadającą ciągowi otrzymanych zer i jedynek, G. J. Chaitin nazwał następnie Ω (*Omega*).[55]

$$\Omega = 00101110110010011000 1...$$

Jak się okazuje, jej kolejne cyfry odpowiadają nieskończonej liczbie zupełnie przypadkowych faktów arytmetycznych. Wiemy wprawdzie, że każda część *Omegi* **musi być** albo zerem, albo jedynką, ale nie wiemy i **nigdy** wiedzieć nie będziemy (!), kiedy rzeczywiście wystąpi w niej zero, a kiedy jedynka. Sytuacja jest więc w maksymalnym stopniu matematycznie nieprzewidywalna. Innymi słowy, *Omega* (Ω), ponieważ stanowi skrajnie nieuporządkowaną sekwencję zer i jedynek, jest nieredukowalna do żadnego algorytmu — jest, jak się powiada, **algorytmicznie nieuprasz-czalna** (niekompresowalna).[56] Znaczy to, że, mówiąc odrobinę inaczej, ewentualny algorytm (program komputerowy), za którego pośrednictwem moglibyśmy tę sekwencję wiernie odtworzyć, musiałby być równie długi, jak ona sama.

Wnioski, które wyprowadza Chaitin w rezultacie odkrycia *Omegi*, są następujące. Cytuję:

> Zazwyczaj przyjmuje się, że jeśli coś jest prawdą, to jest nią z jakiegoś powodu. W matematyce powodem tego, że coś jest prawdziwe, jest dowód, a wobec tego zadaniem matematyków jest odnajdywanie tych dowodów. ...Odkryłem tymczasem, przykładem jest Ω, że pewne **istotne fakty matematyczne zachodzą bez powodu!** One są prawdziwe przez przypadek! W konsekwencji zatem zawsze będą poza zasięgiem matematycznego rozumienia (*mathematical reaso-*

---

[53] G. J. Chaitin, *Randomness and Complexity*..., s. 3. Zobacz też G. J. Chaitin, *Randomness in Arithmetic*, s. 83.

[54] Zauważmy, iż Chaitin wcale tu nie pyta, czy skonstruowane przez niego równanie jest w ogóle rozwiązalne. Poniekąd byłoby to bowiem tylko powtórzenie pytania, z którym już wcześniej zmierzył się Turing.

[55] G. J. Chaitin, *Randomness in Arithmetic*, s. 81.

[56] Tamże, s. 83-85. Zobacz też G. J. Chaitin, *The Limits of Mathematics*, Springer, Singapore 1998, s. 54.

> *ning*). ... Są całkowicie nieredukowalne. Nie ma w nich żadnej struktury. Nie stosują się do nich żadne wzory. 0 lub 1 stają się [odpowiednimi] częściami Ω bez żadnego szczególnego powodu, całkiem przypadkowo. Nawet gdyby Bóg chciał tu coś stanowić na tak lub nie, to każda część Ω wymagałaby osobnego [specjalnego] rozstrzygnięcia, ponieważ nie ma w niej żadnych korelacji, nie ma też redundancji! ... W tym kierunku **nieredukowalność matematycznej informacji** (*Irreducible Mathematical Information*) nie może już pójść dalej, nieprawdaż?[57]

Cóż można jeszcze w tej sprawie powiedzieć? Istnienie nieobliczalnej liczby Ω — *nota bene* liczby, której odkrycie nie byłoby chyba możliwe bez użycia profesjonalnego komputera (ten fakt G. J. Chaitin zawsze podkreśla ze szczególnym naciskiem) — wskazuje nie tylko na doniosłość matematyki nierekurencyjnej, lecz także zwraca uwagę na wszechobecną **przypadkowość:** przypadkowość (losowość) przez środowisko samych matematyków ostentacyjnie lekceważoną czy nawet ignorowaną. Tymczasem, jak to wykazał Chaitin, przypadkowość (losowość) nie omija również matematyki. Zawiera się także w czystej matematyce, a nawet w elementarnej arytmetyce liczb naturalnych (dziedzina równań diofantycznych). Jak pisze:

> Bóg gra w kości nie tylko w mechanice kwantowej i fizyce klasycznej, ale nawet w czystej matematyce, nawet w elementarnej teorii liczb.[58]

Być może mamy więc kolejne twierdzenie limitacyjne, a przynajmniej wyraźny jego przedsmak. To bez wątpienia ważny rezultat. Jak sądzę, nie mniej ważne, choć dla wielu może nieco osobliwe, są także metodologiczne postulaty, które formułuje Gregory J. Chaitin. Mianowicie, dostrzegając nieskuteczność (bezowocność, jałowość) prowadzenia pracy badawczej „w dawnym dobrym stylu"[59], proponuje nowy paradygmat matematyczny — paradygmat opierający się na zwrocie w kierunku „matematyki eksperymentalnej" (*quasi*-empirycznej), nade wszystko zaś uznający przypadkowość (losowość) za istotną i niezbywalną cechę także świata przedmiotów matematycznych.[60] Według Chaitina, ważnym źródłem nowych impulsów jest dzisiaj również informatyka, ponieważ stale rosnące możliwości obliczeniowe komputerów stwarzają całkiem nowe warunki eksperymentowania i testowania. Jakkolwiek paradoksalnie by to nie wyglądało, okoliczności tej nie wolno dzisiaj ignorować lub choćby nie doceniać.[61] Jednak, co chciałem wyraźnie odnotować, wbrew komputa-

---

[57] G. J. Chaitin, *The Limits...*, s. 54-55 (*Conclusion*). Przekład własny ad hoc; podkreślenie moje — J. D.

[58] G. J. Chaitin, *Randomness and Complexity...*, s. 12.

[59] Chaitin zauważa w tym kontekście, iż poszukiwania matematyków, którzy pracują jeszcze w starym stylu, a więc ignorują twierdzenia Gödla i jego własne ustalenia (nie uwzględniają przypadkowości w świecie matematycznym), przypominają próbę wydedukowania z praw Newtona np. całej teorii względności albo równań Maxwella czy Schrödingera. Tamże, s. 13 i n.

[60] Tamże, s. 12-15 (*sub. 5, Experimental mathematics*).

[61] Ewentualny paradoks polega tu na tym, że to, co ze swej istoty niealgorytmizowalne, nieobliczalne i niesekwencyjne — np. okazana przez Chaitina przypadkowość świata matematycznego, a nadto dynamika nieliniowa, kwantowa teoria pola, geometria fraktalna itp. — usiłuje się wytropić

cjonistom i mimo swych wieloletnich związków z IBM-em, Chaitin nie uważa, by komputery mogły zastąpić w myśleniu samych matematyków.[62] Podobnie jak J. R. Searle (i zgodnie z programem słabej AI) sądzi tylko, że maszyny obliczeniowe stanowią dzisiaj dla matematyków wyjątkowo skuteczne **narzędzie** pracy badawczej. Tedy wielkim błędem z ich strony byłoby tego faktu należycie nie zdyskontować.[63] Wszelako zdaniem Chaitina, nowe czasopismo matematyczne — czasopismo, którego jeszcze nie ma, a które dobrze odpowiadałoby programowi nowej matematycznej szkoły — winno się ukazywać pod nazwą *Journal of Experimental Mathematics*.[64]

10. Być może, powątpiewać można w rzeczywistą doniosłość (przełomowość, rewolucyjność) faktycznie dokonanych przez Chaitina ustaleń oraz w lansowany przez niego (w stylu charakterystycznym dla amerykańskiego rynku idei) obraz przyszłej matematyki i program nowej szkoły matematycznej. Jednak jedno wydaje się tu być niewątpliwe. Osiągnięte przez Chaitina wyniki, a przynajmniej częściowo także ich filozoficzna interpretacja, znakomicie współbrzmią z niektórymi poglądami Rogera Penrose'a. Oczywiście, choćby z powodu rozległości swoich zainteresowań naukowych, Penrose idzie dalej zarówno od Chaitina, jak i od Searle'a. Nie wchodząc w zbyt wiele szczegółowych rozstrzygnięć, chciałem zwrócić uwagę jeszcze tylko na jeden problem — problem specjalnie zaakcentowany przez Penrose'a w *Epilogu* książki *Nowy umysł cesarza*... Chciałem zwrócić uwagę na coś, co w szerokiej recepcji wymienionej książki Penrose'a nie jest, jak myślę, należycie doceniane, a już chyba najczęściej jest po prostu pomijane. Krótko mówiąc, chodzi mi o tak zwany **„problem *qualiów*"**.

Do uświadomienia sobie istoty tego problemu można dzisiaj dojść w różny sposób i na różnych drogach.[65] Na przykład można wyjść od twierdzenia Turinga i stanowiska wszystkich pozostałych zwolenników silnej AI, że dwie dowolne maszyny obliczeniowe (uniwersalne maszyny Turinga), których *hardware* osiągnął dostateczny poziom technologicznego zaawansowania i dostateczny stopień złożoności, są sobie

---

i opisać właśnie za pomocą algorytmów i obliczeń.

[62] Na odnalezienie dowodu i genialnych autorów tego przedsięwzięcia czeka wszak szereg nowych twierdzeń matematycznych — twierdzeń równie interesujących jak wielkie twierdzenie Fermata czy hipoteza Riemanna. G. J. Chaitin, *The Limits*..., s. 55.

[63] G. J. Chaitin, *Randomness and Compelxity*..., s. 14-15.

[64] Tamże, s. 15.

[65] By uświadomić sobie istotę tzw. „problemu *qualiów*", dobrze jest wyjść dzisiaj od prostego pytania Thomasa Nagla „Jak to jest być nietoperzem?". Th. Nagel, *Pytania ostateczne*, tłum. A. Romaniuk, Fundacja Aletheia, Warszawa 1997, s. 203-220: Jak *to jest być nietoperzem?* Jednak można też wyjść od jeszcze innych pytań, np.: Czy genialny neurolog, który posiada niemal kompletną wiedzę na temat budowy i funkcjonowania centralnego układy nerwowego (w tym ludzkiego mózgu), lecz niestety jest od urodzenia niewidomy, będzie w stanie kiedykolwiek zrozumieć, że koperkowy kolor jego krawata zupełnie nie pasuje do wiśniowego koloru jego marynarki? Czy ów genialny, lecz niewidomy neurolog, w ogóle kiedykolwiek dowie się, co to znaczy, że pewien przedmiot jest zielony, czerwony, pomarańczowy, szafirowy *etc.*?

równoważne.[66] W następnym kroku wystarczy zapytać, czy ten sam znak równoważności wolno nam postawić pomiędzy umysłami dwojga ludzi (niechby i bliźniąt jednojajowych albo ludzkich klonów), nie mówiąc już o sytuacji, w której ów znak równoważności mielibyśmy postawić pomiędzy umysłem pewnego konkretnego człowieka a umysłem np. pewnego orangutana lub innego ssaka? Wszak do uznania wymienionej równoważności dodatkowo winno nas zachęcać także i to, że z punktu widzenia anatomii i fizjologii mózgu mamy tu bodaj pełną identyczność. Tę identyczność (lub przynajmniej istotne podobieństwo) jeszcze łatwiej można odnaleźć na coraz to głębszych poziomach strukturalnych materii (ożywionej i nieożywionej) — aż po jej struktury atomowe.

Jakby odwracając kierunek myślenia zawarty w pierwotnej argumentacji Searle'a, można też pytać dalej. Można mianowicie pytać, czy ta niemal pełna (lub porównywalna) identyczność podłoża, w którym implementowane są poszczególne programy (ewolucyjnie, kulturowo...), niechybnie zaowocuje niemal pełną (lub porównywalną) **funkcjonalną** identycznością wymienionych „urządzeń"? Czy zatem doprowadzi ona do sytuacji, w której te same dane „na wejściu" zawsze i nieodmiennie prowadzić będą do otrzymywania tych samych danych „na wyjściu". Nie trzeba wielkiej przenikliwości, by zauważyć, że — przynajmniej w obrębie myślenia ludzkiego i ludzkiej aktywności umysłowej — tak wcale nie jest i tak wcale być nie musi. Wobec tego, wypada postawić jeszcze jedno (ostatnie) pytanie: dlaczego? Dlaczego, mimo identycznych zasad działania i identycznych programów (*software*), a nadto jeszcze mimo identyczności (lub istotnej zbieżności) w budowie i działaniu warstwy *hardware'owej*, przynajmniej czasami mamy do czynienia z drastyczną rozbieżnością rezultatów otrzymywanych „na wyjściu"?

Nie chodzi mi przy tym o rozbieżność jedynie, by się tak wyrazić, „międzygatunkową" (choć to również interesujący problem dla siebie). Chodzi mi tu przede wszystkim o rozbieżność „międzyosobniczą". Jak się zdaje, rozbieżności tej nie sposób wyjaśnić inaczej niż przez odwołanie się do **subiektywnej** treści ludzkich doznań zmysłowych, **subiektywnej** treści także innych przeżyć świadomych, wreszcie **subiektywnych** i zawsze **zindywidualizowanych** sposobów doświadczania świata transcendentnego i samego siebie. Z tego właśnie powodu, by na przykład zrozumieć „jak to jest być nietoperzem?", trzeba po prostu **być** nietoperzem. Innej drogi nie ma. Co ciekawe, w ostatnich dziesięcioleciach fakt ten musiał uznać nawet Hilary Putnam.[67]

---

[66] Oczywiście, abstrahuje się tu od czasu i prędkości działania tych maszyn, a nawet rozmiarów pamięci, dopuszczając możliwość korzystania przez obie maszyny z zewnętrznych jednostek pamięci. Według komputacjonistów, są to charakterystyki zupełnie nieistotne dla wymienionej równoważności.

[67] H. Putnam, *Wiele twarzy realizmu i inne eseje*, przeł. A. Grobler, Wydawnictwo Naukowe PWN, Warszawa 1998, s. 304-311. Idzie o to, że bez uwzględnienia *qualiów* (subiektywnych treści umysłowych i ich numerycznego zindywidualizowania) nie można by wykluczyć, iż jesteśmy np. „mózgami w naczyniach", których główna właściwość polega na zdolności produkowania słów lub zdań, stosownie do odpowiednich reguł (programów).

Naturalnie, filozof umysłu, nawet ten wyedukowany na behawioryzmie, nie może na tym poprzestać i powinien pytać dalej. Pytając dalej, prędzej czy później stanie zaś przed problemem jaźni — problemem osobniczej tożsamości i jej podstaw ontycznych. Prędzej czy później zapyta tedy o czynnik **integrujący** całą tę zmienną i wielce zróżnicowaną aktywność umysłową człowieka (lub innej istoty świadomie żyjącej) — aktywność skrajnie zindywidualizowaną i zawsze napiętnowaną subiektywnie.[68]

Czy jednak na pytanie to kiedykolwiek będzie w stanie odpowiedzieć, jeśli zarazem z góry założy — tak jak zakłada się to w neobehawioryzmie, programach badawczych sztucznej inteligencji, kognitywistyce i różnej proweniencji redukcjonizmach — że możliwy i dopuszczalny jest wyłącznie **trzecioosobowy punkt widzenia**? Wątpię. Zatem, cóż jeszcze może uczynić nasz badacz umysłu? Być może postara się o to, by samo pytanie **unieważnić**! To przecież najprościej i skądinąd najwygodniej. Dzięki zaś rozmaitym dwudziestowiecznym „modernizmom", dalibóg, już prawie nie widać w tym nic zdrożnego. Zapewne więc poczuje się wyraźnie ośmielony. Jak myślę, prawdziwa bieda czeka nas jednak wtedy, gdy do unieważnienia tego pytania poczuje się też całkowicie **uprawniony**. Wtedy wszak, i to *lege artis*, unicestwione zostanie nie tylko *ego*, lecz także i *alter ego*. Tylko... Co dalej?

[68] Jak zauważa Penrose, z punktu widzenia silnej AI, „świadomość osobowa" może być uznana za część oprogramowania, natomiast „jej konkretna realizacja w ludzkiej postaci za efekt wykonania programu (*software'u*) przez mózg i ciało (*hardware*)". R. Penrose, *Nowy*..., s. 41-43.