Tytuł: **Kilka słów w kwestii Chińskiego Pokoju i dwu innych argumentów przeciw SI: *sir* Rogera Penrose'a i z redukcjonizmu**

Autor: Marek Kasperski / mjkasperski@kognitywistyka.net

Recenzent: prof. dr hab. Włodzisław Duch (UMK)

Źródło: http://www.kognitywistyka.net / mjkasperski@kognitywistyka.net

Data publikacji: 07 I 2002

SŁOWEM WSTĘPU: Eksperyment Chińskiego Pokoju Johna Searle'a jest tak już znany, że przywoływać go tutaj szczegółowo bliżej nie było by sensu. Jednakże zamieszanie, jakie on wywołał (i dalej wywołuje) jest tak zgubne w skutkach dla niektórych myślicieli (zaćmiewa im trzeźwość umysłu), że wydaje mi się, że należałoby się z nim spróbować rozprawić. Osobiście trudu tego zadania podjąłem się w innej zgoła pracy *Wczoraj, dzisiaj, jutro Sztucznej Inteligencji*[1].

Zacznę od przypomnienia – w syntetycznym skrócie – tez argumentu Searle'a, a potem przejdę do analizy poszczególnych kontrargumentów, które relacjonuje J. Kloch w tekście *Chiński Pokój. Eksperyment myślowy Johna Searle'a. Studium historyczno-filozoficzne (cz. 2)*. Po co, skoro już sam wspomniany autor tego dokonał? Pretekstem ku temu może być fakt, iż odnoszę osobiste wrażenie (być może złudne, co tekst poniższy mam nadzieję rozsądzi), iż relacja J. Klocha nie pozbawiona jest pewnego rodzaju osobistej sugestywności, składającej się w ostateczności na niezbyt rzetelne studium pro-Searle'owskiej filozofii. Do tego wszystkiego dochodzi osobiste poczucie specjalnego rodzaju mistyfikacji wprowadzonej do tekstu przez przywoływanego tu Klocha. Zresztą tekst poniższy pokaże, jakiego rodzaju.

1. Chiński Pokój

1.1. Argument

> Aksjomat 1: Mózgi są przyczynami umysłów.
> Aksjomat 2: Syntaktyka nie wystarcza dla semantyki.
> Aksjomat 3: Program komputerowy całkowicie określa jego formalna lub syntaktyczna struktura.
> Aksjomat 4: Umysły zawierają treści psychiczne, mówiąc dokładniej, treści semantyczne.

[1] Praca ta, po poprawkach i okrojeniu, wyszła drukiem w formie książki: M. J. Kasperski, *Sztuczna Inteligencja. Droga do myślących maszyn*, Helion, Gliwice 2003.

> Wniosek 1: Nie ma takiego programu komputerowego, który sam w sobie wyposażyłby system w umysł. Mówiąc krótko, programy nie są umysłami ani same w sobie nie wystarczą dla powstania umysłu.
> Wniosek 2: Czynności mózgu ograniczone tylko do realizowania programu komputerowego nie wystarczą, by funkcjonowanie mózgu doprowadziło do powstania umysłu.
> Wniosek 3: Cokolwiek, co mogłoby być przyczyną umysłu, musiałoby mieć moc oddziaływania przyczynowego porównywalną z możliwościami mózgu.
> Wniosek 4: Wyposażenie jakiegoś zbudowanego przez nas artefaktu w program komputerowy nie wystarcza, by miał on stany umysłowe porównywalne z ludzkimi. Artefakt taki powinien oczywiście mieć zdolność przyczynowego oddziaływania porównywalną z możliwością ludzkiego mózgu.
>
> [J. R. Searle, *Umysł, mózg i nauka*, s. 35-37. Org. *Minds, Brains, Programs and Persons*, w: "The Behavioral and Brain Sciences" (dalej w pracy będę używał skrótu "BBS") 5(2)/1982), ss. 339-341].

Tak brzmi Chiński Pokój, ni mniej, ni więcej i wnioski z niego wypisane przez samego Searle'a.

1.2. Jasny Pokój. Kontrargument [systemowy]

Czytelnicy strony też doskonale wiedzą, jak skonstruowany jest argument jasnego Pokoju małżeństwa P. M. i P. S. Churchland. Przypomnę tu tylko w skrócie:

> Aksjomat 1: Elektryczność i magnetyzm są formami energii.
> Aksjomat 2: Zasadniczą cechą światła jest jasność.
> Aksjomat 3: Energia nie jest ani konieczna, ani wystarczająca do uzyskania jasności.
> Wniosek 1: Elektryczność i magnetyzm nie są ani konieczne, ani wystarczające do wyjaśnienia istoty światła.
>
> [P. M., P. S. Churchland, *Czy maszyna może myśleć?*, s. 18].

Tak brzmi Jasny Pokój, ni mniej, ni więcej i wniosek z niego wypisany przez samych autorów. Oczywiście należy tu kwestią wyjaśnienia dodać, iż argument ten jest całkowicie sprzeczny a teorią elektromagnetyzmu stworzoną przez Maxwella – teorią, której prawdziwości nie ma co kwestionować!

Jak podsumowuje dyskusję nad tym argumentem Kloch? Następująco:

> Searle odrzuca argument Jasnego Pokoju – by można było w ogóle wnioskować przez analogię należy najpierw sprawdzić, czy w danym przypadku w ogóle zachodzi analogia. Związek między światłem a elektromagnetyzmem wynika z praw fizyki i ma charakter przyczynowy; natomiast symbole jako symbole nie mają ani charakteru przyczynowego ani mocy przyczynowej – mogą jedynie spowodować następny krok programu komputerowego. Są również dobrze znane właściwości systemu zerojedynkowego; nie ma co czekać na wyniki badań odnośnie do ujawnienia fizycznych czy przyczynowych cech zer i jedynek. *Ergo* – analogia Jasnego Pokoju jest chybiona i nie jest kontrprzykładem dla Chińskiego Pokoju. Można ją najwyżej odnotować jako jeszcze jedną nieudaną próbę nawiązywania do idei Searle'a a jednocześnie tworzenia błędnych analogii czy pomysłów, jakich było wiele w odniesieniu do argumentu. (...)

> Churchlandowie w *Jasnym Pokoju* bez sprawdzenia poprawności toku myślenia przyjęli analogię między elektromagnetyzmem a światłem oraz manipulowaniem symbolami chińskimi a semantyką.
>
> [J. Kloch, *Chiński Pokój*..., Internet].

Tylko po przeczytaniu powyższego fragmentu nasuwają się następujące pytania i wątpliwości:

1. Kloch stwierdza, że jakoby Churchlandowie „nie wykazali analogii pomiędzy ich argumentem a argumentem J. Searle'a". A nawet, co czytamy w ostatnim fragmencie, że posunęli się tak daleko, iż nie wykazali, że przyjęli analogię między elektromagnetyzmem a światłem oraz manipulowaniem symbolami chińskimi a semantyką. Tu też pojawia się, moim zdaniem, specjalne zaciemnianie faktów. Kloch nie pisze wprost, że „manipulowanie chińskimi symbolami" jest przecie syntaktyką, czyli 1) dokładnie tym co miał na myśli Searle decydując się na aksjomat 2, 2) dokładnie tym samym, dla analogii, co elektromagnetyzm. Dlaczego Kloch nie raczył tego zauważyć? Pytanie pozostawiam otwartym.
2. Kloch pisze, że „symbole jako symbole nie mają mocy przyczynowej, tylko wywołują kolejny krok w działaniu programu". Zatem jak Kloch rozumie moc przyczynową, bo chyba nie tak, jak klasycznie rozumuje się dzięki wykładni chociażby systemu Hume'a (czytelnikom niniejszego artykułu, jeszcze nie zapoznanym z tą koncepcją, proponuję jako lekturę: D. Hume, *Badania dotyczące rozumu ludzkiego*). Być może przyczynowość, która tutaj się pojawia ma jakieś zabarwienie sprawcze – ociera się o panpsychizm, bądź dualizm Descartesa? Trudno stwierdzić jednoznacznie, tym bardziej jak weźmie się pod uwagę kolejną koncepcję Searle'a, dotyczącą mocy sprawczych neuronów (*sic*!) (Por. J. R. Searle, *Umysł na nowo odkryty*). Przyznać każdy musi, że to oczywisty wyraz bądź dualizm, bądź panpsychizmu koncepcji Searle'a, lecz Kloch zwraca uwagę, że interpretacja taka jest nieprawdziwa!

   > Wydaje się, że część autorów replik w ogóle nie zrozumiała istoty Chińskiego Pokoju; Searle zalicza do nich: Johna Marshalla, E. W. Menzela Jr., Martina Ringle oraz Jerry Sameta; ich artykuły zostały pominięte w niniejszej pracy. Douglas R. Hofstadter jest z kolei przykładem komentatora, który rzuca na Searle'a inwektywy i przypisuje mu dokładnie odwrotne poglądy niż jego własne. Metoda autora słynnego bestselleru *Gödel, Escher, Bach. An Eternal Golden Braid* polega ogólnie na mówieniu *p* tam, gdzie Searle pisze „nie *p*"; przykładem jest odniesienie do dualizmu, który pomysłodawca Chińskiego Pokoju odrzuca zaś według Hofstadtera ma być jego zwolennikiem.
   >
   > [J. Kloch, *Chiński Pokój*...].

   Pytanie oczywiście: to, jaka to koncepcja? I pozostaje tu tylko ewentualnie trzecia odpowiedź – *witalistyczna*. Ale i taki rodzaj interpretacji zdaje się być odrzucany.
3. Kloch pisze, że „nie ma co czekać na wyniki badań odnośnie ujawnienia przyczynowych cech zer i jedynek". W takim razie jak może przyjmować bezzasadnie tezę J. Searle'a o przyczynowej mocy neuronów? Jeżeli zaś ją przyjmuje, to na jakiej podstawie, jeśli nie przyszłych wyników badań, przyjmuje że jedna organizacja materii posiada moce przyczynowa, a inna nie? Przypomnę tylko, bo i Kloch nie porusza tego w swym artykule, że Searle wprowadził taką tezę zapytany o to, że jak to się dzieje, że skoro syntaktyka nie wywołuje semantyki, a mózgi posiadają elementy

semantyczne, to gdzie są te elementy, co je wywołuje. Odpowiedzią Searle'a było: „treści semantyczne powstają dzięki mocom przyczynowym neuronów".

W ostateczności też uważam, parafrazując Klocha, że „*Ergo* – analogia Jasnego Pokoju jest NIE chybiona i JEST kontrprzykładem dla Chińskiego Pokoju. Można ją odnotować jako jeszcze jedną UDANĄ próbę nawiązywania do idei Searle'a a jednocześnie tworzenia NIEbłędnych analogii czy pomysłów, jakich było wiele w odniesieniu do argumentu".

1.3. Problem sensorium. Kontrargument

> Człowiek zamknięty w Chińskim Pokoju jest odizolowany od świata zewnętrznego – trzeba więc mu umożliwić kontakt z otoczeniem, by mógł rozumieć i mieć inne stany umysłowe. Taka ogólna idea, z pewnymi modyfikacjami, przyświeca kolejnym komentatorom. Można do niej zaliczyć takich autorów jak Bruce Bridgeman[2], Daniel Dennett[3], Jerry A. Fodor[4] oraz Aaron Sloman z Moniką Croucher[5].
>
> [J. Kloch, *Chiński Pokój*...].

Tak przedstawia się pokrótce ten argument, który to sam Kloch nazwał „kwestią semantyki z robota-i-komputera". O co w nim chodzi? Autorzy ci chcą pozbyć się z teorii Sztucznej Inteligencji modelu czarnej skrzynki (którą świetnie scharakteryzowali N. Block, w: *Encyklopedia filozofii*, red. T. Honderich i S. Lem, w *Summa technologiae*, Interart, ss. 124-129), który pokutuje do dziś. Powiadają oni, jeżeli nasze stany umysłowe współtworzą w znacznej części informacje ze świat, to system musi posiadać zdolność otwartości na świat poprzez zaimplementowanie mu sensorium – koncepcja bardzo nie głupia. Lecz Kloch znowu wydaje się ją zepchać na boczne tory, żeby nie powiedzieć, zinfantylizować:

> Kamera umożliwiałaby kontakt ze światem zewnętrznym. Komputer sterowałby czynnościami robota np. chodzeniem, wbijaniem gwoździ czy spożywaniem pokarmów *(sic!)*. Taka konfiguracja sprzętowo-programowa nie byłaby już jedynie prostym urządzeniem Schanka – komputer wewnątrz robota rozumiałby a nawet posiadał inne stany umysłowe.
>
> [J. Kloch, *Chiński Pokój*...].

Przy czym naprawdę nie wiem skąd u autora bierze się owe „sic!". Wystarczy pojąć pożywanie się jako umiejętność do przedłużania swojego funkcjonowania, by pobieranie prądu przez maszyny nazwać pożywaniem się! Robotyka kognitywna i Sztuczna Inteligencja zna już takie projekty, m.in.:

- ZÓŁW, Greya Waltera;
- po części DARWIN III, G. Edelmana;
- ryboidalna ROBOPIKE, projekt z MIT;
- homoidalny COG, Rodneya Brooksa;

[2] Chodzi o: B. Bridgeman, *Brains + Programs = Minds*, w: "BBS", Nr 3/1980, ss. 427-428.
[3] Chodzi o: D. Dennett, *The Milk of Human Intentionality*, w: "BBS", Nr 3/1980, s. 430.
[4] Chodzi o: J.A. Fodor, *Searle on What only Brains Can Do*, w: "BBS", Nr 3/1980, s. 431.
[5] Chodzi o: A. Sloman, M. Croucher, *How to Turn an Information Processor into an Understander*, w: "BBS", Nr 3/1980, ss. 447-448.

- projekty komercyjne, przeznaczone na rynek zabawek: PooChi (z kością, której „ssanie” zapewnia mu odpowiednią ilość energii), MeowChi (kocia kuzynka PooChi, z myszą zamiast kości);
- a nawet, chciałoby się powiedzieć – do pewnego stopnia niektóre odmiany Tamagochi (np. te wykorzystywane w *Artificial Life*).

Searle’owi i Klochowi też wydają się nie podobać te koncepcje. Całą tą skądinąd ciekawą ideę, Kloch podsumuje lapidarnie:

> (...) jeśli robot w ludzkim ciele zacząłby się tym sposobem uczyć chińskiego, to w rzeczywistości nadawałby formalnym symbolom semantyczne znaczenie (czy zgodne z rzeczywistością?) i nadal nie miałoby miejsca rozumienie języka chińskiego.
>
> [J. Kloch, *Chiński Pokój*...].

Przy czym, niechybnie odwołując się do alchemicznej idei ludzika w mózgu (homunkulusa), który się gdzieś tam uczy i zarządza nami, Kloch nie podaje skąd taka koncepcja? Po co jakieś pudełka w pudełku – mnogość niepotrzebnych bytów (kłania się brzytwa Ockhama). Z artykułu można oczywiście wywnioskować (lecz aż boję się to napisać): Kloch nie traktuje jako całość (robota) systemu myślącego (centrali) i systemu odczuwającego (sensorium). Oczywiście aż samo nasuwa się pytanie, na jakiej podstawie Kloch traktuje jako jedność człowieka (system pewnie trochę bardziej skomplikowany niż tylko wyżej wymienione funkcje)?

1.4. [Kontr]Argument symulatora mózgu

Ten rodzaj argumentacji można zamknąć w jednej tezie – żeby maszyna mogła myśleć prócz funkcjonalnego podejścia należy podjąć próbę podejścia strukturalnego do mózgu (tzn. badać i naśladować jego strukturę). Jest to teza, która streszcza koneksjonizm. Pogląd ten zdają się wyrażać: Douglas Hofstadter (patrz niżej), John Haugeland (patrz niżej), William Lycan[6], Steven Savitt (patrz niżej) oraz Donald Walter[7]. Zgodnie powiadają oni, że żeby usprawnić i znaturalizować system przekształcania/dopasowywania/interpretowania informacji należy zbudować strukturę mózgopodobną, która posiadałaby m.in. możliwość pracy równoległej czy modułowej, a nawet brałby pod uwagę różnicę w mocach informacyjnych tak skrzętnie przywoływaną w dyskusji nad różnicą działania komputera („wszystko-albo-nic”), a mózgu (biochemia mózgu wydaje się odzwierciedlać zasadę, którą nazwę umownie „a-może-tak-a-może-nie-a-kto-to-wie”). Co na to Kloch?

> Patrząc z zewnątrz może się to wydawać niemożliwe; ale tak jak zdawałoby się, że termostat sam z siebie nie może utrzymać temperatury w odpowiednich granicach a jednak to robi, tak i rozumie symulator mózgu. Obserwator nie jest w stanie tego stwierdzić – jest zbyt mały w stosunku do wielkiego układu symulatora mózgu, by dostrzec, że on rozumie.
>
> [J. Kloch, *Chiński Pokój*...].

Dlaczego zaś stwierdzenie Klocha wydaje się co najmniej brakiem argumentu? Z prostej przyczyny, jaką jest istnienie problemu „drugiego ja” – „Skąd ja (jak człowiek) wiem, że ktoś

[6] W. G. Lycan, *The Functionalist Reply*, w: "BBS", Nr 3/1980, ss. 434-435.
[7] D. O. Walter, *The Termostat and the Philosophy Professor*, w: "BBS", Nr 3/1980, s. 449.

(jakieś drugie ja) w ogóle myśli?" Problem o tyle warty przywołania i na miejscu, że jedyną sensowną próbą jego rozwiązania jest ta, która stoi w opozycji do fragmentu z artykułu *Chiński Pokój*. *Eksperyment myślowy Johna Searle'a*.

Pomijam przy tym dalsze rozróżnienia argumentu z symulatora mózgu, do których należą:

- Model rurkowy, Douglasa Hofstadtera[8],
- Demon, Johna Haugelanda[9],

pozwalając sobie jednakże na zatrzymanie się przy ostatnim należącym do tej listy, postawionego przez Stevena Savitta[10], nazwanego w artykule Klocha „myśleniem o mózgu jako *n-1* neuronach".

Argument ów przedstawia się następująco:

> Wyobraźmy sobie trzy przypadki: Chińczyka o *n*-neuronach, o *n-1* neuronach oraz człowieka z Chińskiego Pokoju. Pierwszy rozumie wszystko i posiada stany intencjonalne, ostatni – według Searle'a – ani nie rozumie, ani nie ma intencjonalności. Co można powiedzieć o człowieku bez jednego neuronu, którego pracę symuluje *demon Searle'a*? Czy jego rozumienie spada w miarę eliminowania kolejnych neuronów i co miałoby oznaczać, że ktoś ma zdolność rozumienia w 73 % czy 4,5%? Sam Savitt odrzuca takie rozwiązanie i nie widzi powodu, by odmawiać intencjonalności Chińczykowi o *n-1* neuronach, jeśli ów brakujący neuron jest symulowany przez demona. Człowiek o *n* neuronach i człowiek z eksperymentu myślowego, którego pomysłodawca wariantu uznaje za symulator mózgu, winni mieć taką samą intencjonalność i zdolność rozumienia chińskiego.
>
> [J. Kloch, *Chiński Pokój*...].

Szczerze mówiąc to faktycznie ten argument nie jest najlepszym przeciw pomysłowi Searle'a. Całościowo przypomina on trochę późniejszą debatę Davida Chalmersa[11] nad tym, kiedy przestanie być świadomym człowiek, kiedy poddać go procedurze eliminowania po jednym poszczególnych neuronów tak, że w ostateczności można skonstruować taki ciąg: <osobnik z *n* liczbą neuronów, osobnik z *n-1* liczbą neuronów, osobnik z *n-2* liczbą neuronów, ..., osobnik z *n-(n-1)* liczbą neuronów>. Pytanie Chalmersa w stosunku do takiego ciągu brzmiało: Kiedy osobnik, albo który z osobników będzie pozbawiony świadomości, przy założeniu, że świadomość na pewno posiada osobnik pierwszy. Oraz drugie, nieco innej konstrukcji: Czy będzie tak, że świadomość będzie *stopniowo* wygasać wraz z traceniem neuronów przez osobnika, czy też tak, że okaże się, iż całkowicie *nagle* zgaśnie u pewnego z osobników?

Dlaczego te pytanie, i w rezultacie ów kontrargument nie jest najlepszym? Odpowiedź jest dosyć trywialna – nie wiadomo jak jest, wszystko, zatem co można powiedzieć o takich przypadkach to czyste spekulacje. Samo podmienianie neuronów jedne drugimi – co może sugerować chęć podmieniania neuronu przez symulację demona – w przypadku argumentu S. Savitta, czy podmienianie neuronów konstruktami krzemowymi – jak ma to miejsce w analogicznym eksperymencie D. Chalmersa – też na niewiele pewnie by się zdało. Dlaczego?

[8] D. R. Hofstadter, *Reductionism and Religion*, w: "BBS", Nr 3/1980, ss. 433-434.
[9] J. Haugeland, *Programs, Causal Powers and Intentionality*, w: "BBS", Nr 3/1980, ss. 432-433.
[10] S. Savitt, *Searle's Demon and the Brain Simulator*, w: "BBS", Nr 2(5)/1982, ss. 342-343.
[11] *The Conscious Mind*..., 1996.

Gdyż 1) odmienność strukturalna powoduje odmienność funkcjonalną[12] i 2) odmienność struktur atomowych wpływa na odmienność całości [mózg przed przemianą ≠ mózg po przemianie, ze względu na jego funkcjonalność][13].

Zatem, w ostateczności, próbując być obiektywnym w kwestii możności myślenia/nie myślenia maszyn i biorąc pod uwagę tutaj argument Searle'a, przyznaję, że kontrargument Savitta nie wydaje się być dobrym. Przy czym stwierdzam, biorąc pod uwagę wszystko, co powyżej o argumencie tym i kontrargumentach w stosunku do niego napisałem, jak i w stosunku do tego, co jeszcze niżej napiszę.

1.5. [Kontr]Argument z kombinatoryki

Liczne środowiska uniwersyteckie, w tym Berkeley University, Stanford University, a także osobistości – Wiliam Lycan (patrz wyżej), Thomas Edelson[14] – głoszą tezy, że odpowiednia kombinacja powyższych kontrargumentów zbija całkowicie argument Chińskiego Pokoju. Lecz na tak przygotowany kontrargument Searle wystrzeliwuje również swoją salwę kombinatoryczną. Powtarzać się nie ma sensu, gdyż składa się on dokładnie z: obrony Searle'a przeciw 1.3, plus obrony przeciw 1.4. W rezultacie też, jak lapidarnie konkluduje Kloch:

> Tak więc i połączenie odpowiedzi systemowej z zastosowaniem robota i symulatora mózgu nie jest w stanie zanegować eksperymentu myślowego Johna Searle'a.
>
> [J. Kloch, *Chiński Pokój*...].

Oczywiście, bo wedle tego, co obaj sądzą (tzn. Searle i Kloch), a co już wyżej zostało przedstawione, to jakże może i mieć?

1.6. Kontrargument z przyszłych rozwiązań. Czyli poczekajmy, zobaczymy

W Chińskim Pokoju można zauważyć pewnego rodzaju wąską użyteczność (czy raczej nieużyteczność, o czym w rozdz. końcowym) – argument dopóki wydaje się być sprawnym, dopóty dotyczy on klasycznej algorytmicznej metody wprowadzonej przez koncepcję maszyny Turinga, a rozwijanej chociażby przez takich klasyków Sztucznej Inteligencji jak Marvin Minsky, Claude Shannon czy John McCarthy. Zatem obiecującym kontrargumentem mogą okazać się przyszłe rozwiązania tak techniczne, jak i teoretyczne. Tak w skrócie brzmi kontrargument, które reprezentują: Ned Block[15], Aaron Sloman z Moniką Croucher (patrz wyżej), Daniel Dennett (patrz wyżej), Bruce Bridgeman (patrz wyżej), William Lycan (patrz wyżej) oraz Roger Schank[16] oraz niektórzy pracownicy z Berkeley University.

Searle popada też w pewnego rodzaju pułapkę formułując swoje aksjomaty i wnioski. I tak we wniosku 3 pisze:

---

[12] Zob. W. Duch, *Jaka teoria umysłu w pełni nas zadowoli?*
[13] Zob. S. Lem, *Brain Chips* oraz *Brain Chips III*.
[14] T. Edelson, *Stimulating Understanding*: *Making the Example Fit the Question*, w: "BBS", Nr 2(5)/1982, ss. 338-339.
[15] N. Block, *What Intuitions about Homunculi Don't Show*, w: "BBS", Nr 3/1980, ss. 425-426.
[16] R. C. Schank, *Understanding Searle*, w: "BBS", Nr 3/1980, ss. 446-447.

> Cokolwiek, co mogłoby być przyczyną umysłu, musiałoby mieć moc oddziaływania przyczynowego porównywalną z możliwościami mózgu.
>
> [J. R. Searle, *Umysł, mózg i nauka*, 36].

Jeżeli tak, wystarczy zatem skonstruować taką maszynę, która mocom przyczynowym dorównywałaby mózgowi wytwarzającemu umysł. Fakt ten spostrzegają, i na swój sposób przeciw Searle'owi wykorzystują, P. M. i P. S. Churchlandowie. Piszą oni:

> Czy nauka może stworzyć sztuczną inteligencję wykorzystując to, co wiadomo o układzie nerwowym? Wydaje nam się, że nie ma żadnych pryncypialnych powodów, żeby to miało być niemożliwe. Searle wygląda na rozgniewanego, ale on sam złagodził swoje stwierdzenia mówiąc, że „każdy inny system, zdolny do wytworzenia rozumu, musi mieć zdolność sprawczą równą (przynajmniej) zdolności mózgu". (...) Zakładamy, że Searle nie próbuje twierdzić, że udany sztuczny rozum musi mieć *wszystkie* zdolności sprawcze mózgu, na przykład zdolność do zapadania na infekcje wirusowe, zdolność do alergii i tak dalej. Żądanie pełnej równoważności jest podobne do żądania, by każde sztuczne urządzenie latające znosiło jajka.
>
> [P. M., P. S. Churchland, *Czy maszyny mogą myśleć?*, s. 23].

Jednak nawet tak silne i nie tylko intuicyjnie prawdziwe argumenty na rzecz możności skonstruowania maszyny myślącej, wydają się być nie przekonujące, a nawet nie zbijające argumentu Searle'a. Kloch pisze:

> Mrzonką jest więc wiązanie nadziei z przyszłymi rozwiązaniami sprzętowo-programowymi. W odniesieniu do programów formalnych oraz komputerów cyfrowych i analogowych eksperyment myślowy rozwiewa nadzieje pokładane w przyszłych rozwiązaniach; szybsze komputery o większej pamięci, wyposażone w najbardziej nawet skomplikowane programy podlegają i podlegać będą nadal argumentowi Searle'a. Można bowiem założyć, że człowiek w Chińskim Pokoju pracuje według już poprawnego, rozwiniętego programu, który nie przypomina w niczym pomysłu Schanka; ten system można powtarzać *n*-razy w dowolnie wielkim układzie – i nadal będzie miał zastosowanie eksperyment myślowy Johna Searle'a.
>
> [J. Kloch, *Chiński Pokój*...].

Skąd wiara w wystarczalność argumentu Searle'a i niewystarczalność rozwoju techniki w dyskusji nad możliwością i niemożliwością skonstruowania SI u J. Klocha? Trudno stwierdzić, bo jak zobaczyć możemy po powyższych rozważaniach ani dzięki dedukcji, ani dzięki indukcji. A przyszłość może nas jeszcze zaskoczyć, jak nie raz zaskakiwała!

Rozważaniem nad tym argumentem kończy J. Kloch swoją drugą część tekstu *Chiński Pokój. Eksperyment myślowy Johna Searle'a*. Lecz by w miarę w pełni zobaczyć toczące się dyskusje nad tym zagadnieniem proponuję jeszcze na koniec rozpatrzyć kilka uwag i kontrargumentów nie poruszanych w wyżej analizowanej publikacji.

1.7. [Kontr]Argument Jasnego Pokoju z innej strony

Po lekturze tekstu J. Klocha dochodzę do wniosku, że nie w pełni został tam uchwycony i zaakcentowany główny wątek argumentacji Churchlandów. Rzecz dotyczy sporności, co do prawdziwości i zasadności nazywania aksjomatem trzeciej tezy Searle'a.

> Być może ten aksjomat jest prawdziwy, jednak Searle nie może udawać, że o tym wie z całą pewnością. Co więcej, założenie, że aksjomat ten jest prawdziwy, jest równoważne spłyceniu pytania o sens klasycznej AI. Założenie programu klasycznej AI polega na bardzo interesującym przypuszczeniu, że jeśli ustali się zestaw ruchów odpowiednio zaprojektowanego wewnętrznego tańca elementów syntaktycznych, jeśli odpowiednio dołączy się te elementy do wejścia i wyjścia systemu, to uzyska się możliwość osiągnięcia takich samych stanów intelektualnych, jakie znajdowane są w ludzkim mózgu.
>
> Spłycający charakter trzeciego aksjomatu Searle'a staje się jasny, jeśli porównamy go bezpośrednio z pierwszym wnioskiem jego pracy: *Programy nie są ani konieczne, ani wystarczające dla myślenia*. Widać wyraźnie, że jego trzeci aksjomat wnosi bezpośrednio 90 procent wagi tego prawie identycznego wniosku. To dlatego myślowy eksperyment Searle'a jest przeznaczony specjalnie do budowania aksjomatu 3. To jest istota Chińskiego Pokoju.
>
> [P. M., P. S. Churchland, *Czy maszyny mogą myśleć?*, s. 20].

Konsensus sam się nasuwa! Należy udowodnić, że zauważalna „semantyczna ciemność" w Chińskim Pokoju jest ciemnością *de facto*. Jak piszą dalej autorzy Jasnego Pokoju, Searle

> nie ma prawa na podstawie tego spostrzeżenia uporczywie twierdzić, że żadna manipulacja symbolami, wykonywana zgodnie z ustalonymi regułami, nigdy nie zdoła wytworzyć zjawisk semantycznych.
>
> [P. M., P. S. Churchland, *Czy maszyna może myśleć?*, ss. 20-21].

Na koniec roztrząsania tego argumentu warto zaznaczyć, iż, dla przykładu, bardzo wielki matematyk Kurt Gödel wierzył w konstrukcję takiej odpowiednio bogatej syntaktyki, w której byłyby dowiedlne zdania, tj. prawdziwe pod względem semantycznym – żargonowo: takiej, gdzie syntaktyka realizowałaby semantykę. Zresztą nie tylko on, również wielki niemiecki filozof G. W. Leibniz chcąc skonstruować *characteristica universalis*!

Zatem w ostateczności, osobiście traktuję argument Chińskiego Pokoju Searle'a, u którego podstaw leży teza trzecia, o niewystarczalności syntaksy dla semantyki, jako spór sympatyków Searle'a z sympatykami Gödela i Leibniza. Przy czym ja osobiście sympatyzuję z tymi drugimi.

1.8. [Kontr]Argument S. Lema z puzzle

Argumenty Lema przeciw argumentowi Chińskiego Pokoju wydają się być dosyć dobrze znane w literaturze polskiej i przyjmowane przez specjalistów (m.in. przez W. Marciszewskiego; por. jego książkę *Sztuczna inteligencja*). Cały wywód przygotowany przez niego i zaprezentowany w *Tajemnicy chińskiego pokoju*, jak pisze sam autor, ucieka od rzeczowych argumentów „za" i „przeciw" na rzecz eksperymentu myślowego pomyślanego na planie eksperymentu Searle'a. Lem pisze:

> Bierzemy tak zwany „puzzle", który jako składna całość ukazuje jakiś naturalistyczny obraz albo jakąś fotografię. To może być kopia *Rycerza* Rembrandta, albo fotografia wieży Eiffla wszystko jedno, CO tam widać, byle naturalistyczna oczywistość obrazu była DANA. Następnie ten obraz rozcina się na takie małe, pokrętne fintifluszki-kawałki, z jakich zazwyczaj takie igraszki (*puzzle*) się składają, bacząc wyłącznie na to, żeby się **formą** każdy kawałek różnił od każdego na tyle, ażeby ich nie dało się dopasować w składną całość, zamieniając elementy błędnie miejscami. Na koniec

> odwracamy obraz i potrząsamy pudlem, w którym leżą te kawałeczki, aby porządnie je wymieszać. Potem przychodzi nasz eksperymentator, widzi same „lewe strony" tych kartonowych kawałeczków i ma z nich złożyć całość taką, ażeby każdy znalazł się tam, gdzie pasuje. Będzie to nieco pracochłonne, ale możliwe, skoro zamiany miejsc zostały udaremnione kształtem indywidualnie nadanym tym kawałkom. Jeżeli teraz odwrócimy całość obrazem do góry, to oczywiście zobaczymy obraz Rembrandta, Ledę z łabędziem czy też wieżę Eiffla, mimo, że ten, kto składał kawałki, nie miał najbledszego pojęcia o tym, że składa nie bezsensowne kształty w całość obrazu, ale że odtwarza pewien bardzo wyrazisty, jednoznaczny OBRAZ. Przecież to jest całkowicie oczywiste i żadnej dodatkowej informacji w ogóle nie wymaga. Zamiast obrazu może tam się równie dobrze znajdować napis w polskim, w albańskim, chińskim czy jakimkolwiek innym z istniejących na tej planecie 5000 języków: może tam pojawić się napis „Kto rano wstaje, temu Pan Bóg daje", albo „Chłop żywemu nie przepuści", albo „*Ich weiss nicht, was soll es bedeuten, dass ich so traurig bin*" (Heine, *Lorelei)* itd. Wszystko jedno, co się pokaże po odwróceniu lewej strony poskładanych należycie kawałków na stronę właściwą, tj. „prawą", ale przecież ewidentne jest, że ten, kto składał, jak „program komputera", **NIC** nie wiedział, czyli nie miał pojęcia, czyli ani na włos nie rozumiał, CO ukaże się na odwrocie, a nawet nie miał wyobrażenia o tym, ŻE tam się jakakolwiek koherentna całość znacząca pojawi. Czy to jest argument przeciw mniemaniu, iż komputer mógłby jednak rozumieć podobnie jak człowiek, co on robi, wykonując kolejne kroki nadane instrukcją? Moim zdaniem, ma to tyle wspólnego z „obaleniem" tezy o AI, co teza, iż z kremowych ciastek można ułożyć napis, negujący szansę wybuchu Etny. Jedno za grosz nie ma nic wspólnego z drugim.
>
> [S. Lem, *Tajemnica chińskiego pokoju*, Internet. Wytłuszczenia tamże].

Żeby jeszcze bardziej uzmysłowić sobie to, co chciał przekazać Lem, przyjrzyjmy się kolejnemu fragmentowi jego artykułu. Tym razem relacji z życia samego autora.

> Obecnie jako skrótowy, autentyczny i oczywisty przykład zastosuję „*argumentum ad hominem*", przy czym ten człowiek, to ja. Jestem już bardzo przygłuchy, bez aparatów wetkniętych w oboje uszu prawie nic nie rozumiem (skoro nie słyszę) z tego, co się do mnie mówi. Jeżeli w piekarni, w której kupuję pieczywo, obcy człek, rozpoznawszy mnie, bo mnie w telewizji widział, zwróci się do mnie, ja nie mogę przez wzgląd na całokształt sytuacji sięgać do guzików płaszcza, potem do kieszeni po pudełko z protezami słuchowymi, lecz usiłuję z tego, co on mówi do mnie, uchwycić chociaż jedno-dwa słowa. To zwykle się udaje, a jak nie, to on powtórzy albo powie coś głośniej i nakieruje mnie na rozmyty, ale może i generalny sens swojej wypowiedzi. Nieporozumienia w rozmowach z głuchymi często się naturalnie zdarzają, ale nikt wtedy, nawet gdy nieporozumienie jest 100%, nie sądzi, że mówi do manekina albo do robota z takim komputerem w czaszce, który „tylko czysto formalnie działa". Czysto formalnie działał ten, kto składał tablicę w obraz (puzzle), ten, kto składa niepojęte symbole chińskie, ale gdy on omyli się, to Chińczyk uzna raczej, że to tylko omyłka, a nie 100-procentowy brak wszelkiego rozumienia. Zresztą na ogół jest tak, że sens dla rozumienia współwyznacza konsytuacja i nie myślę, żeby pytający w piekarni pytał mnie o liczbę gwiazd w mgławicy Andromedy albo o najlepszy przepis na sporządzenie piernika z migdałami.
>
> [S. Lem, *Tajemnica chińskiego pokoju*].

W tym momencie widoczne już staje się, że faktycznie argument Chińskiego Pokoju nie dotyka sedna testu Turinga, więc w rezultacie nie może być istotnym (rzeczowym!) argumentem przeciwko zwolennikom AI. Wszystko dlatego, że pytania, jakie będą dostarczane maszynie będą tak skonstruowane, iż ze zwyczajnych manipulacji na symbolach odpowiedzi na nie maszyna nie udzieli. Sens pomysłu Turinga tkwi właśnie w drugiej formie – „niezwyczajnej" manipulacji na symbolach, z którymi właśnie przecie trudność jest

związana, o którą w rezultacie, stawiając pytanie *Czy maszyny mogą myśleć?*, pytamy! I słusznie, dalej Lem zauważa, iż nie będziemy maszyny pytać się o rzeczy, o których sama nie wie (i nas się nie pytają, a jeśli pytają, to rozumieją fakt, iż odpowiedzieć nie możemy, bo i nie wiemy), ale też nie będziemy maszyny pytać o rzeczy, na których temat ma już gotowe – przez programistów wpisane – odpowiedzi (Co i tak nie jest możliwe!). Z przyczyn oczywiście jasnych: w stosunku do pierwszych, to i „*Salomon z próżnego nie przeleje*", w stosunku do drugich – chodzi nam o maszyny myślące, a nie o myślących programistów (tych już mamy).

> Ludzie parający się AI na razie dowiedzieli się, że trzeba dla gry w pytania i odpowiedzi stwarzać porządne RAMY (*frames*) sytuacyjne, ale wiadomo też z doświadczeń dnia powszedniego, że można pytać kucharkę o to, jak zasmażkę robi, ale się raczej nie należy spodziewać jej sensownej odpowiedzi na pytanie, dlaczego w tłokowych silnikach spalinowych nowszych modeli aut nie ma dwu zaworów (ssania i wydechu), tylko cztery albo chociażby trzy. Nie powie nic, bo pojęcia nie ma, co z „rozumieniem" albo i „nierozumieniem" gramatyki, idiomatyki, składni języka nie ma nic wspólnego. Żeby rozumieć wypowiedź, należy uchwycić jej sens i jej zakres znaczeniowy i *last but not least* jej desygnatywną orientacją specyficzną.
>
> [S. Lem, *Tajemnica chińskiego pokoju*. Wytłuszczenie tamże].

1.8.1. Czy Chiński Pokój Searle'a jest Searle'a?

Na koniec omawiania argumentu Chińskiego Pokoju i powiązania go z osobą S. Lema, chciałbym jeszcze zwrócić uwagę na jedną rzecz. Wśród wielu odpowiedzi na zarzut Searle'a napływających z ośrodków badawczych i uniwersyteckich znalazła się taka, która rozumiejąc argument Chińskiego Pokoju jako pytanie „*Jeżeli zamknięty w pokoju mężczyzna nic nie rozumie, tylko manipuluje znaczkami, to kto rozumie?*" stwierdzała (D. R. Hofstadter, D. Dennett, *The Minds I*):

> (...) człek sam nic nie rozumie, ale on plus instrukcje plus pokój „rozumie chiński".
>
> [S. Lem, *Tajemnica chińskiego pokoju*].

Dlaczego zwracam uwagę na ten fakt? Gdyż, jeżeli tak rozumieć problem przedstawiony przez Searle'a (przypominam, w 1980 roku), to jest on postawiony wtórnie.

Jak relacjonuje Lem przy innej zgoła okazji, w książce z 1964 roku, *Summa technologiae*:

> Fizyk i autor Science-Fiction w jednej osobie, A. Dnieprow, opisał w nowelce eksperyment, mający obalić tezę o „uduchowieniu" maszyny tłumaczącej z języka na język w ten sposób, że elementami maszyny, zastępującymi tranzystory czy przełączniki, stali się rozstawieni odpowiednio na dużej przestrzeni ludzie. Wykonując proste funkcje przekazu sygnałów, przetłumaczyła ta z ludzi zbudowana „maszyna" zdanie z języka portugalskiego na rosyjski, za czym jej konstruktor pytał każdego z ludzi, którzy byli „elementami maszyny", o treść owego zdania. Nikt jej oczywiście z nich nie znał, bo z języka na język tłumaczył ów system jako pewna dynamiczna całość. konstruktor (w noweli) wyciągnął z tego wniosek, że „maszyna nie myśli".
>
> [S. Lem, *Summa technologiae*, s. 118].

Na co, w odpowiedzi

> jeden z cybernetyków radzieckich zareplikował w piśmie, które umieściło opowiadanie, zauważywszy, że gdyby rozstawić całą ludzkość tak, by każdy człowiek odpowiadał funkcjonalnie jednemu neuronowi mózgu konstruktora w noweli, to układ ów myślałby tylko jako całość i żadna z osób, biorących udział w tej „zabawie w mózg ludzki" nie rozumiałaby, o czym ów „mózg" myśli. Z czego jednak doprawdy nie wynika jakoby sam konstruktor pozbawiony był świadomości.
>
> [S. Lem, *Summa technologiae*, s. 118].

1.9. Chiński Pokój, czy Dwa Chińskie Pokoje? Kontrargument autora

Na początek muszę wytłumaczyć pewną rzecz dotyczącą autorstwa argumentu, który poniżę przedstawię.

4 października 2001 roku miałem okazję i przyjemność przedstawić ten argument w towarzystwie profesorów: J. Perzanowskiego, W. Ducha i prof. Tyburskiego na Uniwersytecie Mikołaja Kopernika w Toruniu. Prof. W. Duch stwierdził, że taki rodzaj argumentacji słyszy po raz pierwszy i że wykazuje on swego rodzaju oryginalność. Będąc świadomym autorytetu prof. Ducha i jednocześnie również nie spotykając w żadnej literaturze tematycznej takiego rodzaju zarzutu pod adresem poprawności Chińskiego Pokoju, przyjmuję, że W. Duch miał rację i że argument ów jest pewnego rodzaju moim pomysłem. Jeżeli zaś by tak nie było i okazało się, że do tego czasu już ktoś takiego rodzaju tezę przedstawił, wówczas też osobie, która o tym wie i może ten fakt zaświadczyć byłbym wdzięczny, a i stosownie zmienię autorstwo poniższego pomysłu.

W książeczce Searle'a, *Umysł, mózg i nauka* można znaleźć tak naprawdę dwa nieco różne Chińskie Pokoje. Pierwszy ten w wersji skróconej, który wypisałem na samym początku pracy i drugi – w wersji opisowej. Dlaczego pomimo przyjętej równoważności obu wersji argumentu Chińskiego Pokoju przyjmuję, że są to dwie wersje? Pytanie bardzo interesujące.

W wersji opisowej Searle umieszcza w pokoju człowieka, który nie zna języka chińskiego i który odpowiada poprawnie na pytanie w języku chińskim. Wersje tych pokoi są dwa: albo jest tam kosz, w którym znajdują się pojedyncze kartki, albo jest to biblioteka, w której na półkach stoją opasłe tomiska. Czy to będą kartki w koszyku, czy też książki na półkach jedno jest wspólne i bezdyskusyjne – wszystkie one są zapisane w języku chińskim, niezrozumiałym dla człowieka mieszczącego się w pokoju. Również inna rzecz jest wspólna i niezmiernie istotna dla dalszego ciągu mego zarzutu. Wszystkie te kartki i książki są w rzeczywistości instrukcjami, które mają odzwierciedlać algorytmiczne działanie komputera: „JEŻELI TO, to TO" [*IF... THEN...*]. Pytanie moje i zarzut jednocześnie to: A SKĄD W POKOJU ZNAJDUJĄ SIĘ NIBY *WSZYSTKIE ODPOWIEDZI* NA *WSZYSTKIE PYTANIA*? Czyli, dlaczego Searle nie zauważa, że myślenie zdefiniowane w słynnej pracy Turinga polega właśnie na swego rodzaju generowaniu nowych zdań nie zapisanych w bazie danych. Pozostaje również kwestia możliwości skonstruowania takiej bazy danych, której de facto skonstruować się nie da. Zatem argument Chińskiego Pokoju o tyle nie dotyka problemu myślenia maszyn, że w Chińskim Pokoju nie mamy z czynnościami myślenia (w sensie generowania i transformacji, jak to nakreślił m.in. Chomsky). Z jedyną rzeczą, która mamy w nim do czynienia to z przysłowiowym „papugowaniem"!

1.10. Chiński Pokój – *summa sumarum*

Podsumowując problem Chińskiego Pokoju powołam się na słowa prof. W. Ducha:

> W sztucznej inteligencji argument Searle'a w ogóle jest nie przydatny. Nie jest, jak test Turinga, swego rodzaju testem, przy którego zastosowaniu moglibyśmy dowiedzieć się czy coś myśli, czy nie. Co byśmy nie postawili za kandydata, Chiński Pokój zawsze powie nam: To nie jest maszyna myśląca.
>
> [W. Duch, odczyt: *Czas zamknąć Chiński Pokój*, na Ogólnopolskiej Konferencji Kognitywistycznej w Toruniu].

Mówiąc bardziej metodologicznie, prócz walki na argumenty i kontrargumenty, Searle zdaje się podzielać tezę, że w myśl Chińskiego Pokoju, maszyny za grosz myśleć nie będą. Za nim zaś tą hipotezę broni m.in. J. Kloch w swoim tekście dla pisma "Znak". To, co jest, moim zdaniem, słabą stroną tego argumentu, a co tak ściśle wyraził prof. Duch, jest fakt nie weryfikowalności tezy Searle'a. Krótko mówiąc, uniemożliwia on falsyfikację hipotezy w myśl metodologii tak Poppera, jak i Kuhna.

1.10.1. Kilka słów od prof. W. Ducha

[Poniższy fragment, wraz z recenzją tekstu, przesłał mi prof. Duch. Wydaje się być znakomitym uzupełnieniem wszystkiego tego, co powyżej na temat dyskusji wokół Chińskiego Pokoju zostało przedstawione].

„Czym są symbole, które człowiek rozumie i jaki jest ich stosunek do symboli używanych przez programy? Searle, podobnie jak wielu filozofów, ma trudności w zrozumieniu idei reprezentacji. W popularnej książeczce o filozofii (Popkin i Stroll, 1994) znajdujemy dość absurdalne zarzuty w stosunku do monistycznych rozwiązań: to, co znajdujemy w mózgu nie przypomina wcale obrazów, dźwięków czy kolorów. Podglądanie telewizora czy magnetowidu od środka nie pokaże nam również żadnych obrazków. Do interpretacji informacji, zapisanej w postaci odbijających światło kropek na dysku kompaktowym lub namagnetyzowania powierzchni dysku magnetycznego potrzebny jest system odtwarzający, układ fizyczny, który będzie przez tą informację sterowany. W mózgu człowieka zapis informacji o obrazach, wrażeniach, dźwiękach, słowach czy abstrakcyjnych koncepcjach może przyjmować dowolną postać a postrzeganie nie jest uwarunkowane samą tylko informacją – wymaga odpowiedniego odtwarzacza, a więc fizycznych elementów, które można wprowadzić w odpowiedni stan. W przypadku urządzeń technicznych może to być membrana głośnika czy ekran telewizora, w przypadku mózgu człowieka wydaje się, że do powstania świadomych wrażeń konieczny jest udział płatów czołowych.

Searle ma rację – nie wystarczy sam przepływ informacji, konieczna jest fizyczna materia. Nie chodzi jednak o moc przyczynową neuronów, ale o miejsce i sposób powstawania wrażeń. Symbole mają dla nas znaczenie, bo wibracje sieci neuronów płatów czołowych automatycznie aktywują ślady pamięci związane z danym symbolem, jego rozliczne skojarzenia, co pozwala w płynny sposób na błądzenie myślami, przywodzenie na myśl wrażeń zmysłowych, czyli reprezentacji stanów kory sensorycznej, które wytwarzają się na skutek oddziaływania bodźców fizycznych na nasze receptory zmysłowe. Informacja, jej struktura, program działania umysłu – są bardzo ważne, ale nie mogą zastąpić fizycznego substratu doznającego odpowiednich stanów. Symulacja elektroniczna dźwięku traktowanego jako zjawisko fizyczne, czyli zmiany ciśnienia powietrza, nie prowadzi do powstania dźwięku

a jedynie do jego opisu. Informacja, czyli program i dane, pozwalające odtworzyć dźwięk, mogą być doskonałe, ale bez urządzenia wytwarzającego odpowiednie stany w fizycznym substracie, jakim jest powietrze, czyli bez głośnika czy słuchawek, samo odtwarzanie informacji w komputerze na nic się nie przyda. Mózg nie działa w oparciu o program komputerowy, ale nie ma też powodu, dla którego sztuczne systemy, oparte na mózgopodobnej organizacji przetwarzania informacji, miałyby być określone przez syntaktykę jakiegoś programu. Działanie takich systemów będzie bowiem określone przez procesy uczenia i nieprzewidywalne oddziaływania ze środowiskiem.

Czy zaglądając do mózgu człowieka i umieszczając tam obserwatora możemy coś dostrzec? Czy pojawi się rozumienie w umyśle obserwatora? Podobnie jak z problemem zrozumienia, jak to jest być nietoperzem (lub kimkolwiek innym)[17] mamy tu do czynienia z rzeczywistymi stanami fizycznej materii mózgu, specyficznymi dla danego organizmu. Pomysł Searle'a jest chybiony, gdyż nie można patrząc z zewnątrz osiągnąć zrozumienia prywatnych stanów umysłu. Jakie są warunki powstania zrozumienia w naszym umyśle? Jest to zwykle projekcja własnego stanu umysłu na innych ludzi, dokonana na podstawie obserwacji. Nie mamy jednak argumentów, by uzasadnić, że inni ludzie naprawdę przeżywają świat w podobny sposób, co my. Istnieje, przynajmniej teoretycznie, druga możliwość, jaką jest zestrojenie naszego mózgu z elektrycznymi wibracjami mózgu innego człowieka na tyle, by odczuć zrozumienie przychodzących do nich danych reprezentujących symbole czy wrażenia.

Gdybyśmy zamiast demona, który dostaje dane w postaci symboli wyobrazili sobie demona, którego mózg zestrojony jest z programem, sterującym stanami sztucznego mózgu, gdyby procesy tam następujące zachodziły z odpowiednią dla ludzkiego mózgu szybkością, wówczas demon mógłby odczuć znaczenie przepływu informacji i mieć podobne wrażenia, jak symulowany umysł. Zrozumienie znaczenia, nadawanego przez symulowany umysł dochodzącym do niego symbolom, pojawiłoby się w umyśle demona i nie miałby on już wątpliwości, że sztuczny mózg zdolny jest do prawdziwych stanów intencjonalnych. Program, który za tym stoi, jest warunkiem koniecznym utrzymania odpowiedniego przepływu informacji i odpowiedniej synchronizacji zdarzeń w sztucznym mózgu, ale nie wystarcza do wywołania intencjonalności, zawiera jedynie część syntaktyczną pozwalającą na formalne przetwarzanie symboli. Jednakże takie doświadczenie myślowe jest trudne do wyobrażenia, gdyż nasz umysł musiałby działać całkowicie biernie, a więc być zdominowany przez zestrojony z nim system, w przeciwnym przypadku przeżywalibyśmy jedynie własne interpretacje pobudzeń naszego mózgu, a nie stany zestrojonego z nim umysłu.

Z drugiej strony kognitywiści też mają rację – termostat ma „intencje", chociaż są one bardzo prymitywne[18], niewiele przypominające intencje człowieka. Stany termostatu i jego działanie zależne są od fizycznych właściwości materii, z której zrobiony jest termoelement. Searle rozpisuje się na temat absurdalności takiego poglądu, gdyż według niego nie pozwala to odróżnić umysłu od tego, co umysłem nie jest i prowadzi do panpsychizmu. Oczywista różnica leży w stopniu komplikacji. To właśnie niedostrzeganie ciągłości i stopniowego wzrostu złożoności funkcji prowadzi do sztucznych problemów, z którymi nie potrafi sobie poradzić filozofia umysłu. Jeśli za prawdziwe wrażenia uznamy jedynie te, które swoim stopniem subtelności i możliwością lingwistycznego wyrazu ma normalnie rozwinięty człowiek to oczywiście nie mają takich wrażeń ani zwierzęta, ani sztuczne urządzenia. Na drodze do prawdziwego umysłu są różne nieciągłości, np. pojawienia się autorefleksji, odróżnienie siebie od otoczenia, moment, w którym praczłowiek zdał sobie sprawę z tego, że

[17] Przypominam, postawionym przez T. Nalega. Przyp. M.K.
[18] Tak np. głosi D. C. Dennett. Por. Jego *Natura umysłów* i *Darwin's Dangerous Idea*. Przyp. M.K.

pewne rodzaje dźwięków dochodzą do niego z zewnątrz a pewne są wydawane przez niego samego. Wiele gatunków zwierząt i żadna maszyna nie przekroczyła tego stopnia rozwoju umysłu. Ciągłość rozwoju umysłu widoczna jest u ludzi z różnymi zaburzeniami neurologicznymi i niedorozwojami, widoczna jest w świecie zwierząt i będzie zapewne widoczna w świecie urządzeń sztucznych. Urządzenia przetwarzające informację mogą dobrze symulować procesy myślenia jednakże osiągnięcie intencjonalności, prawdziwego rozumienia i subtelnych skojarzeń będzie prawdopodobnie wymagać wprowadzenia stanów jakiegoś fizycznego substratu, którego subtelne wibracje będą odzwierciedlać treści mentalne, podobnie jak dzieje się to w naszym mózgu.

Test Turinga pozwala odrzucić programy, które na pewno nie można uznać za umysł od pozostałych. Eksperyment z chińskim pokojem zawsze prowadzi do negatywnego wyniku: obserwator lub wielu obserwatorów umieszczonych w mózgu nie osiągnie żadnego zrozumienia. Nie jest to więc test, lecz demagogicznie ustalony punkt widzenia odmawiający intencjonalności wszelkim systemom, w tym również mózgom ludzkim. W tym ostatnim przypadku zupełnie arbitralnie twierdzi się, że wiemy skądinąd, iż mózgi „rozumieją", nie precyzując sensu tego stwierdzenia. Skąd to jednak wiemy i co oznacza stwierdzenie Searle'a „mózgi są przyczyną umysłów"? Oznacza to, że efektów ich działania nie można odróżnić od działania umysłu, w więc są one intencjonalne, odczuwają, pojmują a nie tylko przetwarzają informację. W jaki sposób możemy się o tym przekonać? Powracamy tu znowu do testu Turinga, gdyż żadnego innego nie mamy."

2. Argument Rogera Penrose'a

Nad tym argumentem nie będę się dłużej rozpisywał, bo i szczerze mówiąc nie widzę większego sensu – bardzo dobrze wyjaśniał go sam sir R. Penrose w trzech swoich kolejnych książkach.

Argument ten można podzielić na zgoła dwie tezy:

1. Teza Gödla,
2. Niewyjaśnialność istoty działania umysłu przy pomocy współczesnych nauk (ze szczególnym zwróceniem na ograniczenia współczesnej fizyki). I ostatecznie głoszenie idei powstania nowej fizyki, dzięki której wyjaśnialność byłaby w pełni możliwa i satysfakcjonująca.

Co do pierwszej tezy to na miejscu wydają być słowa Alana Turinga, twórcy teorii rozstrzygalności:

> Krótką ripostą na ten argument jest to, że chociaż ustalono, że istnieją granice możliwości każdej poszczególnej maszyny, to jednak jedynie bez dowodu stwierdzono, że żadne takie ograniczenia nie stosują się do ludzkiego intelektu. – I dalej. – Nie jestem jednak zdania, że ten pogląd można zbyć tak łatwo. Za każdym razem, gdy jednej z tych maszyn zadaje się odpowiednio krytyczne pytanie i daje ona określoną odpowiedź, to wiemy, że ta odpowiedź musi być błędna i daje nam to pewne poczucie wyższości. Czyżby to uczucie było złudne? Jest ono bez wątpienia zupełnie niekłamane, ale myślę, że nie należy zbyt wielkiej wagi do niego przywiązywać. My sami zbyt często dajemy błędne odpowiedzi na pytania, aby można było usprawiedliwić nasze zadowolenie z takiego dowodu omylności części maszyn. (...) Tak więc, krótko mówiąc, mogliby być ludzie zdolniejsi od każdej danej maszyny, ale z kolei mogłyby być inne zdolniejsze maszyny itd.

[A. M. Turing, *Maszyny liczące a inteligencja*, s. 33].

Co do drugiej tezy to na miejscu wydają być słowa Jamesa Trefila, który moim zdaniem zupełnie trafnie zauważył ograniczoność argumentu Penrose'a:

> Przypuśćmy (...), że hipoteza Penrose'a okazuje się w pełni trafna. Przypuśćmy, że (1) mózg rzeczywiście nie jest cyfrowym komputerem i (2) dlatego nim nie jest, że działa według praw nowej nauki, której miejsce znajduje się na przecięciu fizyki klasycznej, mechaniki kwantowej i teorii unifikacji pola.
>
> (...) pomyślmy przez chwilę, co by się stało, gdyby teorię unifikacji pola już napisano i moglibyśmy śmiało wejść w lukę między fizyką kwantową a klasyczną. Wtedy, jeżeli Penrose ma rację, moglibyśmy zrozumieć działanie mózgu na poziomie cząsteczek i komórek.
>
> I co z tego? Najprawdopodobniej nadal moglibyśmy widzieć mózg jako maszynę działającą według znanych praw natury. Tyle że ta maszyna nie byłaby cyfrowym komputerem. Byłaby czymś innym, jak dotąd niewyobrażalnym, działającym według praw natury jeszcze przez nas nie poznanych.
>
> [J. Trefil, *Czy jesteśmy niepowtarzalni?*, s. 144].

W rezultacie, co z tego wynika? Wynikają bezspornie dwa następujące wnioski:

1. Jeżeli argument jest prawdziwy, zatem wystarczy stworzyć nową fizykę (jak chce sam Penrose), by rozwiązać problem poznawalności umysłu i jego symulacji, co w rezultacie doprowadzić może do maszyn myślących.
2. Jeżeli argument jest nieprawdziwy to... nie mamy się czym przejmować!

3. Argument z *prostoty* redukcjonizmu i *złożoności* świata

Przy nazwaniu tego rodzaju argumentu miałem spore trudności, lecz nie wypisanie go tutaj łączyłoby się ze „sporą dziurą" tekstu. Aby móc doświadczyć, o co chodzi w tego rodzaju całej rodzinie argumentów na rzecz niepowtarzalności człowieka i niewystarczalności redukcjonizmu dla możliwości skonstruowania maszyny myślącej, wystarczy przeczytać poniższy cytat:

> (...) Richard Dawkins[19] napisał w swej głośniejszej książce *Rzeka poza rajem*, że „życie to po prostu bajty, bajty i jeszcze raz bajty informacji cyfrowej". Cóż, dla niektórych świat jest wyjątkowo ubogi.
>
> [Ł. Gołębiowski, *Recenzja*: *Charles Jonscher*, *Życie okablowane*, s. 40].

Można by stwierdzić, że Łukasz Gołębiowski odczuwa pewnego rodzaju satysfakcję myśląc, że jego życie, o którym myśli, na pewno nie opiera się na czymś tak prostym jak bajty, jest bardziej bogatym i wartościowym życiem, niż życie Richarda Dawkinsa. Tylko Ł. Gołębiowski raczy nie zauważyć bogatości tworów, które złożone są z owych bajtów, czyli wszelkiego rodzaju informacji cyfrowej (począwszy od pojedynczych plików, przechodząc przez *sofware*, całe olbrzymie niekiedy bazy danych, a skończywszy na Internecie). To

[19] Twórca teorii memów; autor takich książek jak chociażby: *Samolubny gen*, *Rzeka genów*. Przyp. M.K.

złożoność i kombinacja wprowadza nam czynnik różnorodności i bogatości świata. To teza prawdziwa już od Darwina.

Próbując sparafrazować Ł. Gołębiowskiego, co aż się prosi, można by ułożyć następujące twierdzenie:

> Łukasz Gołębiowski napisał w swej recenzji pewnej poczytnej książki, że życie to nie bajty (czyli dwubitowe (0,1) ciągi 8-znakowe), bajty i jeszcze raz bajty, jakby zapominając, że u podstaw każdego życia leży genotyp (odpowiednio skomplikowana struktura czteroznakowa (A, G, T, C – od pierwszych liter aminokwasów).

(Szczerze mówiąc, będąc bardziej złośliwym, lecz i dosadnym (co czasem plusami przyćmiewa ewentualne minusy bycia złośliwym) w określeniu tego, co mam na myśli, można użyć również następującego porównania:

> Niektórzy mawiają, że wszystko w świecie, co nazywamy materią, składa się z atomów. „Cóż, dla niektórych świat jest wyjątkowo ubogi", pewnie skomentowałby Ł. Gołębiowski).

Krótko mówiąc, nie widzę powodu dlaczego np. języka genotypu (który leży u podstaw życia, co jest tezą (zatem zdaniem udowodnionym), a nie hipotezą) nie móc przełożyć na język bitów i bajtów, w dalszej kolejności. I, dodajmy dla Ł. G., to w dodatku tak, żeby oddać całą bogatość świata!

Spłaszczenie zaś teorii redukcyjnych, chociaż tak chwytne dla wielu czytelników o humanistycznym zacięciu (humanistów, w sensie nie potrafiących przełamać trudności matematycznego myślenia!), choć przysparza w dzisiejszych czasach wielu czytelników, to nie stanowi żadnego rodzaju poważnego problemu na rzecz zarzucenia podejść redukcjonistycznych – w tym komputerowo redukcjonistycznych.

Grudzień 2001 – styczeń 2002

Literatura:

[1] D. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press, Oxford 1996.

[2] P. M., P. S. Churchland, *Czy maszyny mogą myśleć?*, w: "Świat Nauki" (*Scientific American*), lipiec 1991, ss. 17-23.

[3] W. Duch, *Jaka teoria umysłu w pełni nas zadowoli?*, w: "Kognitywistyka i Media w Edukacji", Nr 1-2/2000, ss. 29-53; http://www.phys.uni.torun.pl/~duch/.

[4] G. M. Edelman, *Przenikliwe powietrze, jasny ogień. O materii umysłu*, tłum. J. Rączaszek, PWN, Warszawa 1998.

[5] D. Hume, *Badania dotyczące rozumu ludzkiego*, tłum. J. Łukasiewicz, K. Twardowski, DeAgostini & Altaya, Warszawa 2001.

[6] M. Kasperski, *Wczoraj, dzisiaj, jutro Sztucznej Inteligencji*; komputeropis książki: *Sztuczna Inteligencja. Droga do myślących maszyn*, Helion, Gliwice 2003.

[7] M. Kasperski, *A język lata, lata jak łopata. O problemie świadomości w filozofii Sztucznej Inteligencji*, w: „G.N.O.M.", Nr 3/2001, ss. 61-74; http://www.kognitywistyka.net.

[8] J. Kloch, *Chiński Pokój. Eksperyment myślowy Johna Searle'a. Studium historyczno-filozoficzne (cz. 2)*, w: "Znak"; http://www.opoka.org.pl/biblioteka/F/FG/searle_2.html. Całość rozważań w: J. Kloch, *Świadomość komputerów?*, Wyd. Biblos, Tarnów 1996.

[9] S. Lem, *Tajemnica chińskiego pokoju*, w: tenże, *Tajemnica Chińskiego Pokoju*, Universitas, Kraków 1996, ss. 201-208.

[10] S. Lem, *Brain Chips*, w: tenże, *Tajemnica chińskiego pokoju*, Universitas, Kraków 1996, ss. 141-150.

[11] S. Lem, *Brains Chips III* w: tenże, Tajemnica chińskiego pokoju, Universitas, Kraków 1996, ss. 159-166.

[12] W. Marciszewski, *Sztuczna inteligencja*, Wyd. Znak, Kraków 1998.

[13] M. Nowicki, *Syntaktyczne twierdzenia limitacyjne, wyłożone sposobem Turinga, z konkluzjami Chaitina*, http://www.kognitywistyka.net.

[14] R. Penrose, *Nowy umysł cesarza*, tłum. P. Amsterdamski, Wydawnictwo Naukowe PWN, Warszawa 1995; Wyd. II Warszawa 2000.

[15] R. Penrose, *Makroświat, mikroświat i ludzki umysł*, tłum. P. Amsterdamski, Prószyński i S-ka, Warszawa 1997.

[16] R. Penrose, *Cienie umysłu*, tłum. P. Amsterdamski, Zysk i S-ka, Poznań 2000.

[17] J. R. Searle, *Umysł, mózg i nauka*, tłum. J. Bobryk, Wydawnictwo Naukowe PWN, Warszawa 1995.

[18] J. R. Searle, *Umysły, mózgi i programy*, tłum. B. Chwedeńczuk, w: *Filozofia umysłu*, red. B. Chwedeńczuk, Spacja, Warszawa 1995.

[19] J. R. Searle, *Umysł na nowo odkryty*, tłum. Z. Baszniak, PIW, Warszawa 1999.

[20] J. Trefil, *Czy jesteśmy niepowtarzalni?*, tłum. E. Życieńska, Amber, Warszawa 1998.

[21] A. M. Turing, *Maszyny liczące a inteligencja*, tłum. D. Gajkowicz, w: *Maszyny matematyczne i myślenie*, red. E. Feigenbaum, J. Feldman, PWN, Warszawa 1972, ss. 24-47.