

Józef KLOCH

CHIŃSKI POKÓJ  
EKSPERYMENT MYŚLOWY JOHNA SEARLE'A<sup>†</sup>STUDIUM HISTORYCZNO-FILOZOFICZNE  
(cz. 1)

Historia nauki pełni funkcję pomocniczą względem filozofii, może nawet stanowić swego rodzaju przestrzeń testowania mechanizmów rozwoju danej dziedziny poznania<sup>1</sup>. Dzieje eksperymentu myślowego Johna Searle'a stanowią natomiast interesujący przyczynek do badań nad sztuczną inteligencją i do filozofii AI; ukazują też stopniową zmianę nastawienia do dowodu — od wręcz wrogiego do coraz bardziej przychylnego, doceniającego wagę argumentu. Chiński Pokój ma szerokie zastosowanie w badaniach nad sztuczną inteligencją — od prostych programów komputerowych po skomplikowane sieci neuropodobne, w filozofii zaś — w teorii umysłu oraz w kwestii wzajemnego odniesienia semantyki i syntaktyki.

## EWOLUCJA ARGUMENTU SEARLE'A

Pytanie dotyczące myślenia maszyn było wielokrotnie stawiane w ciągu ostatnich kilku dziesięcioleci, zmieniało się jego zabarwienie filozoficzne; na początku więc sprecyzujemy obecną postać problemu. Sięgnijemy następnie do programu Schanka i filozoficznych poglądów na temat rozumienia przez programy komputerowe, a także do koncepcji związanych z modelowaniem ludzkiego umysłu. Chiński Pokój był sprzeciwem profesora z Berkeley wobec tzw. mocnego programu sztucznej inteligencji (*strong Artificial Intelligence*,

---

\*UWAGA: Tekst został zrekonstruowany przy pomocy środków automatycznych; możliwe są więc pewne błędy, których sygnalizacja jest mile widziana (obi@opoka.org). Tekst elektroniczny posiada odrębną numerację stron.

<sup>1</sup>Niniejszy artykuł jest pierwszą częścią większej całości przygotowywanej do druku.

<sup>1</sup>Por. M. Heller, *Filozofia nauki. Wprowadzenie*, PAT, Kraków 1992, 12.

skrót: sAI). Przeanalizujemy też ewoluowanie postaci samego dowodu — od wersji pierwotnej z 1980 do najnowszej z 1990 roku. Przedstawimy też dwa warianty argumentu Searle'a; są one interesującym materiałem w dyskusji nad rolą tzw. przetwarzania równoległego oraz nad stosunkiem syntaktyki do semantyki.

## 1. PRECYZOWANIE PYTANIA

Pytanie Alana M. Turinga: „Czy maszyny mogą myśleć?” przybierało w ciągu dziesięcioleci coraz to nową postać w analizach badaczy AI oraz filozofów. Kolejnymi wersjami powyższej kwestii postawionej w 1950 roku były zagadnienia: „Czy maszyna może uzyskać zdolność myślenia przy pomocy programu uruchomionego na niej?” oraz „Czy program komputerowy może mieć wbudowaną zdolność myślenia?” Zauważmy, że w kolejnych pytaniach punkt ciężkości został przeniesiony z fizycznych własności maszyn istniejących bądź mogących zaistnieć na obliczeniowe, abstrakcyjne możliwości formalnych programów komputerowych. Kolejną postacią wyżej wymienionych zagadnień jest porównanie: „Umysł ludzki jest dla mózgu tym, czym program dla komputera”. Nasuwają się też inne kwestie: „Czy umysł ludzki jest programem komputerowym?” oraz „Czy można powiedzieć, że program komputerowy rozumie?” czy wreszcie „Czy komputer może mieć umysł?”

Zwolennicy tzw. mocnego programu sztucznej inteligencji uznają możliwość wbudowania zdolności myślenia do programu komputerowego. Zgodnie z tym stanowiskiem komputer to nie tylko *narzędzie* w studiowaniu umysłu; poprawnie zaprogramowany komputer o prawidłowych informacjach wejściowych i wyjściowych jest umysłem — rozumie i ma stany umysłowe<sup>2</sup>. Programy natomiast nie są zwykłymi narzędziami umożliwiającymi testowanie psychologicznych wyjaśnień — one są same przez się wyjaśnieniem (*explanation*) sposobu działania umysłu.

Tego typu opinie spowodowały sprzeciw m. in. Johna Searle'a, profesora filozofii z Berkeley. Słynny eksperyment myślowy zwany przez niego „Chińskim Pokojem” wykazuje błędność twierdzenia o posiadaniu przez komputer stanów umysłowych i wyjaśnianiu działania ludzkiego umysłu przy pomocy programów komputerowych. Searle opublikował w 1980 roku artykuł zawierający ów dowód na łamach amerykańskiego pisma „The Behavioral and

---

<sup>2</sup>J. R. Searle, *Minds, Brains, and Programs*, „The Behavioral and Brain Sciences” 3 (1980), 417. W dalszej części pracy do cytowanego artykułu stosuję skrót BBS.

Brain Sciences”, którego zwyczajem jest umieszczanie głównego artykułu wraz z replikami innych naukowców oraz odpowiedzią autora na nie.

### 1.1 Program Schanka

Jako przykład programu realizującego założenia sAI Searle przytacza pracę Rogera Schanka i jego kolegów z Yale<sup>3</sup>, ale zastrzega się jednocześnie, że argument Chińskiego Pokoju odnosi się do jakiegokolwiek symulacji ludzkiego umysłu dokonywanej przez dowolną maszynę Turinga przy pomocy formalnego programu komputerowego. Celem programu Schanka jest naśladowanie rozumienia opowiadań; program odpowiada na pytania dotyczące informacji nie zawartych wprost w określonym zdarzeniu — w tym wypadku dotyczą one zamawiania hamburgera w restauracji. Komputer dysponuje pewnym zakresem wiadomości na temat restauracji oraz opisem sytuacji: klient po otrzymaniu przypalonego hamburgera nie zapłacił za niego, nie dał napiwku i zdenerwowany wybiegł z restauracji. Jeśli postawimy komputerowi pytanie, czy klient zjadł hamburgera, otrzymamy odpowiedź negatywną. Jeśli do komputera zostanie wprowadzony inny opis sytuacji, w której hamburger był smacznie przyrządzony i klient zapłacił oraz dał napiwek kelnerce, na pytanie dotyczące zjedzenia hamburgera komputer odpowie pozytywnie. Takich samych odpowiedzi należałoby się spodziewać od człowieka. Czy to oznacza, że program Schanka ma wbudowaną zdolność myślenia? Czy w ogóle można wbudować tę zdolność do jakiegokolwiek programu komputerowego? Jakie treści filozoficzne związane są z symulowaniem ludzkiego myślenia?

W przypadkach podobnych do opisanych powyżej zwolennicy sAI są zdania, że komputer *rozumie* opowiadanie oraz że działanie komputera wyposażonego w odpowiedni program *wyjaśnia* zasady działania ludzkiego umysłu. Artykuł Searle’a z 1980 roku wykazuje bezzasadność tych twierdzeń.

### 1.2 Searle’a rozumienie *rozumienia*

Zanim przejdziemy do opisu dowodu, warto przyjrzeć się samemu pojęciu rozumienia — co to znaczy że „*a rozumie b*”? Searle rezygnuje w tym wypadku ze złożonych analiz językowych i na konkretnym przykładzie wyjaśnia swoje literalne pojmowanie *rozumienia*:

---

<sup>3</sup>R. C. Schank, R. P. Abelson, *Scripts, Plans, Goals, and Understanding*, Lawrence Erlbaum Press, Hillsdale, N. J., 1977.

„Rozumiem opowiadania po angielsku, w mniejszym stopniu mogę zrozumieć po francusku, w jeszcze mniejszym po niemiecku, a w ogóle [nie rozumiem] po chińsku. Mój samochód, mój kalkulator nie rozumieją niczego — to nie jest ich zakres działania. Często przez analogię lub metaforycznie przypisujemy „rozumienie” i inne kognitywne predykaty samochodom, kalkulatorom i innym artefaktom, ale nic nie zostaje udowodnione poprzez podobne atrybuty”<sup>4</sup>.

Takie pojmowanie *rozumienia* wystarczy Searle’owi do analiz związanych z argumentem Chińskiego Pokoju. Analogiczne czy metaforyczne przypisywanie rozumienia artefaktom ma miejsce w stosunku do różnych maszyn. Mówimy np. o „odczuwaniu” przez termostat zmian temperatury, „umiejętnościach” kalkulatora odnośnie do potęgowania czy też „wiedzy” odtwarzacza płyt kompaktowych co do kolejności utworów w wypadku uruchomienia funkcji *shuffle*. Wykonane przez człowieka narzędzia są związane z celem jego działania, dlatego naturalną rzeczą jest przypisywanie im atrybutów przysługujących ludzkim czynnościom. Musimy jednak pamiętać, że dzieje się tak jedynie na zasadzie analogii, metafory. Mimo tego zastrzeżenia, wielu uczonych zajmujących się badaniami nad sztuczną inteligencją przypisuje tego samego rodzaju poznanie komputerom i ludziom<sup>5</sup>. Searle natomiast dowodzi, że

„w sensie dosłownym zaprogramowany komputer rozumie tyle co samochód lub maszyna licząca a mianowicie dokładnie nic. Rozumienie komputera nie jest ani częściowe, ani niekompletne (jak moje rozumienie niemieckiego) — jest żadne”<sup>6</sup>.

O niemożności zrozumienia czegokolwiek przez zaprogramowany komputer decyduje według filozofa z Berkeley intencjonalność, którą rozumie jako własność stanów umysłowych, przez którą są one ukierunkowane ku lub wokół tego co dzieje się w świecie<sup>7</sup>. Intencjonalność jest produktem przyczynowych własności mózgu i elementem odróżniającym ludzkie myślenie od działania programu komputerowego. Searle zakłada, że jest to fakt empiryczny, biorąc pod uwagę relacje przyczynowe między mózgiem a procesami umysłowymi.

<sup>4</sup>J. R. Searle, *Minds, Brains, and Programs*, BBS 3 (1980), 419.

<sup>5</sup>Patrz np. A. Newell, H. A. Simon, *GPS* — program, który symuluje myśl ludzką, w: MMiM, 278–290.

<sup>6</sup>J. R. Searle, *Minds, Brains, and Programs*, BBS 3 (1980), 419.

<sup>7</sup>Tamże.

W odniesieniu natomiast do relacji komputer — program komputerowy intencjonalność nie będzie miała miejsca<sup>8</sup>.

## 2. CHIŃSKI POKÓJ — OPIS I HISTORIA DOWODU

Głównym celem artykułu *Minds, Brains, and Programs* jest udowodnienie twierdzenia o niewystarczalności pogromu komputerowego do wytworzenia intencjonalności oraz analiza konsekwencji tego twierdzenia.

By to wykazać, Searle przyjął w swym dowodzie metodę zastąpienia konkretnego programu komputerowego przez człowieka i postawił sobie pytanie: co by się stało, gdyby umysł człowieka zaczął pracować na podobnej zasadzie jak program Schanka, tj. odpowiadałby na pytania? Co można byłoby wtedy powiedzieć o intencjonalności?

### 2.1 Pierwotna wersja argumentu (1980, 1982)

Wyobraźmy sobie zamkniętego w pokoju człowieka, który nie zna w ogóle chińskiego, ani w mowie ani w piśmie, i nie potrafi go odróżnić od japońskiego czy od symboli podobnych do obu języków. Człowiek ów ma do dyspozycji mnóstwo chińskich znaków — nazwijmy je zestawem nr 1; następnie zostają mu przekazane teksty chińskie (zestaw nr 2) wraz z zasadami po angielsku — okreśmy je roboczo jako plik  $A^9$ . Instrukcje te pozwalają na wzajemne powiązanie ze sobą zestawu nr 1 i nr 2 jedynie na podstawie kształtu znaków. Następnie człowiekowi jest dostarczony zbiór tekstów chińskich nr 3 również z angielskimi instrukcjami (plik  $B$ ); informują one, w jaki sposób skorelować zestaw nr 3 z zestawami nr 1 i 2, by umożliwić utworzenie pliku znaków chińskich nr 4.

Zamknięty w pokoju człowiek nie wie, że ludzie z zewnątrz przypisali danym zestawom określone nazwy: tekst (*a script*), opowiadanie (*a story*), pytania oraz odpowiedzi na pytania; zasady w języku angielskim określone są jako program. Zakładamy, że instrukcje są bardzo jasne, innymi słowy — program jest napisany tak doskonale, że rodzony Chińczyk nie miałby wątpliwości, iż odpowiedzi daje ktoś, kto od urodzenia mówi po chińsku.

Przeprowadźmy obecnie ów eksperyment myślowy z tym samym człowiekiem, ale w odniesieniu do ojczystego języka, jakim jest dla niego an-

---

<sup>8</sup>Tamże, 417.

<sup>9</sup>W omawianym opisie dowodu Searle stosuje oznaczenia numeryczne w odniesieniu do poszczególnych zestawów symboli chińskich. Dla łatwiejszego ukazania ewoluowania dowodu Chińskiego Pokoju wprowadzam też oznaczenia obu części zasad (programu), czego autor już nie czyni.

gielski; pytania i odpowiedzi również są dawane w tym języku. Dla kogoś z zewnątrz, kto od urodzenia mówi po angielsku, odpowiedzi dawane przez człowieka zamkniętego w pokoju będą doskonałe. Zachodzi jednak istotna różnica między dawaniem odpowiedzi po chińsku i po angielsku; w wypadku tego ostatniego języka miało miejsce pełne rozumienie, podczas gdy odpowiedzi w chińskim były „produkowane” bez zrozumienia, na zasadzie dobierania jednych niezinterpretowanych znaków do drugich według ściśle określonych instrukcji (programu). W tym wypadku człowiek w pokoju zachowywał się jak komputer — wykonywał jedynie operacje na chińskich symbolach, zastępując konkretny program.

Tak wyglądała pierwotna wersja argumentu Chińskiego Pokoju; sprowokował on aż 27 odpowiedzi uczonych z kręgu AI, filozofów oraz psychologów<sup>10</sup>. Searle w replice na te artykuły wyjaśnił zagadnienia wzbudzające polemikę, sprecyzował wiele pojęć, omówił szereg problemów, ale nie wprowadził żadnych zmian do samego dowodu ani w 1980 roku, ani w 1982 roku<sup>11</sup>, kiedy to siedem kolejnych komentarzy do słynnego artykułu zostało opublikowanych w tym samym czasopiśmie „The Behavioral and Brain Sciences”.

## 2.2 Przypowieść o Chińskim Pokoju

Searle powraca do swojego eksperymentu myślowego podczas sympozjum dotyczącego naukowego, intelektualnego i społecznego oddziaływania techniki komputerowej; odbyło się ono w Nowojorskiej Akademii Nauk (5–8 IV 1983).

W panelowej dyskusji<sup>12</sup> na temat roli AI w wyjaśnianiu ludzkiego myślenia Searle wykazuje fałszywość mocnego programu AI, powołując się na swój dowód, określając go tym razem mianem *przypowieści o Chińskim Pokoju*. Searle podkreśla niemożność przejścia od semantyki do syntaktyki — z czysto formalnych operacji na symbolach nie wynika treść umysłowa, na podstawie komputacyjnych procesów człowiek zastępujący program nigdy nie nauczyłyby się chińskiego. Zwraca też uwagę, by w dyskusji na temat sAI

<sup>10</sup> Patrz: BBS 3 (1980), 424–450.

<sup>11</sup> J. R. Searle, *The Chinese Room Revisited*, BBS 5,2 (1982), 345–348.

<sup>12</sup> *Has Artificial Intelligence Research Illuminated Human Thinking?* Panel Discussion with Hubert L. Dreyfus, John McCarthy, Marvin L. Minsky, Seymour Papert, and John Searle, w: *Computer Culture: The Scientific, Intellectual, and Social Impact of the Computer*, ed. by Heinz R. Pagels, „Annals of the New York Academy of Sciences”, 426, New York 1984, 145–148.

podchodzić do zagadnienia z punktu widzenia pierwszej a nie trzeciej osoby, zarówno bowiem świadomość jak i procesy myślowe zachodzące w człowieku dotyczą pierwszej osoby.

W odniesieniu do samego Chińskiego Pokoju Searle wprowadza pewne zmiany — w miejsce dwóch zbiorów chińskich symboli (nr 1 i 2) wprowadza jeden (bazę danych); zamiast o dwóch częściach zasad mówi o wszelkich książkach zawierających zasady dobierania jednych symboli chińskich do drugich. W ten sposób Searle uprościł dowód, rezygnując częściowo z upodabniania programu do Schankowskiego. Stopniowo Searle będzie dążył do nadania argumentowi postaci dowolnego formalnego programu. W wersji dowodu z 1983 roku, jak i zresztą później, nie ma żadnych zmian w dowodzie odnośnie do danych na wejściu i wyjściu — nadal są nimi pytania (obecnie zestaw nr 2), oraz odpowiedzi na pytania (zestaw nr 3).

### 2.3 Umysł — mózg — świadomość (1984, 1986)

W 1984 roku rozgłośnia BBC powierzyła Searle'owi *Wykłady Reith'a*; sześć półgodzinnych audycji<sup>13</sup> dotyczyło krytyki materialistycznego punktu widzenia w odniesieniu do umysłu. W drugim wykładzie zatytułowanym *Can Computers Think?* Searle cytuje z pewnymi zmianami dowód Chińskiego Pokoju w wersji z 1983 roku. O ile nadal jest mowa o trzech zbiorach symboli, o tyle zasady odnoszące się do manipulowania znakami zawarte są w jednej książce. Warto też podkreślić rozwój filozoficznego komentarza do Chińskiego Pokoju — w *Wykładach Reith'a* po raz pierwszy zostały sformułowane cztery logiczne tezy (*thesis*) dotyczące relacji między umysłem, mózgiem a komputerem. Na podstawie tych przesłanek Searle formułuje w tym samym drugim wykładzie<sup>14</sup> również 4 wnioski.

W 1986 roku, od 9 do 12 stycznia, odbywał się w Bombaju światowy Kongres na temat syntezy nauki i religii, na który zaproszeni byli m. in. Joseph Weizenbaum z MIT oraz John Searle; ich prelekcje dotyczyły zagadnienia *Umysł, mózg, świadomość*. W kontekście mocnego programu sztucznej inteligencji oraz testu Turinga Searle przytacza swój argument w wykładzie pod znamienym tytułem: *Turing the Chinese Room*<sup>15</sup>. Zawiera on te

---

<sup>13</sup>W tym samym roku ukazały się one w formie książkowej: J. R. Searle, *Minds, Brains, and Science*, Harvard University Press, Cambridge 1984. Pozycja ta została przetłumaczona na 7 języków.

<sup>14</sup>Na temat tez i wniosków patrz osobny paragraf.

<sup>15</sup>W: *Synthesis of Science and Religion: Critical Essays and Dialogues*. World Coness for the Synthesis of Science and Religion (1986: Bombay, India), ed. T. D. Singh, Ravi

same elementy co w roku 1984; jedynie w odniesieniu do książki z zasadami dobierania symboli zaznacza, że jest ona w języku, który dobrze zna biorący udział w teście, ale modyfikacja ta nie ma większego znaczenia. Pewna zmiana ma natomiast miejsce co do nazwy tez z 1984 roku — nazwane są one tutaj twierdzeniami (*propositions*), ale treść pozostaje zasadniczo ta sama. Zapewne ze względu na temat kongresu Searle dodał jeszcze jedno twierdzenie (umysł przyczynuje zachowanie), którego już nie powtarza w opisie dowodu w 1990 roku. Dodany też został piąty wniosek (mózg przyczynuje zachowanie) z powodu wprowadzenia piątego twierdzenia; treść pozostałych wniosków została zachowana.

#### 2.4 Wersja z 1990 roku

Dowód Chińskiego Pokoju został obszernie opisany przez Searle'a raz jeszcze w artykule opublikowanym w styczniu 1990 roku na łamach „Scientific American”<sup>16</sup>; późniejsze prace<sup>17</sup> mające odniesienie do Chińskiego Pokoju nawiązują do wersji z 1990 roku. Trzy pliki symboli chińskich są nazwane odpowiednio: pierwszy — bazą danych, drugi (zestaw znaków dostarczany na wejściu) — pytaniami, trzeci (zestaw na wyjściu) — odpowiedziami. Programem komputerowym nazwana jest książka z zasadami dobierania symboli na podstawie kształtu, bez znajomości ich znaczenia; autorzy książki określani są jako programiści a człowiek zamknięty w pokoju jest wyobrażeniem komputera. Cztery twierdzenia przyjmują tu nazwę aksjomatów a ich treść, podobnie jak i czterech wniosków, nie zmienia się.

### 3. WARIANTY DOWODU SEARLE'A

Chiński Pokój stał się bodźcem do powstania innych eksperymentów myślowych będących wariantami dowodu Searle'a, a mianowicie „Koreańskiego Pokoju” oraz „Chińskiej Sali Gimnastycznej”.

W 1988 roku William Rapaport, wykładowca informatyki na Uniwersytecie Stanu Nowy Jork w Buffalo, opublikował artykuł<sup>18</sup> dotyczący podstaw komputacyjnego rozumienia języka naturalnego. Praca ta zawiera m. in.

---

Gomatam, Bhaktivedanta Institute, San Francisco 1987, 295–301. W dalszej części pracy w odniesieniu do cytowanej pozycji stosuje skrót SoSaR.

<sup>16</sup> *Is the Brain's Mind a Computer Program?*, January (1990), 26–31.

<sup>17</sup> Np. J. R. Searle, *The Rediscovery of the Mind*, MIT Press, Cambridge 1992.

<sup>18</sup> W. Rapaport, *Syntactic Semantics: Foundations of Computational Natural-Language Understanding*, w: *Aspects of Artificial Intelligence*, ed. James H. Fetzer, Kluwer Academic Publishers, Dordrecht 1988, 81–131.

tw. „Koreański Pokój” zasugerowany profesorowi Rapaportowi przez jego studenta Alberta Hanyoung Yuhana; argument ten był wymierzony przeciw głównej idei Chińskiego Pokoju dotyczącej rozumienia.

Dwa lata później sam Searle opracował inną wersję swojego dowodu; powodem stały się nadzieje związane z równoległym przetwarzaniem wzorowanym na budowie układu nerwowego. Paul M. Churchland i Patricia Smith Churchland, profesorzy filozofii na Uniwersytecie Kalifornijskim w San Diego, twierdzili, że istnieją realne szanse zbudowania myślącej maszyny. By tak się stało, sztuczne systemy muszą naśladować system biologiczny mózgu. Dyskusja na łamach „Scientific American” między Churchlandami a Searlem zaowocowała w postaci m. in. „Chińskiej Sali Gimnastycznej”, wariantu Chińskiego Pokoju dla programów współbieżnych.

### 3.1 Koreański Pokój

Pomysł studenta Alberta H. Yuhana został podjęty i opracowany przez jego wykładowcę Williama Rapaporta. Otóż wyobraźmy sobie światowej sławy autorytet w dziedzinie twórczości Szekspira — koreańskiego profesora literatury angielskiej na uniwersytecie w Seulu; ów wykładowca czytał znakomite, ale wyłącznie tylko koreańskie, tłumaczenia dzieł dramaturga i nie zna w ogóle angielskiego; jest natomiast ekspertem w swojej dziedzinie. Na temat sztuk scenicznych wspomnianego autora napisał po koreańsku szereg artykułów; zostały one przetłumaczone na angielski, opublikowane w wielu liczących się anglojęzycznych czasopismach naukowych i spotkały się ze znakomitym przyjęciem. Pytanie, zadane przez Rapaporta, brzmi: „Czy koreański profesor *rozumie* Szekspira? Zauważmy, że inaczej niż w Chińskim Pokoju problem nie dotyczy tego, czy naukowiec rozumie angielski, bo on go nie rozumie. Ani czy mechanicznie („bezmyślnie”) postępował według algorytmu tłumaczenia, który inni przygotowali dla niego. [...] On *rozumie* Szekspira — intelektualna wspólnota uczonych poświadczyła to — i on rzeczywiście *coś niecoś* (*something*) pojmuje”<sup>19</sup>.

Według Rapaporta w odniesieniu do człowieka w Chińskim Pokoju można powiedzieć, że jednak coś niecoś rozumie, choć nie jest Chińczykiem. Albo inaczej — system AI rozumiejący język naturalny — może coś niecoś rozumieć<sup>20</sup>.

---

<sup>19</sup>Tamże, 114.

<sup>20</sup>Tamże. W niniejszym paragrafie chodziło jedynie o zaznaczenie istnienia i opis wariantu Chińskiego Pokoju w analizie historycznej dowodu Searle’a. Filozoficzne konse-

### 3.2 Chińska Sala Gimnastyczna

Według Searle'a dowód Chińskiego Pokoju odnosi się do każdej wersji mocnego programu sztucznej inteligencji, zarówno do systemów szeregowych jak i równoległych. Nie można naturalnie lekceważyć najnowszych osiągnięć w dziedzinie systemów konekcyjnych, ale nie można też przeceniać znaczenia tych badań. Niemniej jednak prowokują one istotne pytania dotyczące sposobu działania złożonych, równoległych systemów podobnych do struktur w mózgu.

Każde zadanie, jakie ma być wykonane na komputerze równoległym, może być wykonane na tradycyjnym komputerze szeregowym. Współbieżne systemy nie pozwalają więc na obejście eksperymentu myślowego.

Searle chciał jednak czarno na białym dowieść Churchlandom, że nie mają racji pokładając wielkie nadzieje w systemach konekcyjnych oraz w modelach sieci neuropodobnych; przedstawił więc odnoszący się do wspomnianych systemów wariant argumentu nazywając go Chińską Salą Gimnastyczną. W dowodzie tym występują analogiczne elementy, co w dowodzie oryginalnym w brzmieniu z 1990 roku. Całość eksperymentu odbywa się w wielkiej sali gimnastycznej zamiast w pokoju. Cały zespół ludzi, a nie jeden człowiek, ma do dyspozycji zbiór chińskich symboli (zestaw nr 1) oraz instrukcje po angielsku z zasadami manipulowania symbolami (książki). Do sali wprowadzane są pliki chińskich znaków (pytania — zestaw nr 2), które opracowuje wielki zespół ludzi, przez co cały system wymaga pewnego dostrojenia, zgrania. Osoby w sali przekazują sobie nawzajem symbole aż do ostatecznego opracowania zestawu nr 3 (odpowiedzi).

Cała hala ludzi jest więc zorganizowana jak sieć neuropodobna działająca przy zastosowaniu programów równoległych. Dowód został przeprowadzony z zastosowaniem terminów Chińskiego Pokoju. Ponieważ chodzi o eksperyment myślowy należy naturalnie pominąć dwa elementy związane z Chińską Salą Gimnastyczną: problem wielkości sali mogącej pomieścić wszystkich ludzi stanowiących model mózgu oraz czas, jaki byłby potrzebny do przepracowania zestawu nr 2 w zestaw nr 3. Z komputacyjnego punktu widzenia istnieje równoważność szeregowego i równoległego systemu — człowiek w Chińskim Pokoju czy ludzie w Chińskiej Sali Gimnastycznej wskutek manipulowania symbolami nic nie rozumieją. Wielkie rozmiary sali czy większy stopień złożoności systemu nie doprowadzą do interpretacji seman-

---

kwencji Koreańskiego Pokoju oraz porównanie elementów obu argumentów por. w dalszej części pracy, paragraf dotyczący semantyki i syntaktyki.

tycznej, nie stworzą umysłu. Dowód Chińskiej Sali Gimnastycznej zaprzecza mocnemu programowi sztucznej inteligencji również w odniesieniu do systemów konekcyjnych i sieci neuropodobnych; semantyka nie wynika również z równoległych obliczeń, ani z systemów wzorowanych na ludzkim mózgu.

## LOGICZNA STRUKTURA DOWODU CHIŃSKIEGO POKOJU

Zaczątki formułowania aksjomatów i wniosków występują już w streszczeniu artykułu z 1980 roku<sup>21</sup> i są powtórzone bez zmian w 1982 roku<sup>22</sup>. Jak to już zaznaczyliśmy, po raz pierwszy zostały one wyraźnie sformułowane w wykładach dla BBC w roku 1984<sup>23</sup>. Searle zastrzegał się wtedy, że twierdzenia te są być może jeszcze zbyt niedopracowane i wprowadzał do nich pewne zmiany; tak powstawały kolejne wersje Chińskiego Pokoju zawarte w wykładzie z Bombaju (1986) oraz w artykule ze styczniowego numeru „Scientific American” z 1990 roku<sup>24</sup>.

### 1. AKSJOMATY EKSPERYMENTU MYŚLOWEGO

Aksjomaty zawierają logiczną strukturę Chińskiego Pokoju; potraktowane łącznie stanowią koncepcję relacji między umysłem, mózgiem i komputerem. Searle podkreśla prostą strukturę logiczną dowodu, dzięki czemu można ocenić jego prawidłowość.

#### 1.1 Syntaktyka programów komputerowych

Zasadniczą sprawą w argumentacie było dla Searle’a udowodnienie następującej idei: jeśli człowiek zastępujący program nie rozumiał chińskiego, to nie może się to również w żaden sposób powieść komputerowi; pracuje on bowiem na zasadzie manipulowania symbolami będącymi składowymi programami, który jest czysto formalny, syntaktyczny. Autor Chińskiego Pokoju zawarł te myśli w pierwszym aksjomacie.

AKSJOMAT 1: *Programy komputerowe są sformalizowane (syntaktyczne)*

---

<sup>21</sup>J. R. Searle, *Minds, Brains, and Programs*, BBS 3 (1980), 417.

<sup>22</sup>J. R. Searle, *The Chinese Room Revisited*, BBS 5,2 (1982), 338. Por. też: tamże, 345.

<sup>23</sup>J. R. Searle, *Minds, Brains, and Science*, Harvard University Press, Cambridge 1984, 38–41.

<sup>24</sup>Aksjomaty i wnioski cytowane wg: J. R. Searle, *Is the Brain's Mind a Computer Program?*, „Scientific American”, January (1990), 26–31. Por. też: J. R. Searle, *Minds, Brains, and Science*, Harvard University Press, Cambridge 1984, 28–41 oraz J. R. Searle, *Turing The Chinese Room*, w: SoSaR, 295–301.

Powyższe stwierdzenie jest prawdziwe z definicji; wyraża to, co rozumiemy pod pojęciem programu komputerowego — jest to zespół operacji na symbolach. Ta cecha czyni komputer tak potężnym narzędziem — ten sam formalny program może być uruchomiony na różnych maszynach cyfrowych oraz na tej samej maszynie mogą być uruchamiane różne programy. Jak one powstają i w jaki sposób są realizowane? Przyjrzyjmy się nieco bliżej temu zagadnieniu.

Wewnętrzny język komputera zapisuje w pamięci maszyny dane wejściowe, wyniki przetwarzania danych oraz bieżąco realizowany program wynikowy; ten ostatni powstaje wskutek przekształcania programów źródłowych przez tzw. program kompilujący. Cały czas mamy do czynienia z abstrakcyjnym zapisem symbolicznym nie mającym żadnych istotnych własności fizycznych. Działania na symbolach nie mają też odniesienia do jakiegokolwiek znaczenia. Program źródłowy jest napisany w języku programowania, tj. w języku formalnym zapewniającym komunikowanie się człowieka z komputerem. Program ma określoną składnię — zbiór zasad określający rodzaje wyrażen, jakie mogą być wykorzystane w programie podczas operacji na symbolach, które mogą znaczyć cokolwiek. „W tym sensie program ma składnię, ale nie ma semantyki (warstwy znaczeniowej)”<sup>25</sup>. Na wszystkich poziomach powstawania i pracy programu działania odbywają się jedynie na symbolach. Interpretuje je natomiast programista lub użytkownik — oni nadają znaczenie symbolom zastosowanym w programie.

## 1.2 Semantyka ludzkiego umysłu

Inaczej niż w wypadku działania komputera wygląda sytuacja w odniesieniu do myślenia człowieka, o czym mówi aksjomat drugi.

AKSJOMAT 2: *Ludzki umysł zawiera intelektualny składnik (semantyczny)*

Umysł człowieka nie pracuje na zasadzie manipulowania symbolami. Nasze przekonania, pragnienia, rozumienie etc. mają zawartość sematyczną, punkt odniesienia do świata, lub skupiają się wokół czegoś w świecie — nie są więc niezinterpretowanymi symbolami.

---

<sup>25</sup>J. R. Searle, *Is the Brain's Mind a Computer Program?*, „Scientific American”, January (1990), 27.

„Ponieważ umysł pociąga za sobą język, pojawi się też składnia jako uzupełnienie semantyki, poza tym lingwistyczne rozumienie wymaga przynajmniej zrębu znaczeniowego”<sup>26</sup>.

### 1.3 Rozróżnienie semantyki i syntaktyki

Nie można otrzymać semantyki z syntaktyki — o tym mówi kolejny pewnik.

AKSJOMAT 3: *Składnia sama przez się ani nie kształtuje semantyki ani do niej nie wystarcza*

Rozróżnienie między syntaktyką a semantyką jest fundamentalną zasadą dla różnych dyscyplin. Aksjomat trzeci podkreśla rozróżnienie między tym, co czysto formalne i nie posiada żadnego znaczenia a tym, co posiada treść, określone znaczenie. Eksperyment myślowy Johna Searle’a przypomina tę ideę — człowiek zamknięty w pokoju ma do dyspozycji całą składnię a nie ma żadnej możliwości, by przyswoić sobie znaczenia, on ciągle nie rozumie chińskiego. Mało tego, nigdy się go nie nauczy w ten sposób, bowiem posiadanie mnóstwa symboli w połączeniu z zasadami manipulowania nimi nie da mu dostępu do znaczenia żadnego symbolu; *syntaksa-i-symboli* oraz *semantyka-i-znaczenie* łączą się ze sobą ale z samej syntaktyki nie wynika semantyka.

### 1.4 Mózg a umysł

Wiemy coraz więcej na temat funkcjonowania mózgu, ale zakres wiadomości jest ograniczony, zwłaszcza odnośnie do powstawania stanów psychicznych. Naukowcy zbadali niektóre tylko fenomeny umysłu; należą do nich m. in. pragnienie, ból i widzenie. Wnioski były zaskakujące — okazało się bowiem, że zjawiska umysłowe powstają wskutek specyficznych procesów neurobiologicznych. Przykładowo:

„Pragnienie, [...] wywoływane jest przez wyładowania pewnych typów neuronów w podwzgórzu, które z kolei wytwarzane są przez działanie konkretnego peptydu, angiotensyny II. Związki przyczynowe są typu „z dołu do góry” w tym sensie, że procesy neuronowe na niższym poziomie wywołują zjawiska psychiczne wyższego rzędu. W rzeczywistości, o ile wiemy, każde

---

<sup>26</sup>J. R. Searle, *Is the Brain's Mind a Computer Program?*, „Scientific American”, January (1990), 27.

zjawisko psychiczne [...] wywoływane jest przez specjalne wyłączenia nerwowe, zachodzące w konkretnych strukturach neuronalnych”<sup>27</sup>.

#### AKSJOMAT 4: *Mózg jest przyczyną umysłu*

jest podsumowaniem pewnych znanych faktów empirycznych odnośnie do funkcjonowania mózgu. Procesy neurofizjologiczne zachodzące w mózgu są powodem wszelkich stanów psychicznych i Searle myśl tę zawarł w powyższym pewniku.

Piąty aksjomat: *Umysł jest przyczyną zachowania* dodany był do pozostałych jeden jedyny raz — tylko w artykule z kongresu, który się odbył w 1986 roku<sup>28</sup>. System, który posiada odpowiednie stany umysłowe będzie zachowywał się w określony sposób, np. uczucie pragnienia powoduje, że szukamy napojów; jeśli ktoś zada nam ból — krzyczymy. Stany pragnienia lub bólu powstają w umyśle i one powodują odpowiednie zachowanie człowieka.

Analizując prace zawierające aksjomaty, zauważamy, że Searle zmieniał ich kolejność. Przyjmując numerację pewników 1–2–3–4 według wersji Chińskiego Pokoju z 1990 roku w poprzednich wyglądała ona następująco:

- w 1984: 4–3–1–2, czyli od końca w stosunku do 1990 roku z zamianą miejscami aksjomatów nr 1 i nr 2;
- w 1986: 1–3–2–4, tj. podobnie jak w 1990 z zamienieniem pewników nr 2 i nr 3.

O ile więc Searle w 1984 roku podał na początku konkluzję: *Mózg jest przyczyną umysłu*, o tyle w 1990 roku umieścił ją na końcu. Stało się tak zapewne wskutek systematyzowania logiki dowodzenia. Do tych konkluzji prowadzi cały tok myślenia — dlatego więc zapewne autor eksperymentu zmienił kolejność aksjomatów w ten sposób, by stanowiła ona zakończenie rozumowania.

## 2. WNIOSKI SEARLE'A NA TEMAT ARGUMENTU CHIŃSKIEGO POKOJU

Historia formułowania czterech głównych wniosków pokrywa się z historią aksjomatów; zaczątki znajdują się w streszczeniach artykułów z 1980 i 1982 roku, pierwsza wersja w wykładach w BBC w 1984 roku, następne

<sup>27</sup>J. R. Searle, *Czy intelekt mózgu jest programem komputerowym?*, „Świat Nauki”, Lipiec (1991), 13.

<sup>28</sup>J. R. Searle, *Turing The Chinese Room*, w: SoSaR, 299.

— z niewielkimi zmianami — w roku 1986 oraz 1990; podobnie jak w odniesieniu do aksjomatów wykład w Bombaju zawierał piąty wniosek (mózg przyczynuje zachowanie), który pomijany jest w następnych pracach. Poniżej zacytowałem wnioski w brzmieniu z 1990 roku.

## 2.1 Programy komputerowe a myślenie

Aksjomaty 1–3 przypominają, że program komputerowy jest czysto formalny, ludzkie myślenie zawiera składnik semantyczny a składnia nie jest ani konstytutywnym, ani wystarczającym elementem dla semantyki. Searle zaznacza, że szczególnie niezręcznie mu pouczyć o tym jego kolegów — uczonych z kręgu sztucznej inteligencji. Często, jak się wydaje, zapominają oni o tym, że komputer jest jedynie urządzeniem manipulującym symbolami i wyciągają błędne wnioski. Za mijającą się z prawdą uważa Searle koncepcję mocnego programu sztucznej inteligencji. Sam program nie jest zdolny do myślenia, jest jedynie narzędziem do manipulacji formalnymi symbolami, co nie wystarcza do uzyskania rozumienia<sup>29</sup>. Wniosek pierwszy, jaki Searle wyciąga z aksjomatów 1, 2 oraz 3, to — innymi słowy — odmienne brzmienie twierdzenia, że sAI jest błędnie sformułowana — umysłu nie można stworzyć jedynie przy pomocy programów komputerowych.

WNIOSEK 1: *Programy ani nie kształtują myślenia, ani do niego nie wystarczają*

W toku myślenia zawartym w powyższym wniosku oraz trzech pierwszych aksjomatach chodziło też o uwypuklenie następujących idei:

(1) Jeśli dany system jest myślący oraz może być symulowany przy pomocy manipulacji symbolami (a mózg może być naśladowany w ten sposób), to wcale z tego nie wynika, że myślenie jest równoważne formalnej manipulacji.

(2) Jedynym znanym obecnie tworem myślącym jest biologiczny mózg; nie oznacza to jednak, że inne systemy nie mogą myśleć. Być może takie znajdują się we Wszechświecie, bądź też w przyszłości będzie można stworzyć sztuczne systemy myślące, aktualnie jednak jest to niemożliwe.

(3) Chiński Pokój znajduje również zastosowanie w odniesieniu do systemów konekcyjnych<sup>30</sup>; współbieżne obliczenia, również oparte na programach formalnych, nie przyniosą żadnych właściwości intelektualnych. Jeśli więc pozostaniemy na poziomie komputera i programu formalnego nic nie

<sup>29</sup>J. R. Searle, *Turing The Chinese Room*, w: SoSaR, 297.

<sup>30</sup>Por. paragraf 3.2 na temat Chińskiej Sali Gimnastycznej, wariantu Chińskiego Pokoju.

da zwiększanie: pojemności pamięci, szybkości obliczeń i dostępu do danych oraz coraz to większe skomplikowanie programów komputerowych, z systemami konekcyjnymi włącznie.

(4) Test Turinga jest niewystarczającym kryterium do odkrywania obecności treści umysłowych; system może przejść poprawnie TT i nie mieć żadnych elementów związanych z intelektem.

## 2.2 Siła twórcza artefaktów i mózgu

Kolejny wniosek wypływa z czwartego aksjomatu: *Mózg jest przyczyną umysłu.*

WNIOSEK 2: *Dowolny inny system zdolny do spowodowania umysłu musiałby mieć siłę przyczynową (przynajmniej) równoważną mózgowi.*

W oparciu o zasadniczą myśl tego wniosku rozważmy najpierw przykład systemu mechanicznego. Jeśli w wypadku silnika gazowego samochód miałby poruszać się z taką samą prędkością jak z silnikiem Diesla, musiałby w obu przypadkach mieć taką samą moc mechaniczną. W odniesieniu do myślenia zagadnienie nie jest jednak tak proste, chodzi bowiem o procesy chemiczno-biologiczne. Nie oznacza to jednak, że mogą myśleć jedynie systemy biologiczne; drugi wniosek nic nie mówi na temat mechanizmów, ale o tym, że dowolny inny system musi mieć siłę przyczynową równoważną mózgowi. Teoretycznie więc jest możliwe, że będzie myślał system niebiologiczny i będzie on miał siłę sprawczą równoważną ludzkiemu intelektowi. System taki mógłby być przyczyną procesów myślowych przy zastosowaniu zupełnie innych elementów biologicznych czy chemicznych. Tak więc, gdyby myślący Marsjanie przybyli na naszą planetę mogłoby się okazać po otwarciu ich głowy, że zamiast mózgu mają zielony szlam a jednak mają stany umysłowe<sup>31</sup>. Searle uważa jednak za zerowe prawdopodobieństwo, że np. krzemowe *chipy* czy lampy próżniowe mogą mieć tę samą siłę przyczynową co neurony. Szanse równoważności chemicznych właściwości neuronów i układów elektronicznych są takie, jak szanse przeprowadzenia przez krzemowe *chipy* specyficznie biologicznych procesów, jak np. trawienia, fotosyntezy czy laktacji; trzeba jednak podkreślić, że nie zostało to jednak udowodnione przez Chiński Pokój, lecz jest jedynie twierdzeniem empirycznym, choć trudnym do podważenia.

---

<sup>31</sup>J. R. Searle, *Minds, Brains, and Science*, Harvard University Press, Cambridge 1984, 40.

### 2.3 Czy formalne programy wystarczają do wytworzenia siły przyczynowej mózgu?

Odpowiedź na pytanie dotyczące wystarczalności powielenia siły przyczynowej mózgu przez programy formalne jest negatywna. Wynika ona z wniosku trzeciego, wyciągniętego z dwóch pierwszych.

WNIOSEK 3: *Dowolny artefakt wytwarzający zjawiska umysłowe, dowolny sztuczny mózg, który miałby powielić specyficzną siłę przyczynową mózgu nie mógłby tego osiągnąć jedynie przez wykonywanie formalnego programu.*

Ponownie przewija się tutaj myśl związana z niemożnością realizacji mocnego programu AI. Searle uważa natomiast za właściwe założenia złagodzonego programu sztucznej inteligencji — komputer w studiach nad umysłem jest potężnym i skutecznym *narzędziem*. Dzięki niemu o wiele bardziej precyzyjnie można formułować i testować hipotezy na temat ludzkiego intelektu. Autor dowodu Chińskiego Pokoju nie ma zastrzeżeń<sup>32</sup> co do takiego kierunku badań nad AI — komputerowe *modele* są bardzo dobrymi narzędziami w badaniu umysłu; zwłaszcza sieci neuropodobne pozwalają dokładniej symulować procesy mózgu.

### 2.4 Jak mózg wytwarza zjawiska umysłowe?

Wynikiem rozumowania zawartego w pierwszym aksjomacie i pierwszym wniosku jest

WNIOSEK 4: *Sposób, w jaki mózg ludzki wytwarza zjawiska umysłowe, nie może polegać jedynie na wykonywaniu programu komputerowego.*

Gdyby przyjąć stanowisko przeciwne, byłoby ono głęboko antybiologiczne, a nawet więcej — przeczyłoby temu, co nauka wie dziś na temat pracy mózgu. Taki właśnie przeciwny pogląd uznaje sAI — w ludzkim umyśle nie ma nic biologicznego, nie jest on częścią świata natury, ludzki intelekt jest czysto formalny, nie może być analizowany jako konkretny wynik procesów biologicznych<sup>33</sup>. Podsumujmy powyższe myśli słowami z drugiego wykładu z radia BBC:

„Rezultatem tej dyskusji, jest przypomnienie nam wszystkim czegoś, o czym dawno wiedzieliśmy: stany umysłowe są fenome-

---

<sup>32</sup>J. R. Searle, *Minds, Brains, and Programs*, BBS 3 (1980), 417.

<sup>33</sup>Por. J. R. Searle, *Minds, Brains, and Science*, Harvard University Press, Cambridge 1984, 28.

nami biologicznymi. Świadomość, intencjonalność, subiektywność i umysłowa przyczynowość — to wszystko jest częścią historii naszego biologicznego życia wraz z różnicowaniem, rozmnażaniem się, wydzielaniem żółci i trawieniem”<sup>34</sup>.

Wspomnijmy na koniec piąty wniosek, który występuje tylko raz w piśmiennictwie Searle’a — w wystąpieniu na kongresie w Bombaju (1986). Wniosek ten: *Mózg jest przyczyną zachowania*<sup>35</sup> jest wyciągnięty na podstawie prawa przechodniości z aksjomatu czwartego — *Mózg jest przyczyną umysłu* oraz piątego (również pojawiającego się jedynie w tym artykule) — *Umysł jest przyczyną zachowania*.

Poprzez specyficzne biochemiczne właściwości mózgu są powodowane wszystkie umysłowe zjawiska łącznie z fenomenami, które pośredniczą w relacjach wejście–wyjście; fenomeny te są przyczyną zachowania. W wypadku bólu i okrzyku takie właśnie jest następstwo faktów: najpierw ma miejsce odczuwanie bólu (wyładowania neuronowe) we wzgórzu mózgu a następnie w korze somatomotorycznej, w niej powstaje bodziec powodujący okrzyk.

W wypadku wniosku piątego chodziło Searle’owi o rozróżnienie między siłą przyczynową mózgu do powodowania stanów umysłowych i siłą przyczynową do produkowania relacji wejścia i wyjścia<sup>36</sup>.

### 3. ŹRÓDŁA INSPIRACJI

Po zapoznaniu się z dowodem Chińskiego Pokoju, jego historią i wariantami warto pokusić się o odszukanie źródeł inspiracji w odniesieniu do eksperymentu myślowego. Jak się wydaje, istnieją przynajmniej dwa ślady, które zapewne naprowadziły Searle’a na pomysł argumentu — są nimi *Test Turinga* oraz *Wielki Mózg Chin* Neda Blocka. Do takiego wniosku upoważnia analiza porównawcza Chińskiego Pokoju z dwoma pozostałymi eksperymentami; można wyciągnąć interesujące wnioski z podobieństw, różnic oraz założeń i idei filozoficznych, jakie przyświecały ich autorom.

#### 3.1 Test Turinga

Gdy Alan M. Turing postawił pytanie: „Czy maszyna może myśleć?”, zamiast definiować pojęcia „maszyna” oraz „myśleć” zaproponował opisanie problemu przy pomocy eksperymentu myślowego — *gry w naśladow-*

<sup>34</sup>Tamże, 41.

<sup>35</sup>J. R. Searle, *Turing The Chinese Room*, w: SoSaR, 299.

<sup>36</sup>J. R. Searle, *Turing The Chinese Room*, w: SoSaR, 299.

nictwo<sup>37</sup>. W pokoju nr 1 znajduje się dwoje ludzi — mężczyzna (A) i kobieta (B); odpowiadają oni na pytania człowieka (C) z drugiego, oddzielnego pokoju nr 2. Dalekopis jest środkiem porozumiewania się a zadaniem C jest rozpoznanie płci odpowiadających. Jakie będą wyniki w tej części gry, jak często człowiek C będzie decydował poprawnie i błędnie? Zastąpmy teraz mężczyznę A komputerem cyfrowym. Czy będzie on z powodzeniem naśladował żywego człowieka? Czy z taką samą częstotliwością C będzie decydował błędnie względnie zgodnie ze stanem faktycznym? Czy są możliwe komputery cyfrowe, które poprawnie pracowałyby w grze w naśladownictwo? Oto opis i główne kwestie związane z Testem Turinga.

Tabela nr 2 jest zestawieniem elementów dowodu Chińskiego Pokoju i Testu Turinga. Obydwa dotyczą zagadnień: człowiek — język — rozumienie — komputer cyfrowy. Już na pierwszy rzut oka widać znaczne podobieństwo; zresztą sam Searle zatytułował wykład na kongresie w Bombaju „Chiński Pokój Turinga”. W obu testach występują identyczne elementy (pokoje, ludzie, pytania, odpowiedzi, książki z zasadami) lub zbliżone do siebie (magazyn informacji — baza danych). Są jednak i zasadnicze różnice — w Chińskim Pokoju odpowiadający nie zna i nie rozumie w ogóle języka dialogu a nawet nie ma możliwości, by się go kiedykolwiek nauczyć; jest to dla niego jedynie zestaw symboli bez znaczenia; w TT ludzie w pokoju nr 1 rozumieją język, co więcej, rozumie go również komputer (*sic!*). Druga istotna różnica — w Chińskim Pokoju człowiek zastępuje komputer a w TT komputer — człowieka. Gdyby te dwie różnice zniwelować, obydwie testy byłyby niezmiernie podobne<sup>38</sup>. Turing starał się wykazać, że jeśli nikt nie potrafi odróżnić działania komputera od człowieka, to znaczy że komputer posiada ludzkie umiejętności, np. odpowiadania na pytania związane z daną historią. Searle natomiast dowodził, że komputer nigdy nie będzie mógł rozumieć. Test Turinga, po fali entuzjazmu, jest dziś uznawany za niewystarczający, zaś Chiński Pokój jest akceptowany, choć ma również żarty przeciwników. Dwie wyżej wymienione istotne różnice w stosunku do TT zadecydowały o mocnej pozycji eksperymentu Searle’a we współczesnej filozofii sztucznej inteligencji.

---

<sup>37</sup>A. M. Turing, *Computing Machinery and Intelligence*; „Mind” LIX, 2236 (Oct. 1950), 433–460. Korzystam z przedruku w: *The Philosophy of Artificial Intelligence*, ed. Margaret A. Boden, Oxford University Press, New York 1990, 40–66. W opisie Turinga nie ma cyfrowych oznaczeń pokoi, wprowadziliśmy je dla łatwiejszego rozróżniania poszczególnych elementów testu.

<sup>38</sup>Pomijając naturalnie mniej istotne różnice typu: dwa pokoje zamiast jednego; dwoje ludzi a nie jeden człowiek itp.

### 3.2 *Wielki Mózg Chin Neda Blocka*

Analizując historię dowodu Johna R. Searle'a, rozważmy również inny eksperyment myślowy, który można uznać za kolejnego, obok testu Turinga, protoplastę Chińskiego Pokoju.

Na początku lat sześćdziesiątych Hilary Putnam<sup>39</sup> zapoczątkował pogląd w filozofii umysłu, który bazował na analogii między umysłem a komputerem cyfrowym; nazywa ten pogląd *funkcjonalizmem* lub *komputacyjną wizją (view) umysłu*. Zgodnie z tą ideą maszyna, człowiek, czy cokolwiek, co jest zbudowane z krzemowych tranzystorów, pracują na tej samej zasadzie, jeśli opisać je na odpowiednim poziomie abstrakcji. Co więcej, według Putnama błędne jest twierdzenie, że mózg jest czymś istotnym dla naszego umysłu. Według zasad funkcjonalizmu wszystko, co posiada umysł, można opisać jak maszynę Turinga — dany stan umysłu jest identyczny z jednym ze stanów tablicy adresowej MT.

Autorem eksperymentu myślowego wymierzonego przeciw funkcjonalizmowi jest Ned Block, wykładowca MIT na Wydziale Lingwistyki i Filozofii. Wyobraźmy sobie sztuczne ciało, ale takie jak ludzkie, w którego głowie zamieszkują ludziki (*homunculi*). Neurony organów czuciowych połączone są z zestawem świateł, neurony wyjść motorycznych podpięte są do zespołu przycisków. Inteligencja ludzików jest niska ale wystarcza do wykonania powierzonego im zadania. Polega ono na odpowiednim adresowaniu impulsów pochodzących z wejścia i na zmienianiu tzw. karty stanu na tablicy ogłoszeń; karta zawiera symbol oznaczający dany stan wyszczególniony na tablicy adresów. Kolejność czynności ludzika wygląda następująco:

„Jeśli na tablicy jest karta ‘G’ i zapali się światło wejścia I17, naciska on guzik wyjścia O191 i zmienia stan karty na ‘M’”<sup>40</sup>.

Wyobraźmy sobie teraz symulację pracy mózgu człowieka. Każdy neuron będzie zastępowany przez jednego człowieka, potrzeba więc ok. 1 miliarda ludzi. Załóżmy, że rząd chiński został przekonany do założeń funkcjonalizmu i zgodził się, by miliard Chińczyków symulowało przez godzinę mózg ludzki. Każdy z nich

<sup>39</sup>Por. jego: *Brains and Behavior*, w: *Analytical Philosophy*, second series, ed. Ronald Butler, Oxford 1965, 1–20; *The Mental Life of Some Machines*, w: *Intentionality, Minds and Perception*, ed. Hector-Neri Castaeda, Detroit, 1967, 177–200; *Minds and Machines*, w: *Dimensions of Mind*, ed. Sidney Hook, New York 1960, 148–179.

<sup>40</sup>N. Block, *Troubles with Functionalism*, „Minnesota Studies in the Philosophy of Science”, vol. IX (1978), 278.

„został wyposażony w urządzenie nadawczo-odbiorcze umożliwiające mu łączenie się [...] z innymi osobami i sztucznym ciałem wspomnianym w poprzednim przykładzie. Zamieniamy ludziki w głowie sztucznego ciała na radiowe odbiorniki i nadajniki połączone z wejściami i wyjściami neuronów. Zamiast na tablicy ogłoszeń doprowadzamy do wyświetlania liter na szeregu satelitach umiejscowionych w ten sposób, by były widziane z każdego miejsca w Chinach”<sup>41</sup>.

Zgodnie z zasadami funkcjonalizmu Chińczycy i sztuczne ciało byłoby metafizycznym modelem pracy mózgu człowieka. Zachodzi pytanie: czy system ten ma w ogóle jakieś stany umysłowe? Funkcjonalizm maszynowy mówi, że każdy stan umysłowy jest identyczny ze stanem komputera, np. stan  $U$  jest identyczny ze stanem  $K_u$ . Jak się wydaje, Wielki Mózg Chin, bo tak nazwany jest ów eksperyment, przeczy temu — neguje zdanie  $U = K_u$ , a co za tym idzie, przeczy zasadom funkcjonalizmu<sup>42</sup>.

Hilary Putnam zaczął się wycofywać ze stanowiska funkcjonalizmu podczas wykładów roku akademickiego 1982/83, które zaowocowały książką *Representation and Reality*<sup>43</sup>; ostatecznie w 1992 roku rozstał się z wiodącymi ideami funkcjonalizmu w artykule pod znamienym tytułem: *Why Functionalism Didn't Work?*

W tabeli nr 2 zestawione są elementy wchodzące w skład Chińskiego Pokoju i Wielkiego Mózgu Chin. Obydwa eksperymenty koncentrują się wokół problemu wykazania stanów umysłowych w programie komputerowym. Z ich porównania wynika szereg mniej ważnych rozbieżności co do nazewnictwa; można do nich zaliczyć następujące pary: pokój (sala gimnastyczna) — głowa (teren Chin), baza danych — karty stanu, książka z zasadami — zadania do wykonania przez Chińczyków. Są też i istotne różnice: w Chińskim Pokoju pytania są stawiane przez ludzi, a odpowiedzi dawane przez człowieka, natomiast w eksperymencie wykładowcy MIT elementem odpowiadającym ludziom są sygnały docierające z zewnątrz do sztucznego ciała poprzez organy zmysłu a pytaniom i odpowiedziom przyporządkować należy impulsy neuronowe na wejściu i wyjściu.

<sup>41</sup>Tamże, 279.

<sup>42</sup>Pomijamy tu zarówno zasadność argumentu Neda Blocka, jak i funkcjonalizmu; obydwie zagadnienia stanowią temat do odrębnej, obszernej pracy. W niniejszym paragrafie chodzi jedynie o poszukiwanie źródeł zainspirowania Searle'a do utworzenia dowodu Chińskiego Pokoju.

<sup>43</sup>MIT Press, Cambridge 1988.

Zbieżność odniesienia do kultury chińskiej jest przypadkowa; Block wybrał Chiny, gdyż mają więcej niż miliard mieszkańców, a właśnie miliard ludzi potrzebnych byłoby do modelu ludzkiego mózgu; Searle zaś nie znał w ogóle języka chińskiego i dlatego ten właśnie wybrał do swojego eksperymentu.

Między obydwojema eksperymentami są jednak i doniosłe podobieństwa. Jak już wyżej wspomnieliśmy, oba dowody koncentrują się wokół takiego samego problemu; jest nim stwierdzenie obecności myślenia, rozumienia lub bardziej ogólnie — stanów umysłowych w maszynach. W obu wypadkach odpowiedź jest negatywna i zarówno w jednym, jak i drugim eksperymencie chodzi o testowanie teorii na podobnej zasadzie — poprzez metaforyczne modele. W Chińskiej Sali Ginnastycznej (wariant Chińskiego Pokoju) ludzie zastępują komputer i manipulują symbolami nie znając ich znaczenia, w Wielkim Mózgu Chin miliard osób; zastępuje maszynę i przekazuje niezinterpretowane sygnały. Działanie Chińczyków odzwierciedla pracę neuronów, ludzie z dowodu Searle'a pracują jak system konekcyjny wzorowany na sieciach neuronowych.

Gdyby więc, pomijając mniej istotne różnice, w eksperymencie Block'a sygnały zamienić na ludzi, impulsy neuronowe na wejściu i wyjściu na pytania i odpowiedzi — otrzymalibyśmy wersję Chińskiego Pokoju w postaci Chińskiej Sali Ginnastycznej.

### 3.3 Wnioski

O ile w wypadku testu Turinga i Chińskiego Pokoju uderza podobieństwo elementów, o tyle w odniesieniu do eksperymentów Blocka i Searle'a zwraca na siebie uwagę przede wszystkim pokrewieństwo idei i wniosków. Tytuł wykładu z Bombaju *Turingowski Chiński Pokój* jest wyraźnym nawiązaniem przez samego jego autora do myśli słynnego angielskiego matematyka i logika<sup>44</sup>. Wielki Mózg Chin jest postrzegany jako protoplasta, prekursor dowodu Searle'a. Twórca Chińskiego Pokoju konkretniej sprecyzował jego logiczną strukturę; pytania, odpowiedzi oraz sam pomysł programu „rozumiejącego” język chiński koncentrują się wokół rozumienia, a nie zaś jedynie przejmowania bliżej nieokreślonych impulsów neuronowych na wejściu i przekazywania ich, po przetworzeniu, na wyjście.

---

<sup>44</sup>J. R. Searle, *Turing The Chinese Room*, w: SoSaR, 295–301; por. też: W. E. Smythe, *Simulation Games*, BBS 3 (1980), 448 — autor artykułu wprost nazywa Chiński Pokój rodzajem Testu Turinga.

Można więc powiedzieć z wielkim prawdopodobieństwem, że Test Turinga i Wielki Mózg Chin były źródłem inspiracji dla autora artykułu *Minds, Brains and Programs*. TT schodzi z forum filozofii sztucznej inteligencji i jest już raczej tylko etapem historii badań nad AI; o eksperymencie Block'a pisze się niewiele. Natomiast Chiński Pokój jest uznawany za klasyczny dowód w filozofii sztucznej inteligencji i pokrewnych dziedzinach. Nowe, istotne elementy wprowadzone przez Searle'a do jego argumentu, różnice w stosunku do obydwu testów oraz logika wywodu wpłynęły z biegiem czasu na coraz bardziej przychylny stosunek filozofów i naukowców do Chińskiego Pokoju.

Struktura eksperymentu zawarta w aksjomatach jest bardzo jasna i prosta. Z 27 autorów artykułów polemicznych z 1980 roku oraz z 7 autorów z 1982 roku *żaden (sic!)* nie zdołał wykazać błędnej logiki czy uproszczeń w argumentacie. W literaturze na temat eksperymentu Searle'a nie można dostrzec ani jednego krytyka, który by stawiał czoła logicznej strukturze Chińskiego Pokoju<sup>45</sup>; komentatorzy podawali różne pomysły, ale żaden z nich nie zachwiał. Najwięcej zastrzeżeń wzbudził trzeci aksjomat i podejmowane były próby wyprowadzania semantyki z samej tylko syntaktyki, by zaprzeczyć temu, że składnia sama przez się ani nie kształtuje semantyki, ani do niej nie wystarcza. Próby takie trwają nadal; jak dotąd — są nieskuteczne.

TABELA NR 1: EWOLUCJA ELEMENTÓW CHIŃSKIEGO POKOJU

1980, 1982 BBS	(1983) 1984 Dyskusja panelowa	1984 BBC	1986 Bombaj	1990 Scientific American
zbiór chińskich symboli tekst ( <i>script</i> ) ( <i>data base</i> ) (zestaw nr 1)	zbiór chińskich symboli baza danych ( <i>data base</i> ) (zestaw nr 1)	Pzbiór chińskich symboli baza danych ( <i>data base</i> ) (zestaw nr 1)	chińskie symbole ( <i>zestaw nr 1</i> )	zbiór chińskich symboli baza danych ( <i>data base</i> ) (zestaw nr 1)
zbiór chińskich symboli opowiadanie ( <i>story</i> ) (zestaw nr 2)				
pytania (zestaw nr 3)	pytania (zestaw nr 2)	pytania (zestaw nr 2)	chińskie symbole na wejściu pytania (zestaw nr 2)	plik symboli — pytania (zestaw nr 2)

<sup>45</sup>Por. J. R. Searle, *Turing The Chinese Room*, w: SoSaR, 301.

odpowiedzi na pytania (zestaw nr 4)	odpowiedzi (zestaw nr 3)	odpowiedzi na pytania (zestaw nr 3)	chińskie symbole na wyjściu (zestaw nr 3)	plik symboli — odpowiedzi (zestaw nr 3)
zestaw zasad (po angielsku) do wajemnego łączenia zbioru symboli nr 2 i nr 1 (plik A programu)	wszelkie książki z zasadami do mieszania chińskich symboli (poszczególne pliki programu)	książka zasad (po angielsku) do manipulowania symbolami (programu)	zasady dobierania jednych symboli do drugich, w języku który zna człowiek w pokoju (program)	książka (po anelsku) zasad dobierania jednych symboli do drugich (program komputerowy)
wstępne zdefiniowanie czwartej tezy (późniejszego aksjomatu) w streszczeniu artykułu		4 tezy ( <i>thesis</i> ) w kolejności 4-3-1-2 (w stosunku do 1990)	4 propozycje ( <i>propositions</i> ) w kolejności 1-3-2-4 (w stosunku do 1990) dodatkowo — piąta propozycja	4 aksjomaty ( <i>axioms</i> ) w kolejności 1-2-3-4
wstępne sformułowania wniosków w streszczeniu artykułu		4 wnioski w kolejności 1-4-2-3 (w stosunku do 1990)	4 wnioski w kolejności 1-2-4-3 (w stosunku do 1990) dodatkowo — piąty wniosek	4 wnioski w kolejności 1-2-3-4
wersja dowodu w odniesieniu do programu Schanka (odpowiadającego na pytania dotyczące opowiadania)	zmiana: zredukowanie zbiorów znaków chińskich z czterech do trzech	zmiany: jeden plik programu, po raz pierwszy wyraźnie sformułowane tezy i wnioski	zmiany: dodane — piąta propozycja i piąty wniosek	ogólna wersja dowodu w odniesieniu do formalnego programu komputerowego uruchomionego na dowolnym komputerze cyfrowym

TABELA NR 2: ZESTAWIENIE ELEMENTÓW CHIŃSKIEGO POKOJU, TESTU TURINGA ORAZ WIELKIEGO MÓZGU CHIN

TEST TURINGA (1950)	CHIŃSKI POKÓJ (1980) 1990*	WIELKI MÓZG CHIN (1978)
pokój nr 1 — dla odpowiadających: mężczyzny (A) i kobiety (B); pokój nr 2 — dla człowieka (C) zadającego pytania	pokój — przebywa w nim człowiek dający odpowiedzi	głowa sztucznego ciała; teren Chin
magazyn informacji	baza danych — zestaw chińskich symboli nr 1	wszystkie karty stanu
mężczyzna (A) a następnie zastępujący go komputer cyfrowy — odpowiada na pytania, rozumie po angielsku ; kobieta (B) — odpowiada na pytania, rozumie po angielsku	człowiek, który zastępuje komputer odpowiada na pytania, nie rozumie po chińsku	ludziki w głowie, wykonujące proste operacje; w czasie symulowania mózgu ludzkiego — zastąpione nadajnikami i odbiornikami radiowymi; 1 mld Chińczyków zastępujących neurony komunikujących się ze sobą i ze sztucznym ciałem przy pomocy urządzeń nadawczo-odbiorczych
człowiek (C) — zadaje pytania, rozumie po angielsku	ludzie na zewnątrz pokoju — zadają pytania; rozumieją po chińsku	sygnały z zewnątrz sztucznego ciała docierające poprzez organy zmysłu
pytania zadawane po angielsku, przekazywane delekopisem	pytania — zestaw chińskich symboli nr 2	neuronowy impuls na wejściu docierający za pośrednictwem organów zmysłu
odpowiedzi — dawane po angielsku, przekazywane delekopisem	odpowiedzi — zestaw chińskich symboli nr 3	impuls na wyjściu neuronów motorycznych
książka z zasadami (domyślnie: napisana po angielsku, program)	książka z zasadami — napisana po angielsku (program)	zadania do wykonania przez wszystkich ludzi i Chińczyków
(domyślnie: autorzy książki — programiści)	autorzy książki — programiści	(domyślnie: autor zadań — programista)

problem ogólny: czy jest możliwe posiadanie przez komputer cyfrowy umiejętności ludzkich, np. dodawania, rozumienia, uczenia się itp.; odpowiedź Turinga: TAK	problem ogólny: czy komputer cyfrowy mógłby myśleć dzięki zaimplementowaniu w nim programu komputerowego? odpowiedź Searle'a: NIE	problem ogólny: czy opisany system (Wielki Mózg Chin) ma jakiegokolwiek stany umysłowe? odpowiedź Block'a: NIE
---	---	--

\* — 1980: pierwsza wersja opisu Chińskiego Pokoju, 1990: ostateczna, najpełniejsza wersja