

Józef KLOCH

CHIŃSKI POKÓJ. EKSPERYMENT MYŚLOWY
JOHNA SEARLE'A. STUDIUM HISTORYCZNO –
FILOZOFICZNE

(cz. 2)

Argument Chińskiego Pokoju wydawał się Searle'owi tak jasny, że nie spodziewał się żadnych zastrzeżeń; był przygotowany na zaakceptowanie jego idei przez kręgi uczonych. Zanim opublikował swoją pracę przedstawił ją środowiskom uniwersyteckim Berkeley, Yale, Stanford oraz Massachusetts Institute of Technology (MIT). Z kolei amerykańskie czasopismo *The Behavioral and Brain Sciences*, w którym miał się ukazać artykuł, ma zwyczaj rozsyłania nieopublikowanej jeszcze pracy do wybranych uczonych zajmujących się danym zagadnieniem; ich komentarze drukowane są łącznie z kluczowym artykułem.

Okazało się, że eksperyment myślowy Searle'a stał się bodźcem do długiej i burzliwej dyskusji. Jej początkiem było 27 komentarzy¹, w większości negatywnych. W dwa lata później 7 kolejnych prac na temat Chińskiego Pokoju zostało opublikowanych na łamach tego samego czasopisma². Obie serie stanowią najobszerniejszą dyskusję nad argumentem. Naturalnie nie były to jedyne artykuły — po 1982 roku spory nie wygasły i trwają do dziś; raz podjęty temat na stałe zagościł w dyskusjach filozofów nauki, psychologów, naukowców z kręgu sztucznej inteligencji, lingwistów, neurologów itp.

Poniżej przedstawiliśmy najbardziej reprezentatywne komentarze i najczęściej powtarzające się idee w odniesieniu do eksperymentu myślowego. Można je podzielić na następujące grupy: odpowiedź systemowa,

*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.

¹Patrz: *The Behavioral and Brain Sciences*, Cambridge University Press, BBS 3 (1980), 424–450; repliki Searle'a do komentarzy: *Tamże*, 450–457. W dalszej części artykułu w odniesieniu tego periodyku stosuję skrót BBS.

²Patrz: BBS 5, 2 (1982), 338–345; odpowiedzi Searle'a: *Tamże*, 345–348.

semantyka a robot i komputer, symulator mózgu, propozycja kombinacyjna oraz przyszłe rozwiązania. Searle ustosunkowując się do komentarzy jeszcze lepiej wyjaśnia istotę argumentu, wprowadza do niego pewne zmiany, tworzy jego wariant — Chińską Salę Gimnastyczną, szerzej omawia szereg zagadnień, jak np. rolę intencjonalności, jego poglądy na temat mocnego programu sztucznej inteligencji, koncepcję umysłu, wzajemne odniesienie syntaktyki do semantyki etc.

Niniejszy artykuł przedstawia historię pierwszej fazy (1980–1982) bardzo ostrej dyskusji nad eksperymentem myślowym oraz wprowadza w zagadnienia filozoficzne³. W latach 60–tych i 70–tych uczeni rozwijali koncepcję komputacyjnego pojmowania procesów, jakie odbywają się w mózgu człowieka; komentarze z 1980 i 1982 roku nawiązują do tych idei. Argument Searle’a przeczy takiej wizji — między działaniem programu komputerowego a intelektualnym funkcjonowaniem biologicznego mózgu zachodzi głęboka różnica.

1. PROBLEM HOLISTYCZNEGO ROZUMIENIA SYSTEMU

Człowiek w Chińskim Pokoju stanowi tylko jeden z elementów całego systemu, do którego należą m.in. symbole chińskie, książka z zasadami, pytania, odpowiedzi etc. Co prawda człowiek jako część systemu nie ma zdolności rozumienia, ale posiada ją cały system. Trzeba wziąć pod uwagę wszystkie elementy razem i wtedy ma miejsce rozumienie — tak brzmi główna idea najczęstszej repliki zwanej odpowiedzią systemową⁴.

Zwolennikami tego poglądu są m.in. Drew McDermott, John McCarthy i Robert Wilensky oraz, w odniesieniu do analogii tzw. *Jasnego Pokoju*, wykładowcy filozofii z San Diego — Paul Churchland i Patricia Smith Churchland.

1.1 Niewystarczalność argumentacji

Autor Chińskiego Pokoju poddaje krytyce tok rozumowania zwolenników wyżej wymienionego poglądu — po pierwsze brak pozytywnego argumentu, że cały system rzeczywiście rozumie, po drugie gdyby zaakceptować odpowiedź systemową należałoby wyciągnąć absurdalne wnioski.

Popierający odpowiedź systemową poza testem Turinga, nie podają żadnego argumentu, który skłaniałby do przyjęcia stanowiska: cały układ

³Ponieważ analogia tzw. *Jasnego Pokoju* z 1990 roku jest wersją odpowiedzi systemowej — została włączona do tej części pracy.

⁴Reakcja na te poglądy — por. J.R. Searle, *Minds, Brains, and Programs*, BBS 3 (1980), 419–420; Tenże, *Intrinsic Intentionality*, BBS 3 (1980), 453; Tenże, *The Chinese Room Revisited*, BBS 5, 2 (1982), 346.

rozumie. System co prawda potrafi dać pozytywną odpowiedź na pytanie czy rozumie po chińsku⁵, ale to nie daje wiarygodności twierdzeniu, że ma tutaj miejsce jakiegokolwiek rozumienie tego języka. Odwoływanie się do metody testu Turinga jest zakładaniem, że to, co jest dyskutowane zostało udowodnione. Co więcej — można wykazać brak rozumienia w całym systemie przez zinternalizowanie wszystkich elementów w głowie człowieka.

Wyobraźmy sobie, że człowiek z Chińskiego Pokoju opuścił go, nauczył się na pamięć kształtów wszystkich symboli oraz zasad z książki, wszelkie operacje komputacyjne przeprowadza w głowie i działa na otwartej przestrzeni. Kompletny system jest więc wewnątrz człowieka. Można nawet powiedzieć, że działają w nim dwa podsystemy: angielski — i ten posiada zdolność rozumienia — oraz chiński manipulujący znakami. Ten ostatni jest nawet w tym wypadku częścią angielskiego podsystemu i nadal nie przyporządkowuje żadnej treści semantycznej do formalnych znaków⁶. W subsystemie angielskim słowo *a tree* oznacza drzewo a w chińskim symbol w kształcie krzaczka nic nie oznacza — nadal wszystkie znaki są pozbawione znaczenia.

Przedstawmy w inny sposób ideę systemowej odpowiedzi — cały system myśli, jeśli jest w nim poprawny program i dane na wejściu oraz wyjściu. Gdyby na serio potraktować to twierdzenie można by dojść do absurdalnego wniosku — wszystkie rodzaje niekognitywnych systemów (np. trawienie) musiałyby być zaliczone do rozumnych.

Drew McDermott⁷ w 1982 roku zmienił nieco odpowiedź systemową — wprowadził jeszcze jedną osobę do Chińskiego Pokoju, Wangę rozumiejącego język chiński; chciał przez ten zabieg wprowadzić intencjonalność do całego układu. Nie przedstawił jednak adekwatnej motywacji co do obecności drugiej osoby i nie dodał żadnego nowego argumentu do tej grupy komentarzy⁸.

Niewystarczalność argumentacji w rozumowaniu zwolenników odpowiedzi systemowej skłania do jej odrzucenia.

⁵R. Wilensky, *Computers, Cognition and Philosophy*, BBS 3 (1980), 449–450; por. też: J. McCarthy, *Beliefs, Machines and Theories*, BBS 3 (1980), 435.

⁶J.R. Searle, *Minds, Brains, and Programs*, 453.

⁷*Minds, Brains, Programs and Persons*, BBS 5, 2 (1982), 339–341.

⁸J.R. Searle, *The Chinese Room Revisited*, 346.

1.2 Wariant odpowiedzi systemowej — Jasny Pokój

Analogia Churchlandów⁹ jest bardzo sugestywną i przy tym najnowszą repliką wymierzoną przeciw trzeciemu aksjomatowi omawianego eksperymentu myślowego; określić ją można jako wariant odpowiedzi systemowej. Filozofowie z San Diego odrzucają pewnik Searle'a: *Składnia sama przez się ani nie kształtuje semantyki ani do niej nie wystarcza*; uważają go za bardzo splotony i sądzą, że cały Chiński Pokój został specjalnie skonstruowany po to, by obudować wyżej wymieniony aksjomat. Podają też — przejrzysty według nich — kontrprzykład zwany *Jasnym Pokojem*. „Wyobraźmy sobie ciemny pokój a w nim człowieka trzymającego duży magnes (...). Jeśli ten człowiek będzie wymachiwał magnesem do góry i na dół, to (...) powinno to wywołać koliste rozchodzenie się fal elektromagnetycznych, a zatem powinno zrobić się jasno. Jednak każdy z nas wie, że (...) jest nie do pomyślenia, byśmy mogli uzyskiwać światło prosto przez poruszanie magnesem w kółko”¹⁰.

Ktoś może twierdzić na podstawie powyższego przykładu, że światło nie ma związku z falami elektromagnetycznymi, gdyż potrząsanie magnesem nie daje światła w ciemnym pokoju. Odwołując się zaś do Chińskiego Pokoju można wykazywać, jak Searle, że potrząsanie, mieszanie chińskich symboli nie da „światła” rozumienia języka chińskiego. Analogicznie do argumentu z 1980 roku Churchlandowie również podają aksjomaty związane z Jasnym Pokojem:

Aksjomat 1: Elektryczność i magnetyzm są formami energii.

Aksjomat 2: Zasadniczą cechą światła jest jasność.

Aksjomat 3: Energia nie jest ani konieczna, ani wystarczająca dla uzyskania jasności.

Wniosek 1: Elektryczność i magnetyzm nie są konieczne ani wystarczające dla wyjaśnienia istoty światła.

Gdyby taki tok myślenia przedstawiono po ogłoszeniu koncepcji Maxwella na temat elektromagnetycznej natury światła a przed uznaniem faktu tożsamości światła i fal elektromagnetycznych, to aksjomaty te mogłyby stanowić podstawę odrzucenia teorii Maxwella — konkludują Churchlandowie. A przecież jednak wskutek badań teoria ta została potwierdzona. Toteż Churchlandowie radzą czekać powołując się na precedensy w historii nauki, kiedy to poglądy odrzucane były następnie akceptowane. Według nich być może późniejsze badania wykażą, że można

⁹P.M. Churchland, P. Smith Churchland, *Czy maszyna może myśleć?*, Świat Nauki, Lipiec (1991), 17–23. Por. też J.R. Searle, *Is the Brain's Mind a Computer Program?*, Scientific American, January (1990), 26–31.

¹⁰P.M. Churchland, P. Smith Churchland, *Tamże*, 20.

zbudować semantykę jedynie przy pomocy samej syntaktyki i wnioskii wyciągnięte z Chińskiego Pokoju uważają za przedwczesne.

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.

2. KWESTIA SEMANTYKI Z ROBOTA–I–KOMPUTERA

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 komentatorom¹¹. Można do niej zaliczyć takich autorów jak Bruce Bridgeman, Daniel Dennett, Jerry A. Fodor oraz Aaron Sloman z Moniką Croucher.

2.1 Znaczenie informacji ze świata zewnętrznego

Przyjęli oni założenie — komputer nie powinien ograniczać się do jedynie przyjmowania danych na wejściu i przekazywania ich na wyjściu; należy mu ułatwić kontakt ze środowiskiem, w jakim przebywa przez umieszczenie go wewnątrz robota. 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.

¹¹Searle określa tę grupę zarzutów mianem *robot reply*. Określenie to trudno przetłumaczyć na język polski; ze względu na treść replik można ten termin oddać opisowo jako *zastosowanie robota*. Odpowiedzi na zarzuty: J.R. Searle, *Minds, Brains, and Programs*, 420–421 oraz tegoż autora *Intrinsic Intentionality*, BBS 3 (1980), 451–455.

Komputer potrzebuje więcej informacji i to dwojakiego rodzaju — a priori (bazy danych) oraz o środowisku¹²; między symbolami formalnymi a rzeczami w świecie musi być odpowiednie połączenie przyczynowe — i te elementy wystarczą do intencjonalności¹³. System powinien być wystarczająco duży i należy do niego dołączyć wszystkie rodzaje operacji komputacyjnych¹⁴. Można rozważyć jeszcze jedną możliwość podaną przez Daniela Dennetta¹⁵ — co stałoby się, gdyby robot miał ciało człowieka; czy nadal nie zachodziłoby rozumienie? Po pierwsze należy pamiętać, by opisywać sytuację z punktu widzenia pierwszej osoby, a nie jak chcą operacjoniści — trzeciej. Załóżmy więc, że na dane dwa znaki chińskie program komputera zareaguje poleceniem podania soli; wtedy nie trudno się domyślić, jakie było znaczenie chińskiej instrukcji. Ale 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 miałyby miejsca rozumienie języka chińskiego.

2.2 Percepcja a manipulowanie symbolami

We wszystkich powyższych wariantach połączenia robota i komputera Searle dostrzega milczące uznanie następującego poglądu — percepcja nie jest jedynie manipulowaniem formalnymi symbolami, skoro muszą być dodawane przyczynowe połączenia ze światem zewnętrznym. Ale dodanie możliwości motorycznych czy percepcyjnych nic nie zmienia w najobszerniejszym nawet systemie — nadal brak rozumienia i intencjonalności. Robot–i–komputer jest jedynie odmianą eksperymentu myślowego — zamiast człowieka w pokoju mamy komputer w robocie, chińskie symbole na wejściu to bodźce z kamery a odpowiedź na wyjściu jest odpowiednikiem polecenia wydanego robotowi, by wykonał daną operację. Informacje są pobierane i wydawane bez znajomości ich znaczenia. Nadal nie widać możliwości powstawania stanów umysłowych w wyżej opisanych systemach.

Chiński Pokój może być powtarzany w dowolnie wielkim systemie i skutek pozostanie ten sam. Wprowadzanie nowych elementów nic nie da, jeśli system nie może połączyć treści semantycznych ze zbiorem syntaktycznych czynników. Jerry Fodor pisze, że dzieje się tak, gdyż dotąd

¹²B. Bridgeman, *Brains + Programs = Minds*, BBS 3 (1980), 427–428.

¹³*Tamże* oraz J.A. Fodor, *Searle on What only Brains Can Do*, BBS 3 (1980), 431.

¹⁴A. Sloman, M. Croucher, *How to Turn an Information Processor into an Understanding*, BBS 3 (1980), 447–448.

¹⁵*The Milk of Human Intentionality*, BBS 3 (1980), 428–430.

nie osiągnęliśmy *właściwego* rodzaju połączenia przyczynowego między symbolami a światem zewnętrznym; nie należy więc szczyć wysiłków, by osiągnąć zamierzony efekt. Można przecież jednak pomyśleć o dowolnym połączeniu między np. symbolem jajka fu jung a samym jajkiem fu jung. Połączenie to nigdy nie będzie zdolne do zinterpretowania tego symbolu jako symbol jajka fu jung. Fodor sugeruje też w sposób zaskakujący, by myśleć o mózgu czy komputerze jako o wykonujących formalne operacje na symbolach *zinterpretowanych* a nie tylko formalnych. Jednakże program komputerowy nie jest w stanie tego dokonać, interpretacji dokonuje człowiek.

Zastosowanie robota do całego systemu nie zaprzecza Chińskiemu Pokojowi — przy pomocy innych terminów nadal są wyrażane idee Searle'a, wbrew zamiarom ich autorów. Przysporzenie możliwości percepcyjnych bez interpretacji sygnałów docierających z zewnątrz nie zamieni możliwości motorycznych czy operacji komputacyjnych w stany umysłowe podobne do ludzkich.

3. MODEL SYMULATORA MÓZGU

Kolejna grupa komentarzy do Chińskiego Pokoju proponuje symulowanie nie tylko rozumienia języka ale samych połączeń mózgowych; coraz wyraźniej rysuje się tendencja do odchodzenia od maszyny odpowiadającej na pytania dotyczące restauracji. Taki sam kierunek obrał i Searle — o ile w 1980 roku oparł swój argument o idee zaczerpnięte z programu Schanka, o tyle kolejne wersje eksperymentu myślowego coraz bardziej od niego odchodzą; w 1990 roku Chiński Pokój odnosi się już do dowolnego programu formalnego i dowolnego komputera cyfrowego.

3.1 Naśladowanie działania formalnej struktury mózgu

Symulator mózgu — pod taki wspólny mianownik terminologiczny można wziąć analizy Douglasa Hofstadtera, Johna Haugelanda, Williama Lycana, Stevena Savitta oraz Donalda Waltera; symulacja pracy połączeń mózgowych rodowitego Chińczyka odbywałaby się na poziomie synaps. Na wejściu przyjmowane byłyby pytania i opowiadania w języku chińskim; kolejny krok to naśladowanie formalnej struktury mózgu Chińczyka podczas opracowywania chińskich odpowiedzi i w końcu przekazanie teże na wyjściu komputera. Cały system miałby działać w oparciu o programy równoległe wzorowane na sieciach neuronowych. Symulator mózgu schodzi więc do bardzo niskiego poziomu tj. synaps a dokładniej wyładowań neuronowych w synapsach; według opinii wyżej

wymienionych uczonych w tym wypadku musi zachodzić rozumienie — cóż bowiem mogłoby różnić mózgowy program Chińczyka od programu komputerowego?

Układ ten działałby automatycznie bez czyjegokolwiek wspomagania; gdyby symulator znał dany kontekst społeczny, w jakim przyszedłoby mu działać — należałoby, zgodnie z ideami funkcjonalizmu, zaakceptować opinię, że system ów rozumie język chiński¹⁶. 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 mózgu¹⁷. Obserwator nie jest w stanie tego stwierdzić — jest zbyt mały w stosunku do wielkiego układu symulatora mózgu, by dostrzec, że on rozumie¹⁸.

3.2 Replika — model rurkowy

Searle odpowiada na pomysł symulatora wyobrażeniem rurkowo-zaworowego modelu mózgu¹⁹. Zamiast manipulować symbolami człowiek ma do dyspozycji wypełniony wodą zestaw rurek z zaworami; każde połączenie to synapsa w mózgu rodowitego Chińczyka. Według autora słynnego artykułu *Minds, Brains and Programs* ani rurki ani człowiek nie mają zdolności rozumienia; co więcej — nie posiada jej także system człowiek-i-rurki-z-wodą. To czysty absurd sądzić, że człowiek i rurki coś wiedzą; w symulacji cieczerw-elektrycznej naśladowane jest nie to, co trzeba. Douglas Hofstadter jednak przypuszcza²⁰, że kilka trylionów rurek z wodą mogłoby funkcjonować w ten sposób, by wyprodukować świadomość — tak skomplikowany system mógłby myśleć wykonując program działania mózgu. Nie uchwycił on jednak istoty argumentu, bo jeśli nawet w tym wypadku miałyby zaistnieć świadomość, to urzeczywistnienie siły przyczynowej mózgu musiałyby nastąpić przy pomocy rurek wodnych a nie przez zastąpienie formalnego programu²¹.

¹⁶J.C.Eccles, *A Dualist-Interactionist Perspective*, BBS 3 (1980), 430-431.

¹⁷D.O.Walter, *The Termostat and the Philosophy Professor*, BBS 3 (1980), 449. Searle uważa porównanie do termostatu za zbyt słabą analogię — por. J.R. Searle, *Intrinsic Intentionality*, 455.

¹⁸W.G. Lycan, *The Functionalist Reply*, BBS 3 (1980), 434-435.

¹⁹J.R. Searle, *Minds, Brains, and Programs*, 420-421.

²⁰D.R. Hofstadter, *Reductionism and Religion*, BBS 3 (1980), 433-434.

²¹J.R. Searle, *Minds, Brains, and Programs*, 454-455.

3.3 Demon Haugelanda

Najwięcej uwagi poświęca Searle argumentowi Johna Haugelanda²², który uważa za prawdziwie oryginalny. Założmy, że siła przyczynowa u rodowitego Chińczyka jest zdolnością przylegających neuronów do stymulowania siebie nawzajem przez synapsy. Powstrzymajmy teraz wyładowania neuronowe przez pokrycie synaps cienką warstwą lakieru; brak ten zastępuje człowiek z Chińskiego Pokoju, zwany przez Haugelanda *demonem Searle'a* — zamiast symbolami będzie on manipulował stanami fizycznymi synaps pobudzając neurony w ten sposób, jakby zachodziły w nich prawdziwe wyładowania. Przez obserwowanie stanu neuronów i stymulowanie ich bądź nie, demon będzie symulował siłę przyczynową oraz urzeczywistniał intencjonalność, którą autor pomysłu określa w tym wypadku jako pochodną. Według Haugelanda wyżej wymienione rozumowanie jest zaprzeczeniem argumentu. Zastrzeżenia Searle'a wzbudza rzekome posiadanie przez demona siły przyczynowej z powodu symulowania wyładowań neuronowych. Według Haugelanda: „[Chińczyk] nadal działa tak jakby rozumiał chiński. Obecnie choć żaden z jego neuronów nie ma właściwej siły przyczynowej — ma ją demon, który ciągle rozumie tylko angielski”²³.

Twórca idei *demonia Searle'a* sądzi, że systemy AI mogą być rozumne, jeśli ich działalność jest bogata semantycznie. Twierdzi on również, że istnieje możliwość takiego zaprojektowania tych układów sztucznej inteligencji, by rzeczywiście były semantycznie aktywne i posiadały oryginalną intencjonalność. Haugeland myli się jednak mówiąc, że tego typu intencjonalność jest wystarczająco bogata w jej semantycznej działalności — semantyka jest nadal zależna od obserwatora, to on przypisuje znaczenia. Sam demon nie potrafi urzeczywistnić przyczynowej siły, więc jego symulowanie wyładowań neuronowych nie będzie miało skutku w postaci rozumienia języka przez Chińczyka z synapsami powleczonymi lakierem²⁴.

3.4 Myślenie mózgu o n-1 neuronach

W 1982 roku Steven Savitt²⁵ w kolejnej serii komentarzy w *The Behavioral and Brain Sciences* wyraził pogląd, że argument Haugelanda nie jest wykorzystany i zanalizowany do końca. Opublikował więc swoją

²²J. Haugeland, *Programs, Causal Powers and Intentionality*, BBS 3 (1980), 432–433.

²³Tamże, 432.

²⁴J.R. Searle, *Minds, Brains, and Programs*, 452–453.

²⁵*Searle's Demon and the Brain Simulator*, BBS 5, 2 (1982), 342–343.

wersję wprowadzając szereg zmian w stosunku do oryginalnego pomysłu. 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 miałyby 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.

Savitt w rezultacie wcale nie omija dylematu, jaki Searle przedstawił Haugelandowi — albo demon urzeczywistnia siłę przyczynową mózgu i otrzymujemy intencjonalność (ale siła ta nie polega przecież na mieszanii formalnych symboli) albo demon manipuluje znakami bez reprodukcji wspomnianej siły. Pierwszą możliwość Searle odrzuca — zgodnie ze swoim eksperymentem myślowym, druga nie daje oczekiwanych przez Savitta rezultatów. Żadne próby tuszowania braków²⁶ nie są w stanie obronić tego wariantu Savitta, a dla Searle'a bardziej interesujący jest argument w jego oryginalnej postaci — autorstwa Johna Haugelanda.

4. PROPOZYCJA KOMBINACYJNA

Środowiska uniwersyteckie Berkeley i Stanford oraz Wiliam Lycan a także Thomas Edelson wyrażają nadzieję, że połączenie omówionych powyżej trzech idei wymierzonych przeciw Chińskiemu Pokojowi stanowi jego zaprzeczenie; stanowisko to można określić mianem propozycji kombinacyjnej.

4.1 Problem intencjonalności w systemie robot + komputer + symulator

Trzy poprzednie idee traktowane osobno są nieprzekonywujące, ale wzięte razem przesądają — według autorów — o odrzuceniu argumentu Johna Searle'a: systemowi złożonemu z robota, komputera i symulatora mózgu należy przypisać intencjonalność. Istotnym czynnikiem w wypadku umiejętności rozumienia jest możliwość kontaktu ze światem ze-

²⁶Por. *Tamże*, 347.

wnętrznym²⁷. Na zastrzeżenie Searle'a — gdzie jest rozumienie w tym systemie? — Wiliam Lycan²⁸ odpowiada, że obserwator o wymiarach pojedynczej komórki jest w mózgu zbyt mały, by je dostrzec.

Jeśli byłoby możliwe zbudowanie takiego układu robot–komputer–symulator, którego zachowanie byłoby nieodróżnialne od ludzkiego, wtedy takiemu systemowi można przypisać intencjonalność. Gdy bowiem coś wygląda i zachowuje się jak my — musi mieć stany umysłowe. Podkreślmy, że jest to stanowisko behawioryzmu, opisywanie sytuacji z pozycji obserwatora, trzeciej osoby. Tak można byłoby sądzić, gdybyśmy nic nie wiedzieli o składowych układu, tymczasem mamy przecież wgląd do niego i możemy dalej prowadzić analizy nie zatrzymując się jedynie na zewnętrznym zachowaniu.

4.2 Odpowiedź Searle'a — mechaniczny manekin

Komputer w robocie, nawet jeśli jest symulatorem mózgu na poziomie synaps działającym na podstawie systemów konekcyjnych, manipuluje symbolami bez znajomości ich znaczenia. Stąd należy uznać robota za pomysłowego mechanicznego manekina. Nie ma podstawy, by przypisywać mu intencjonalność; nic nie potwierdza hipotezy, że układ robot–komputer–i–symulator myśli. John Searle²⁹ porównuje go w tym wypadku do fauny i wskazuje na powody, dla których należy robotowi odmówić intencjonalności. U zwierząt zauważamy takie same elementy składowe jak u ludzi: kończyny, skórę, oczy, mózg etc. Nie możemy uznać ich zachowania za sensowne bez przypisania im intencjonalności, ich stany umysłowe muszą powstawać na zasadach podobnych do naszych. Biorąc pod uwagę powyższe porównanie widać różnice w odniesieniu do robota z komputerem i symulatorem mózgu — nie ma podstaw do przypisania intencjonalności takiemu systemowi. Wiemy bowiem, że jego działanie jest wynikiem formalnego programu a przyczynowe własności fizycznych substancji robota nie mają nic wspólnego z intencjonalnością.

Tak więc i połączenie odpowiedzi systemowej z zastosowaniem robota i symulatora mózgu nie jest w stanie zanegować eksperymentu myślowego Johna Searle'a.

²⁷T. Edelson, *Stimulating Understanding: Making the Example Fit the Question*, BBS 5, 2 (1982), 338–339.

²⁸*The Functionalist Reply*, BBS 3 (1980), 434–435.

²⁹*Minds, Brains, and Programs*, 421. Por. też jego *The Chinese Room Revisited*, 346 oraz stanowisko przeciwników propozycji kombinacyjnej: B. Libet, *Mental Phenomena and Behavior*, BBS 3 (1980), 434 a także T. Edelson, *Stimulating Understanding: Making the Example Fit the Question*, 338–339.

5. PERSPEKTYWA MOŻLIWYCH PRZYSZŁYCH ROZWIĄZAŃ

Argument Chińskiego Pokoju zakłada, że badania nad sztuczną inteligencją odnoszą się tylko do komputerów analogowych i cyfrowych — w przyszłości natomiast być może nowe rozwiązania technologiczne umożliwią sztuczne tworzenie procesów przyczynowych, czego Searle nie bierze już pod uwagę. Program Schanka jest zbyt prymitywny — przyszłe zdystansują go na tyle, że eksperyment myślowy nie będzie się do nich odnosił. Tak brzmią zasadnicze zarzuty stawiane przez niektórych uczonych z Berkeley oraz takich naukowców jak Ned Block, Aaron Sloman z Moniką Croucher, Daniel Dennett, Bruce Bridgeman, William Lycan oraz Roger Schank.

5.1 Redefiniowanie mocnego programu sztucznej inteligencji

System musi być o wiele większy, należy umożliwić mu wykonywanie innych operacji³⁰. Trzeba wprowadzić do programów umiejętność uczenia się systemu i ułatwić mu czerpanie informacji z relacji ze światem zewnętrznym³¹ — słowem *obecne* programy są niewłaściwe³². Jeśli komputer wydziela ciepło i szumi, dlaczego nie miałby mieć intencjonalności³³?

Mocny program sztucznej inteligencji został zdefiniowany przez wyżej wymienionych autorów i jest przez nich pojmowany jako jakiegokolwiek poznanie sztucznie osiąganego i wyjaśniane. Gdyby tak rozumieć AI to eksperyment myślowy nie miałby rzeczywistego zastosowania — replikuje Searle; mocny program w swym oryginalnym brzmieniu jest precyzyjnie określony: „Procesy umysłowe są procesami komputacyjnymi których granice określają formalne składowe”³⁴.

Powróćmy na chwilę do argumentu Searle’a — człowiek z Chińskiego Pokoju nie rozumie chińskiego ale rozumie angielski, komputer nie pojmuje żadnego z nich. Wobec tego w człowieku musi tkwić jakaś zdolność powodująca rozumienie angielskiego; czy jest sposób, by przekazać ją maszynie? Jej operacje są określane działaniami komputacyjnymi a czło-

³⁰A. Sloman, M. Croucher, *How to Turn an Information Processor into an Understanding*, 447–448.

³¹B. Bridgeman, *Brains + Programs = Minds*, 427–428.

³²W.G. Lycan, *The Functional Reply*, 434–435. Por. też R.C. Schank, *Understanding Searle*, BBS 3 (1980), 446–447: wiąże on też pewne nadzieje z przyszłymi rozwiązaniami odnośnie do programów komputerowych; podkreśla także, że nawet jeśliśmy mogli układać programy, które działałyby tak, jak gdyby rozumiały i świadomie myślały — to czy rzeczywiście miałyby one takie stany umysłowe?

³³D. Dennett, *The Milk of Human Intentionality*, BBS 3 (1980), 430. Por. też N. Block, *What Intuitions about Homunculi Don't Show*, BBS 3 (1980), 425–426.

³⁴J.R. Searle, *Minds, Brains, and Programs*, 422.

wiek ma strukturę biologiczną z odpowiednimi procesami chemicznymi i fizycznymi; jest ona przyczynowo zdolna do wytwarzania takich zjawisk intencjonalnych jak uczenie się, rozumienie, percepcja itp. Jeśli cokolwiek innego miałyby taką siłę przyczynową, mogłyby mieć również intencjonalność — o tym mówi argument Chińskiego Pokoju. Czy inne procesy fizyczne i chemiczne mogą wywołać takie skutki? — jest to pytanie empiryczne.

5.2 Modele formalne a intencjonalność

Z eksperymentu myślowego Searle'a wynika, że żaden czysto formalny model sam przez się nie wystarczy do wytworzenia intencjonalności — dla niej formalne własności nie są konstytutywne. Jedynym ich skutkiem jest spowodowanie kolejnego kroku programu nie zaś siła przyczynowa. Wszelkie inne własności przyczynowe są nieistotne dla modelu formalnego, ponieważ może on być realizowany na różnego typu maszynach nie posiadających tych własności.

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.

6. OCENY KRYTYKI ARGUMENTU SEARLE'A

Wydaje się, że część autorów replik w ogóle nie zrozumiała istoty Chińskiego Pokoju³⁵; Searle zalicza do nich: Johna Marshalla, E.W. Menzela Jr., Martina Ringle oraz Jerry Sameta; ich artykuły zostały pominięte w niniejszej pracy³⁶. Douglas R. Hofstadter jest z kolei

³⁵Por. J.R. Searle, *Intrinsic Intentionality*, 455; Tenze, *The Chinese Room Revisited*, 345.

³⁶Można je znaleźć w obydwu seriach komentarzy w *The Behavioral and Brain Sciences* z 1980 i 1982 roku: J. Marshall, *Artificial Intelligence — The Real Thing?*, BBS 3 (1980), 435–437; E.W. Menzel Jr., *Is the Pen Mightier than the Computer?*, BBS 3 (1980), 438–439; M. Ringle, *Mysticism as a Philosophy of Artificial Intelli-*

przykładem komentatora, który rzuca na Searle'a inwektywy³⁷ i przypisuje mu dokładnie odwrotne poglądy niż jego własne. Metoda autora słynnego bestselleru *Gödel, Escher, Bach. An Eternal Golden Braid* polega ogólnie na mówieniu $\sim p$ tam, gdzie Searle pisze p ; przykładem jest odniesienie do dualizmu, który pomysłodawca Chińskiego Pokoju odrzuca zaś według Hofstadtera ma być jego zwolennikiem³⁸.

Większość autorów podjęła jednak mniej czy więcej rzeczową dyskusję — podsumujmy ją.

6.1 Nieadekwatność zarzutów i metody komentatorów

John Searle jednoznacznie ocenia komentarze z lat 1980–1982: żaden z nich nie zaprzeczył jego eksperymentowi myślowemu³⁹. Swojej opinii nie zmienia i podobną wydaje w 1990 roku: „Wszystkie (...) argumenty mają wspólną cechę: są zupełnie nieadekwatne, ponieważ błędnie ujmują istotę (...) dowodu Chińskiego Pokoju. Opiera się on na różnicy między manipulacją formalnymi symbolami, która jest wykonywana przez komputer a umysłowymi działaniami biologicznymi mózgu; tę różnicę skróciłem — mam nadzieję nikogo nie wprowadzając w błąd — do różnicy między syntaktyką a semantyką”⁴⁰.

Mimo błędów, nieuzasadnionych zarzutów i niezrozumienia, jakich pełno w komentarzach, stanowiły one bodziec dla Searle'a do ponownego przeanalizowania i uściślenia wielu elementów argumentu — m.in. jasno określił on aksjomaty i wnioski, zmienił ich kolejność oraz uogólnił postać Chińskiego Pokoju. Stosunek Searle'a do poszczególnych grup replik został już opisany w niniejszym artykule. Dokonajmy teraz ogólniejszego podsumowania krytyki pod adresem eksperymentu myślowego.

Najbardziej Searle ceni sobie pomysł Johna Haugelanda dotyczący stymulowania przez *demoną Searle'a* wyładowań w synapsach — uważa go za bardzo oryginalny⁴¹. Żaden z komentatorów nie zdołał osłabić logicznej struktury argumentu⁴² — ich artykuły są natomiast pełne wła-

gence, BBS 3 (1980), 444–445; J. Samet, *Understanding and Integration*, BBS 5,2 (1982), 341–342.

³⁷W komentarzu poświęconym kluczowemu artykulowi Searle'a *Minds, Brains, and Programs* w ten sposób określa tę pracę: „Jeden z najgorszych, najbardziej doprowadzających do szału artykułów jaki przeczytałem w moim życiu”; Por. D. R. Hofstadter, *Reductionism and Religion*, 433.

³⁸Por. J.R. Searle, *Intrinsic Intentionality*, 454.

³⁹*Tamże*, 456.

⁴⁰J.R. Searle, *Is the Brain's Mind a Computer Program?*, 30.

⁴¹*Intrinsic Intentionality*, 452.

⁴²Por. J.R. Searle, *Turing the Chinese Room*, w: *Synthesis of Science and Religion: Critical Essays and Dialogues*. World Congress for the Synthesis of Science and Reli-

snych uzupełnień, dogmatycznych stwierdzeń, założeń, błędnych analogii oraz powtarzania idei z pewnymi zmianami.

By utworzyć kontrprzykłady na podstawie samego Chińskiego Pokoju wyliczali oni liczne uzupełnienia systemu: przetwarzanie równoległe, drugiego człowieka wewnątrz pokoju, symulator mózgu, czy robota mającego możliwość dostarczania informacji do komputera i który pozwalałby układowi poruszać się i wykonywać inne czynności; czasem są również łączone różne elementy jak np. w propozycji kombinacyjnej. Proponujący wszelkie dodatki zapominają przy tym, że milcząco uznają niewystarczalność operacji komputacyjnych w tworzeniu stanów umysłowych. Dodatkowe elementy w systemie przysparzają co prawda nowych danych, ale są one niezinterpretowane i w żaden sposób nie zwiększają szansy na powstanie intencjonalności. Zwykle też po opisanu nowego układu nagle pojawia się bezpodstawne dogmatyczne stwierdzenie typu: *teraz musi zachodzić rozumienie*. Autorzy nie podają konkretnych powodów dla których miałyby powstawać rozumienie, podczas gdy argument Searle'a ma solidną logiczną strukturę, aksjomaty i wnioski.

W symulatorze mózgu mamy do czynienia z akceptacją poglądu (bez dowodu), że ludzki mózg działa według programu komputerowego. Snute porównania nie są poprzedzone analizą, np. Churchlandowie w *Jasnym Pokoju* bez sprawdzenia poprawności toku myślenia przyjęli analogię między elektromagnetyzmem a światłem oraz manipulowaniem symbolami chińskimi a semantyką. Ci sami autorzy nie są odkrywcy co do zastosowania w układzie przetwarzania równoległego, bowiem powielają jedynie pomysł zwolenników symulatora mózgu z 1980 i 1982 roku. Podobna uwaga dotyczy Savitta wprowadzającego zmiany do jednak wyeksploatowanego pomysłu Haugelanda, przez co wcale nie unika zarzutów Searle'a, jakie miały miejsce w odniesieniu do *demonia Searle'a*.

Na końcu artykułu zawierającego odpowiedzi do komentarzy na temat *Minds, Brains and Programs* Searle zaznacza, że Chiński Pokój przetrwał wszystkie ataki; jednocześnie stara się on skierować uwagę naukowców z kręgu sztucznej inteligencji w odpowiednią, jego zdaniem, stronę — ku złagodzonemu programowi sztucznej inteligencji (Weak Artificial Intelligence). Według niego jest to odpowiednia dziedzina badań — odkrywcza, interesująca i wystarczająco trudna⁴³.

gion (1986: Bombay, India), ed. T.D. Singh, Ravi Gomatam, Bhaktivedanta Institute, San Francisco 1987, 301.

⁴³J.R. Searle, *Intrinsic Intentionality*, 456.

6.2 Istotne dyskusje

Część komentarzy z *The Behavioral and Brain Sciences* z roku 1980 oraz 1982 nie została dotąd omówiona. Dotyczą one różnych zagadnień: sposobu działania mózgu, wewnętrznej intencjonalności, przypisywania intencjonalności z pozycji obserwatora oraz mocnego programu sztucznej inteligencji a także złagodzonego podejścia do AI, wzajemnego odniesienia się do siebie semantyki i syntaktyki oraz koncepcji umysłu. Wokół tych zagadnień koncentrują się również artykuły na temat eksperymentu myślowego Searle'a opublikowane po 1982. Pojawia się coraz więcej wyważonych opinii na temat Chińskiego Pokoju, wzrasta liczba uczonych akceptujących argument, przestają grać rolę ambicje i nerwy — zastępują je rzeczowe rozważania. Należą do nich również krytyczne głosy pod adresem niektórych poglądów autora Chińskiego Pokoju; zaliczyć do nich należy m.in. opinię Penrose'a na temat pojmowania ludzi jako maszyn szczególnego rodzaju⁴⁴.

Wymienione powyżej zagadnienia dotyczą filozoficznych konsekwencji eksperymentu myślowego; będą przedmiotem analiz drugiej części niniejszej rozprawy. Czy można przyjąć, że program komputerowy jest nie tylko modelem umysłu, ale samym umysłem? A może jednak należy traktować komputerowy model jako użyteczne narzędzie w badaniach nad intelektem? W jaki sposób na podstawie samej syntaktyki dojść do znaczenia? Czy to w ogóle jest możliwe? Czy dopuszczalne jest twierdzenie, że semantyka nie istnieje i wystarcza tylko syntaktyka? Tymi kwestiami zajmiemy się bliżej w kolejnych częściach artykułu. Nawiążemy tam również do poglądów komentatorów z 1980 i 1982 roku, którzy bardziej czy mniej zgadzają się z Searlem. Należą do nich Arthur C. Danto⁴⁵, B. Libet⁴⁶, Grover Maxwell⁴⁷, Ronald Pucetti⁴⁸ oraz Thomas Natsoulas⁴⁹.

⁴⁴R. Penrose, *The Emperor's New Mind. Concerning Computers, Minds, and the Laws of Physics*, Penguin Books, New York 1991, 22–23.

⁴⁵*The Use and Mention of Terms and the Simulation of Linguistic Understanding*, BBS 3 (1980), 428.

⁴⁶*Mental Phenomena and Behavior*, BBS 3 (1980), 434.

⁴⁷*Intentionality: Hardware, not Software*, BBS 3 (1980), 437–438.

⁴⁸*The Chess Room: Further Demythologizing of Strong AI*, BBS 3 (1980), 441–442.

⁴⁹*The Primary Source of Intentionality*, BBS 3 (1980), 440–441.