Przegląd Filozoficzny – Nowa Seria
R. 20: 2011, Nr 2 (78), ISSN 1230–1493

Marcin Jażyński

Qualia w chińskim pokoju. Obliczanie i świadomość fenomenalna

Słowa kluczowe: *qualia, świadomość, umysł, pierwszoosobowa/trzecioosobowa perspektywa, system poznawczy, funkcjonalizm, redukcjonizm, eksperymenty myślowe, obliczeniowość*

1. Co ma wspólnego chiński pokój z qualiami?

Eksperyment myślowy Searle'a ma wykazać, że sztuczny system, jak chiński pokój, nigdy nie będzie rozumiał tego, „co robi". Biorąc za dobrą monetę stanowisko Searle'a, można pójść krok dalej i w ten sam sposób argumentować, że stosując strategię kognitywistyczną nie da się wyjaśnić nie tylko intencjonalności, ale też świadomości, w tym świadomości fenomenalnej. Do takiego samego wniosku ma skłaniać inny eksperyment myślowy, znany jako młyn Leibniza. Oba eksperymenty mają podobną strukturę. Jednak ich konkluzywność jest dyskusyjna. Postaram się to pokazać, by następnie na tym tle rozważyć stanowisko Chalmersa. Muszę przyznać, że wydaje mi się ono dziwne. Chalmers z jednej strony stanowczo odpiera zarzuty Searle'a, z drugiej zaś wniosek Leibniza o tym, że świadomości nie można wyjaśnić w mechanicystyczny sposób, jest, jak sądzę, zgodny ze stanowiskiem Chalmersa.

1.1. Searle

Poglądy Searle'a dotyczą zarówno natury stanów umysłowych, jak i sposobów ich opisywania. Jego stanowisko w ogólnym zarysie przedstawia się tak oto.

Obliczeniowe teorie funkcjonalistyczne traktują umysł jako swego rodzaju maszynę syntaktyczną. Pogląd taki oznacza, że „w umyśle nie ma nic biologicznego". To znaczy, że własności mózgu nie wpływają na własności umy-

słu. W istocie, konsekwencją tezy o możliwości wielorakiej realizacji może być twierdzenie o swego rodzaju niezależności umysłu od mózgu. W tym świetle mózg byłby jedną z możliwych fizycznych realizacji „programu" umysłowego, który równie dobrze mógłby być realizowany w innej strukturze fizycznej o dostatecznie dużej złożoności. Teza, wedle której umysł jest programem, a mózg swoistym hardware'em, jaki go realizuje, jest postulatem mocnej wersji sztucznej inteligencji (AI)[1]. AI jest radykalną teorią umysłu, opierającą się na tezach funkcjonalistycznych. Funkcjonalizm głosi, że jeśli możliwa jest różnoraka realizacja fizyczna[2], to własności mentalne jednego typu mogą przysługiwać różnego typu strukturom fizycznym lub też stany mentalne identycznego typu mogą być realizowane przez struktury fizyczne różnych typów. Zatem to, jaki pod względem fizycznym jest dany system, jest nieistotne. Musi on być jedynie na tyle złożony, by mógł funkcjonować w pewien sposób, tzn. realizować pewien program. Z tezy tej wynika, że tym, co konstytuuje umysł, jest funkcjonalność, tj. pewien sposób zorganizowania i funkcjonowania elementów danego systemu fizycznego. Sposób ten i organizację funkcjonalną można scharakteryzować obliczeniowo jako zbiory sekwencji procesów obliczeniowych przebiegających wedle określonych reguł syntaktycznych. Zatem istotna jest tu funkcjonalna organizacja procesów umysłowych, a nie własności pewnej struktury fizycznej, np. mózgu. Tymczasem wydaje się rozsądne i, łagodnie mówiąc, potwierdzone przez doświadczenie, że to, co się dzieje z naszym mózgiem, ma zasadniczy wpływ na przebieg procesów umysłowych. Słowem, oczywiste wydaje się to, że charakter procesów mózgowych ma znaczenie dla przebiegu czynności umysłowych. Zdaniem Searle'a występowanie własności umysłowych – jak intencjonalność – jest wynikiem przebiegania charakterystycznych procesów neurofizjologicznych. Wedle niego umysł jest skutkiem i własnością procesów mózgowych: „psychiczność jest skutkiem procesów w neuronach lub zespołach neuronalnych mózgu (modułach), lecz jednocześnie jest ona rzeczywistą cechą systemów nerwowych". I dalej: „(...) sposobem wyjaśnienia intencjonalności jest dokładne pokazanie, jak zjawiska intencjonalne mogą być cechą systemów żywych, a zarazem skutkiem procesów biologicznych. Doznania słuchowe, wzrokowe, wrażenia dotykowe, pragnienia, pożądania seksualne są skutkiem procesów mózgowych, są realnymi właściwościami struktury mózgu, są jednocześnie procesami intencjonalnymi" (Searle 1999: 45).

Searle jednak, moim zdaniem, krytykuje teorie obliczeniowe w niewłaściwy i nietrafny sposób. Twierdzi bowiem, że umysł nie może być niezależny od mózgu, ponieważ nie jest maszyną obliczeniową. Gdyby nią był, wówczas

[1] Choć może się to wydawać dziwne, przynajmniej na pierwszy rzut oka, Chalmers, mimo swojego nieredukcyjnego funkcjonalizmu broni mocnej wersji sztucznej inteligencji (i możliwości wytworzenia przez sztuczny system świadomości).

[2] U Chalmersa jest to tzw. zasada niezmienności organizacyjnej.

nie miałby własności semantycznych i przyczynowych, a jedynie syntaktyczne, tymczasem pewne jest to, że stany umysłowe posiadają swoje treści semantyczne i są przyczynami działań. Innymi słowy: albo obliczeniowość, albo treść i przyczynowość umysłowa, a teorie obliczeniowe z zasady nie mogą uwzględniać semantyki stanów umysłowych, czy też nie mogą ujmować tej ich własności, jaką jest treść. Searle pisze:

> Mieć umysł to coś więcej niż realizować formalne czy też syntaktyczne operacje. Nasze stany umysłowe na mocy definicji mają zawsze jakąś treść. Jeżeli myślę o Kansas City, życzyłbym sobie wypić szklankę zimnego piwa bądź zastanawiam się, jaki będzie spadek notowań giełdowych, w każdym wypadku mój stan umysłowy, niezależnie od tego, jakie formalne właściwości mu przypisujemy, ma jakieś psychiczne treści. To znaczy, nawet jeśli moje pragnienia są ciągiem symboli, musi być w myśleniu coś więcej niż abstrakcyjne symbole, gdyż ciągi symboli same w sobie nie mają żadnego znaczenia. Jeżeli myśl jest zawsze myślą o czymś, przeto dany ciąg symboli musi mieć jakieś znaczenie, by stał się myślą. Mówiąc krótko, umysł ma coś więcej niż syntaktykę. Powód, dla którego komputerowy program nie może być umysłem, jest prosty, komputerowy program ma cechy syntaktyczne, umysły mają coś więcej niż syntaktyka. Umysły są semantyczne w tym sensie, że poza strukturą formalną mają jeszcze jakieś treści. (...) Procesy umysłowe nie są obliczeniowe w żadnym interesującym sensie. Gdyż obliczeniowość wyklucza semantyczność. A stany umysłu charakteryzują się tym, że mają własności semantyczne, w oparciu o które działamy (Searle 1995: 45).

Zatem, twierdzi Searle, umysł nie jest komputerem lub procesy umysłowe nie są obliczeniowe. Obliczeniowość bowiem wyklucza semantykę, a także możliwość fizycznego oddziaływania przyczynowego, jak w przypadku myśli na inne myśli i zachowanie, jako że z definicji struktury obliczeniowe posiadają tylko własności syntaktyczne. Działanie struktur czy systemów obliczeniowych nie wystarcza, by można je stawiać na równi z umysłami, takimi jak ludzkie. Znaczy to, że umysły nie są systemami obliczeniowymi ani procesy umysłowe nie są procesami obliczeniowymi. „Artefakt taki, by mogły w nim zachodzić procesy umysłowe, powinien mieć zdolność przyczynowego oddziaływania porównywalną z możliwością ludzkiego mózgu" (Searle 1995: 52). Żeby zaś to było możliwe, musiałby posiadać semantykę, tj. systemy obliczeniowe musiałyby mieć obok własności syntaktycznych także semantyczne.

Argumentacja Searle'a przebiega, ogólnie mówiąc, tak: po pierwsze, systemy obliczeniowe są z definicji syntaktyczne, a jako takie nie mogą posiadać własności semantycznych i przyczynowych. Po drugie, nasze stany umysłowe cechuje to, że zawsze mają one pewną treść semantyczną i są przyczynami innych stanów umysłowych oraz zachowań. Występowanie treści semantycznych oraz przyczynowość mentalna należy do ich istotnych własności. Zatem stany umysłowe nie są stanami obliczeniowymi, a umysł (mózg) nie jest maszyną obliczeniową.

Problem więc polega na tym, czy można bez sprzeczności utrzymywać że: (1) procesy umysłowe mają charakter obliczeniowy, oraz że (2) procesy umysłowe rozumiane jako obliczeniowe mogą posiadać własności semantyczne i przyczynowe. Krótko mówiąc, pytanie dotyczy tego, czy pewnym procesom obliczeniowym, a więc syntaktycznym, mogą przysługiwać te dwa rodzaje własności? I jak jest możliwe powiązanie i współwystępowanie tych cech w stanach umysłowych?

1.2. Chiński pokój

Chciałbym pokazać, że argumentacja i zarzuty Searle'a nie stanowią większego problemu dla funkcjonalizmu komputacyjnego. Po pierwsze, jego wątpliwości dotyczące możliwości występowania czy nawet symulowania własności semantycznych stanów umysłowych w systemach obliczeniowych dają się uchylić na kilka sposobów. Po drugie, mamy dobre metody na pokazanie mechanizmu współwystępowania syntaktyki i semantyki w obliczeniowym modelu umysłu. Prawdziwe trudności, jakie ma metafora komputerowa, tkwią gdzie indziej.

Istota tego argumentu polega na tym, że „chiński pokój" – czysto syntaktyczny system – ma być równoważny funkcjonalnie pod względem organizacji modelowi umysłu, jaki proponuje metafora komputerowa. I dalej, jeżeli chińskiemu pokojowi nie można przypisywać cech rozumienia, intencjonalności zachowania – a wedle Searle'a nie można ze względu na czysto syntaktyczny sposób jego działania – to także nie można modelować tych zjawisk tak, jak proponuje funkcjonalizm komputacyjny. Inaczej mówiąc, choć wedle Searle'a mózgi „wytwarzają" umysły lub są przyczynami ich istnienia, to nie „robią" tego w taki sposób, jaki proponuje funkcjonalizm komputacyjny.

Dlaczego mamy tak sądzić? Prześledźmy kilka kroków argumentacji Searle'a:

(1) Searle twierdzi, że mózgi „wytwarzają" umysły. To znaczy, że procesy mózgowe o specyficznej, neurobiologicznej charakterystyce są przyczynami występowania stanów i procesów umysłowych. Stany umysłowe mają własności semantyczne, są intencjonalne.

(2) Zatem z (1): procesy mózgowe są przyczynami występowania tych własności – posiadania semantycznej treści przez stany umysłu oraz ich intencjonalności.

(3) Zachowanie komputera polega na czysto syntaktycznych manipulacjach na symbolach przebiegających wedle reguł. A operacje syntaktyczne nie mogą prowadzić do wyłonienia się własności semantycznych ani nie mogą poprawnie naśladować procesów, którym można przypisać własności semantyczne, jak na przykład rozumienie, myślenie, działania intencjonalne. Ogólnie: syntaktyka nie wystarcza dla semantyki.

(4a) Zatem maszyny syntaktyczne, na przykład komputery, nie mogą posiadać istotnych cech, jakie przysługują umysłom. Nie mogą bowiem mieć stanów umysłowych o treściach psychicznych.

(4b) Poprzez działania syntaktyczne nie można także naśladować działań o semantycznych cechach.

Z (4a) i (4b) mamy wnioskować, że maszyna syntaktyczna nie może być umysłem ani nie może naśladować umysłu. Model umysłu proponowany przez *cognitive science* i metaforę komputerową jest (i musi być) fałszywy, nieadekwatny do zjawisk, jakie modeluje.

(4c) Model obliczeniowy nie jest także modelem działania mózgu. Przyczyna jest taka sama, gdyż jest to model syntaktyczny, a syntaktyka nie wystarcza dla semantyki. Natomiast procesy mózgowe są przyczyną wystąpienia stanów umysłowych, które posiadają treści semantyczne/intencjonalne.

1.3. Kontrargumentacja

Zmierza ona w kilku kierunkach:

(1) Związek między syntaktyką i semantyką jest inny, niż sądzi Searle. A przynajmniej syntaktyka może naśladować semantykę (Fodor, Lem, Churchlandowie). Nas szczególnie interesuje punkt drugi, który mówi, że:

(2) W pewnym sensie chiński pokój jest metaforycznym modelem działania mózgu/umysłu. Jeśli godzimy się na to, że procesy mózgowe są odpowiedzialne za, na przykład, rozumienie, to tym, „co rozumie", nie są poszczególne części mózgu, ale cały system, na którego działanie składają się „bezrozumne" działania neuronów. Paradoksalnie, mózg działa jak chiński pokój.

Przypomnijmy, że chiński pokój jest równoważny funkcjonalnie maszynie syntaktycznej. Być może więc działania maszyny syntaktycznej (jak maszyna Turinga) są równoważne funkcjonalnie działaniom mózgu/umysłu. Jeśli na daleko posuniętym poziomie ogólności chiński pokój i mózg wykazują funkcjonalne podobieństwo działania, to możemy sądzić jedno z dwojga: system jako całość „wytwarza" semantykę i intencjonalność (mózg i „pokój"). Albo także mózg (obok „pokoju") nie „wytwarza" intencjonalności, co jest niezgodne z założeniem Searle'a.

Gdzie leżą podobieństwa między mózgiem i maszyną Turinga? W funkcjonalnej charakterystyce. Mówiąc wprost, mózg i komputer funkcjonują w ten sam sposób, czy jeszcze prościej: „robią to samo". Jest to wyraz silnej wersji tezy o wielorakiej fizycznie realizacji stanów umysłowych i silnej wersji AI. Według Turinga: „Nie interesuje nas fakt, że mózg ma konsystencję zimnej owsianki: Nie powiemy: Ta maszyna jest całkiem twarda, czyli nie jest mózgiem, a więc nie może myśleć" (Hodges 1998: 38). Fizyczne własności obu systemów nie grają większej roli. Dla procesów umysłowych i tego, jak działa

umysł, istotne jest to, co ujawnia opis funkcjonalno-komputacyjny, w tym obliczeniowy charakter tych procesów. „Jeśli chcemy znaleźć takie podobieństwo (tj. istotne podobieństwa między mózgiem a komputerem), powinniśmy szukać raczej matematycznych analogii funkcjonalnych" (Hodges 1998: 42). Wedle tego ujęcia mózg i komputer są systemami funkcjonalnie równoważnymi.

W swoim artykule *Maszyna licząca a inteligencja* Turing analizuje i odpiera różnego rodzaju argumenty przeciw tezie, że maszyna może myśleć. Wśród nich znajduje się argument Jeffersona, przypominający stanowisko Searle'a. Jest to tak zwany argument ze świadomości. Ma on prowadzić do wniosku, że maszynie nie może przysługiwać myślenie i świadomość. Czytamy tam między innymi:

> Dopóki maszyna nie ułoży sonetu czy nie napisze koncertu, nie dzięki przypadkowemu zestawieniu symboli, ale z potrzeby myśli i uczuć, dopóki wiec napisawszy takie dzieło nie będzie wiedziała, że je napisała – dopóty nie możemy uznać, że maszyny mogą równać się z mózgiem. Żaden mechanizm nie potrafi odczuwać (a nie tylko wysyłać sygnatury, bo to po prostu sztuczka) zadowolenia (...), smaku itd. (Turing 1995: 284).

Także Searle zwraca uwagę na to, że maszyna operuje wyłącznie symbolami, a jako taka nie może niczego wiedzieć czy rozumieć, a zatem nie można jej przypisywać intencjonalności. Maszyny syntaktyczne po prostu nie mają i nie mogą przejawiać ludzkiej intencjonalności, jaką w naszym przypadku „wytwarza" mózg.

Sytuacja w chińskim pokoju jest nieco paradoksalna, ponieważ – jak mógłby powiedzieć Searle – kiedy przyglądamy się i badamy wnikliwie każdą część chińskiego pokoju, poznajemy jej funkcje, które z grubsza biorąc polegają na przesyłaniu dalej informacji, jaką ta część otrzymała na swoim wejściu. Funkcja całego chińskiego pokoju sprowadza się do przetwarzania informacji. Oprócz „wysyłania sygnałów" nie znajdujemy w żadnej z jego części ani śladu rozumienia, świadomości czy intencjonalności, nie wspominając o semantyce i znaczeniu. Jak zatem całość, to znaczy cały chiński pokój, mógłby przejawiać takie cechy choćby w szczątkowej postaci? A chiński pokój jest wszak funkcjonalnie równoważny modelowi umysłu opartemu na maszynie Turinga.

Problem jednak w tym, że szukając intencjonalności, rozumienia i świadomości w poszczególnych częściach, podsystemach i prostych elementach całego systemu, nie znajdziemy jej także w mózgu. To tak, jakby spodziewać się, że poszczególne neurony okażą się intencjonalne i świadome. Żaden element sieci neuronowej czy istoty szarej, ogólnie mówiąc, żadna część mózgu i układu nerwowego brana z osobna nie posiada intencjonalności czy świadomości. Intencjonalność stanów umysłu/mózgu czy świadomość jest natomiast wynikiem aktywności i współdziałania bardzo wielu struktur mózgowych. Paradok-

salność chińskiego pokoju znika, jeśli pomyślimy o tym, jak działa mózg, który jest (wedle słów Searle'a) przyczyną występowania stanów intencjonalnych. Trudność stanowiska Searle'a polega na tym, że w pewnym sensie chiński pokój jest bardzo ogólnym, niemniej trafnym metaforycznym opisem działania nie tylko maszyny Turinga, ale i mózgu. Z tego punktu widzenia można powiedzieć, że funkcjonalna równoważność zachodzi między chińskim pokojem i maszyną Turinga, ale też między chińskim pokojem i działającym mózgiem. Dlatego chiński pokój nie wydaje się dobrym pomysłem na krytykę funkcjonalizmu komputacyjnego.

Podobnie argumentuje A. Hodges, wskazując na poglądy samego Turinga:

> Ciekawe jest, nawet jeśli to kwestia anachroniczna, co powiedziałby Turing na Searle'owską przypowieść o chińskim pokoju, która z kolei sama była czymś w rodzaju odpowiedzi na inscenizację gry w udawanie. Searle zakłada, (1) że istnieje algorytm przekładu chińskiego na angielski, (2) że algorytm ten jest wykonywany nie przez maszynę, lecz jednego człowieka lub wielu ludzi w pokoju, pracujących bezrefleksyjnie. Dokonuje się wówczas przekładu z chińskiego, ale żaden z tłumaczy nie dysponuje jakąkolwiek wiedzą ani rozumieniem – powstaje paradoks. Pogląd Turinga jest, jak sądzę, taki, że gdyby się to osiągnęło, nie byłby to żaden paradoks, lecz po prostu prawdziwy stan rzeczy ubrany w formę dramatyczną. Byłby to obraz mechanizmu działania mózgu, w którym neurony nie dysponują indywidualnie rozumieniem, a jednak cały system, jak się wydaje, dysponuje nim; liczyłby się wówczas tylko ten zewnętrzny efekt. (...) Można by pójść dalej mówiąc, że sytuacja w Bletchley Park była niezwykle zbliżona do chińskiego pokoju, gdyż gwoli zachowania tajemnicy ćwiczono ludzi w wykonywaniu algorytmów kryptograficznych bez znajomości ich przeznaczenia. Prawdopodobnie ta właśnie wizja, trafnych sądów wypływających z bezrefleksyjnych rachunków, była pozytywną inspiracją dla Turinga w jego wyobrażaniu sobie mechanicznej inteligencji około roku 1941. Osnową poglądów Turinga jest to, że pewność co do istnienia świadomości jest iluzją – w znaczeniu nieredukowalnej własności poddającej się modelowaniu obliczeniowemu – pojawiającą się i dającą się ostatecznie wytłumaczyć przez odwołanie do jej złożoności (Hodges 1998: 98).

2. W młynie Leibniza[3]

Eksperyment Searle'a miał pokazać, że na sposób kognitywistyczny nie możemy wyjaśnić intencjonalności umysłu. Eksperyment Leibniza ma z kolei pokazać, że na sposób mechanicystyczny nie można wyjaśnić fenomenu świadomości. Biorąc pod uwagę te dwa pomysły, należałoby uznać, że naturalistyczny program wyjaśnienia umysłu, w tym świadomości fenomenalnej, jest skazany na porażkę. Najpierw krótko scharakteryzuję pomysł Leibniza, a potem spróbuję na tej mapie sporu umiejscowić stanowisko Chalmersa.

[3] Eksperyment ten szerzej analizuję gdzie indziej (*Świadomość w młynie*, „Przegląd Filozoficzny" 2 (74) 2010). Tu krótko streszczę moje rozumowanie.

Jak powiedziałem, eksperyment myślowy Leibniza ma pokazywać niedostępność poznawczą świadomości dla funkcjonalno-naturalistycznego podejścia. Jest on bez wątpienia bardzo sugestywny i na pierwszy rzut oka wiarygodny. Jednak po nieco bliższym oglądzie kontrowersyjne staje się to, czy spełnia on swoje zadanie. W 17 paragrafie swojej *Monadologii* Leibniz mówi:

> Należy wszakże przyznać, że postrzeżenie i to, co do niego należy, nie da się wytłumaczyć racjami mechanicznymi. To znaczy przez kształty i ruchy. Przypuściwszy zaś, że istnieje maszyna, której budowa pozwala, aby myślała, czuła, miewała postrzeżenia, będzie można pomyśleć ją, z zachowaniem tych samych proporcji, tak powiększoną, aby można do niej wejść jak do młyna. Założywszy to, odnaleźlibyśmy wewnątrz przy zwiedzaniu jej tylko części, które popychają się wzajemnie, nigdy jednak nic, co tłumaczyłoby postrzeżenie. (...) trzeba szukać tego właśnie w substancji prostej, a nie w rzeczy złożonej, czy też w machinie.

Czym jest ta „maszyna" czy „młyn"? O czym mówi Leibniz? Ważne jest tu to, że rzeczona maszyna posiada elementarne cechy *świadomości*. Czy jest to coś w rodzaju sztucznego systemu poznawczego, który myśli, czuje, postrzega itp.? Stanowisko Leibniza w tej kwestii jest przecież w dużym stopniu zbieżne z krytyką idei silnej wersji sztucznej inteligencji.

A może to mózg? Dlaczego nie mielibyśmy zinterpretować Leibnizjańskiego młyna jako wnętrza mózgu? Mam nadzieję, że pod pewnym warunkiem taka zamiana może być uprawniona. Jest nim podtrzymywanie tezy o mechaniczności umysłu, a mówiąc bardziej współcześnie, o jego funkcjonalno-komputacyjnej strukturze działania. Zamieniamy więc młyn Leibniza na mózg, a towarzyszyć temu będzie teza, że jest to maszyna biologiczna. Samego zaś filozofa, który wchodzi do młyna i przygląda się działaniu jego mechanizmów, możemy zinterpretować jako zespół badaczy prowadzących badania nad poznawczymi funkcjami mózgu, którzy posługują się metodami nauk o poznawaniu ze szczególnym uwzględnieniem neurobiologii i psychologii poznawczej. Mówiąc w terminologii Leibniza, szukają oni czysto mechanicznego wyjaśnienia. Idąc za nim, powinniśmy dojść do wniosku, że nawet jeśli mózg jest maszyną i jeśli w jakiś sposób wytwarza świadomość, nie uda się nam wyjaśnić tych procesów w fizykalno-funkcjonalny sposób.

Dlaczego jednak mamy się zgodzić na konkluzję Leibniza? Nie jest ona wcale oczywista i jego wniosek wydaje się nieuprawniony. Bo kto wie, może byśmy znaleźli. Trzeba jednak wiedzieć, czego szukać, i dysponować odpowiednimi narzędziami. Poza tym, jeśli tego nie znajdziemy w mózgowym młynie, jak wyjaśnimy rolę procesów neurobiologicznych w świadomości? Cudem? Równie dobrze możemy postawić na taką strategię: jeżeli w młynie nie znajdujemy nic innego, jak tylko jego części skonstruowane i zorganizowane w pewien sposób, oraz widzimy działanie tego systemu i zakładamy przynaj-

mniej możliwość, że maszyna ta w jakiś sposób „wytwarza" czy jest odpowiedzialna za powstawanie pewnego zjawiska, to rozsądne wydaje się rozważenie tego, że za powstanie tego zjawiska odpowiada budowa, organizacja i działanie tej maszynerii.

Przytoczone rozumowanie Leibniza można interpretować dwojako. Po pierwsze, jako twierdzenie epistemologiczne. Przy tej słabszej interpretacji mówi ono tyle, że dysponując jedynie naturalistycznymi narzędziami nauki mechanicystycznej – dziś powiedzielibyśmy o naukach poznawczych – nie możemy i nie będziemy mogli wyjaśnić fenomenu świadomości i, szerzej, stanów i własności mentalnych. Przy drugiej, silniejszej interpretacji przybiera ono postać mocnego twierdzenia metafizycznego, które można zinterpretować tak: świadomość jest czymś niefizycznym, nie jest też wytwarzana przez funkcjonowanie złożonych procesów fizycznych, na przykład neurobiologicznych. Inaczej mówiąc, świadomość nie jest skutkiem, emergentną własnością działania funkcjonalnie zorganizowanych systemów i podsystemów fizycznych – maszynerii młyna. Należy ją raczej określić mianem czegoś w rodzaju niefizycznej „substancji prostej", której istnienia i własności nie da się odkryć ani wyjaśnić badając działanie złożonej organizacji funkcjonalnej maszyny – mózgowego młyna.

Sądzę, że z dwojga złego większe nadzieje na spójną teorię umysłu rokuje ponowne zbadanie maszynerii młyna niż próba wyjaśnienia przyczynowego związku niefizycznej „substancji prostej" z zachowaniem, o ile chce się uniknąć epifenomenalizmu świadomości (czego nie unika na przykład Chalmers).

Jak powiedziałem, ten eksperyment myślowy wbrew pozorom nie spełnia swojego zadania, tj. nie pokazuje, że po pierwsze, nie ma trzecioosobowego dostępu do świadomości, i po drugie, że nie jest ona czymś fizycznym, a przynajmniej czymś, czego nauki fizykalne nie mogą uczynić przedmiotem wyjaśnienia. Podobnie rzecz się ma z innymi eksperymentami tego typu, jak „chiński pokój" Searle'a. Nie znaczy to jednak, że teoria umysłu i świadomości, jakiej mają służyć, jest fałszywa. Jeśli jednak świadomość jest czymś w rodzaju „substancji prostej", to jak wyjaśnić jej przyczynowe własności, na przykład możliwość przyczynowego wpływania na procesy decyzyjne i wybory strategii zachowania czynione przez podmioty oraz na samo zachowanie organizmów w fizycznym świecie? Jak wspomniałem, alternatywę stanowi epifenomenalny charakter świadomości, ale wówczas odmawiamy jej rzeczywistego wpływu na działanie. Istnieją ważne powody – o których na końcu – by nie iść tą drogą.

Podsumowując: co łączy eksperymenty Searle'a i Leibniza?

Wedle Searle'a: na sposób komputacyjno-maszynowy nie można wyjaśnić semantyki, intencjonalności i rozumienia. To, co widzimy, to tylko syntaktyczne operacje maszyny na symbolach.

A według Leibniza: na sposób mechanicystyczny nie można wyjaśnić fenomenu świadomości; to, co widzimy, to tylko ruchy maszynerii młyna.

Można na to odpowiedzieć: po pierwsze, intencjonalność, a też świadomość, jest wynikiem działania całego systemu, jakim jest chiński pokój, młyn lub mózg. Nie zaś tej czy innej jego części.

I po drugie: chiński pokój i młyn są modelami działania mózgu, więc jak najbardziej możemy mieć nadzieję i spodziewać się wyjaśnienia fenomenu świadomości, „qualiów", ze strony nauk o poznawaniu.

Jak na tej mapie sporu przedstawia się stanowisko Chalmersa?

3. Rozwiązanie Chalmersa

W rozdziale *Świadomego umysłu* poświęconym sztucznej inteligencji z pewnym zdziwieniem spostrzegłem, że po pierwsze, autor na swój własny, oryginalny sposób zbija argumentację i wnioski Searle'a[4], a po drugie, wysuwa tezę

[4] „Dlatego rozsądny jest wniosek, że ostateczny system ma dokładnie te same przeżycia świadome co oryginalny system. Gdyby układ nerwowy wytwarzał przeżycia jasnej czerwieni, tak czyniłby też system demonów, a więc i sieć kartek papieru za pośrednictwem demona. Oczywiście jednak ten ostatni przypadek jest tylko kopią systemu w chińskim pokoju. Dlatego też daliśmy pozytywny powód, by sądzić, że system naprawdę ma przeżycia świadome, takie jak rozumienie chińskiego lub przeżywanie czerwieni. (...) kartki w pokoju nie są jedynie stosem formalnych symboli. Stanowią one konkretny system dynamiczny z organizacją przyczynową odpowiadającą bezpośrednio organizacji pierwotnego mózgu. Zaciemnia to wolne tempo, które kojarzy nam się z taką manipulacją symbolami, podobnie jak obecność demona nimi manipulującego, ale to właśnie konkretna dynamika kartek papieru wytwarza przeżycia świadome. Po drugie, rola demona jest całkowicie drugorzędna. Ciekawa dynamika przyczynowa zachodzi między kartkami papieru, które odpowiadają neuronom w pierwotnym przypadku. Demon po prostu działa jak pewnego rodzaju mediator przyczynowy. (...) Fakt, że demon jest świadomym podmiotem działającym, może skłaniać do przypuszczenia, że jeśli gdziekolwiek są przeżycia systemu, to w demonie; ale w rzeczywistości świadomość demona jest całkowicie bez znaczenia dla funkcjonowania systemu. Zadanie demona mogłoby być wykonywane przez prostą tabelę przeglądową. Kluczowym elementem systemu jest dynamika zachodząca między symbolami. (...) Gdy tylko jednak spojrzymy poza obrazy przywoływane na myśl przez obecność nieistotnego demona i przez powolne tempo przekładania symboli, to zobaczymy, że dynamika przyczynowa w pokoju dokładnie odzwierciedla dynamikę przyczynową w czaszce. Dzięki temu tak niewiarygodne nie wydaje się przypuszczenie, że system wytwarza przeżycia. Niektórzy mogą sądzić, że ponieważ mój argument opiera się na powielaniu organizacji na poziomie neuronów w mózgu, wykazuje [prawdziwość] tylko słabej formy AI, która jest związana ściśle z biologią. Byłoby to jednak niedocenianiem siły argumentu. Program symulacji mózgu służy jedynie jako pierwszy krok. Skoro już wiemy, że jeden program może wytwarzać umysł nawet wtedy, gdy jest implementowany w stylu chińskiego pokoju, siła zasadniczego argumentu Searle'a całkowicie znika: wiemy, że demon i papier w chińskim pokoju mogą faktycznie urzeczywistniać niezależny umysł. Otwiera się wówczas droga do całego szeregu programów, które mogą potencjalnie generować przeżycia świadome. Wielkość tego szeregu jest kwestią otwartą, ale chiński pokój nie jest przeszkodą" (Chalmers 2010: 265).

o możliwości implementacji tudzież wytworzenia świadomości w sztucznych systemach. Co więcej, teza ta dotyczy także świadomości fenomenalnej. Pisze:

> Wniosek jest taki, że wydaje się, że co do zasady nie ma barier dla ambicji sztucznej inteligencji. (...) mamy dobre, pozytywne powody, aby sądzić, że implementacja stosownego obliczenia przyniesie ze sobą przeżycia świadome. Perspektywy świadomości maszynowej są więc dobre w zasadzie, jeśli jeszcze nie w praktyce. (...) Niezależnie od tego, jaka organizacja przyczynowa okaże się kluczowa dla poznania i świadomości, możemy oczekiwać, że podejście obliczeniowe będzie w stanie ją uchwycić. Można nawet argumentować, że to ta elastyczność kryje się za często przywoływaną uniwersalnością systemów obliczeniowych. Zwolennicy sztucznej inteligencji nie są ograniczeni do przyjmowania jednego rodzaju obliczeń, który mógłby wystarczyć do umysłowości; teza AI jest tak wiarygodna właśnie dlatego, że klasa systemów obliczeniowych jest tak duża. Pozostaje więc kwestią otwartą, jaka konkretnie klasa obliczeń wystarczy do powielenia umysłowości człowieka; mamy jednak dobre powody, by sądzić, że klasa ta nie jest pusta (Chalmers 2010: 270).

Wspomniałem już, że stanowisko Chalmersa wydało mi się dziwne. Oto, jak odpowiedziałby on na następujące pytania:

(a) Czy świadomość dostępu daje się badać w sposób naukowy? Odpowiedź: tak.

(b) Czy świadomość zjawiskowa powstaje jako wynik działania fizycznych systemów? Odpowiedź: tak.

(c) Czy zatem poddaje się ona badaniu w taki sam sposób jak świadomość dostępu? Odpowiedź: nie.

Interesująca jest odpowiedź (c). Krótko mówiąc, specyfika poglądu Chalmersa i powód mojego zdziwienia polega na tym, że twierdzi on, iż świadomość – także zjawiskowa – jest możliwa w sztucznych systemach, ale nie można jej scharakteryzować funkcjonalistycznie. Dlaczego nie? Jak sądzę, odpowiedź Chalmersa opierałaby się na pewnym argumencie modalnym. Pisze on:

> Argumenty w związku z tym nie uzasadniają silnej postaci funkcjonalizmu, zgodnie z którą organizacja funkcjonalna jest *konstytutywna* dla przeżycia świadomego; ale uzasadniają słabszą postać, która nazwałem *funkcjonalizmem nieredukcyjnym*, zgodnie z którym organizacja funkcjonalna wystarcza dla przeżycia świadomego z koniecznością przyrodniczą. Z tego punktu widzenia przeżycia świadome są wyznaczane przez organizację funkcjonalną, ale nie muszą być redukowalne do organizacji funkcjonalnej.
>
> W każdym razie wniosek jest wciąż mocny. Zasada niezmienności organizacyjnej mówi nam, że co do zasady systemy poznawcze realizowane w ośrodkach wszelkiego rodzaju mogą być świadome. W szczególności wniosek daje silne wsparcie dla ambicji naukowców w dziedzinie sztucznej inteligencji (...). Jeśli funkcjonalizm nieredukcyjny jest poprawny, nieredukowalność świadomości nie stanowi bariery dla ewentualnej budowy świadomych urządzeń obliczeniowych (Chalmers 2010: 228).

3.1. Co możemy zrobić? Propozycja pragmatyczna

Czy jest to dobry powód do zawieszenia naturalistycznego programu wyjaśnienia świadomości fenomenalnej? Sądzę, że nie. A nawet, nie rozstrzygając argumentu Chalmersa, z powodzeniem możemy poprzestać na „konieczności przyrodniczej" bez ambicji do logicznej, a skoro „przeżycia świadome są wyznaczane przez organizację funkcjonalną", choć – jak twierdzi Chalmers – nie redukują się do niej, spróbować funkcjonalistycznie wyjaśnić rolę świadomości dostępu i świadomości fenomenalnej. Dlaczegóż porzucać nadzieję na kognitywistyczne wyjaśnienie świadomości? Odchodzę tu od poglądu Chalmersa. Sądzę bowiem, że jeżeli inżynierom sztucznej inteligencji udałoby się skonstruować „świadome urządzenie obliczeniowe", dałoby to nam mocne podstawy do odpowiedzi na pytanie, czym jest świadomość, także fenomenalna.

4. Ranga i przyczynowość qualiów

Na zakończenie chciałbym postawić dwa pytania, które od dawna mnie nurtują.

(a) Różni autorzy, w tym Chalmers, podkreślają, że qualia są czymś najważniejszym lub najbardziej tajemniczym w życiu psychicznym. Otóż, bez cienia ironii, chciałbym powiedzieć, że nie rozumiem, dlaczego właśnie qualiom przyznaje się tak wysoką rangę. Równie dobrze można uznać to za kwestię perspektywy czy gustu. Na przykład dla biologa pewnie nie będą najważniejsze.

(b) Dlaczego qualiom odmawia się własności przyczynowych? Przecież są odczuwane, postrzegane i ujmowane pojęciowo? Czy te odczucia grają jakąś rolę w życiu psychicznym? Jeśli tak, to qualia miałyby swoje nieepifenomenalne miejsce w łańcuchu przyczynowym. Czy zatem byłaby to przyczynowość fizyczna? Czemu nie, jeśli mogłyby one wpływać na zachowanie? Jednak przyjęło się zaprzeczać takiej możliwości. Wydaje mi się to nieco dziwne, jeśli wziąć pod uwagę to, że jakość doznania wpływa na chęć jego powtórzenia albo unikania w przyszłości – na przykład wstręt lub uczucie przyjemności. Przeżycie doznania „żółte, ciepłe, przyjemne" – to znaczy w pewnej sytuacji: właśnie zaczęły się wakacje – wpłynie na poszukiwanie takiego doznania w przyszłości, tj. na unikanie pracy. Podobnie jakość doznania zjedzenia garści ziemi wpłynęła na zachowanie wielu z nas, tj. tych, którzy to zrobili w dzieciństwie; reszta powinna nam w tej kwestii zaufać. Krótko mówiąc, sądzę, że to, *jakie* jest moje doznanie, wpływa na moje zachowanie. A skoro tak, qualia, wbrew epifenomenalistom, wchodzą w interakcje przyczynowe z innymi stanami umysłu i grają swoją rolę w zachowaniu. Zaś jako takie stają się przedmiotem badań nauk o poznawaniu i zachowaniu. Zapewne zwolennicy nieredukowalności qualiów powiedzą, że nie trafiam tu w interesujące filozoficznie jakości doznań. Czym zatem one są? W co trafiam?

Bibliografia

Chalmers, David (2010), *Świadomy umysł*, przeł. M. Miłkowski, PWN.
Hodges, Andrew (1998), *Turing*, przeł. J. Nowotniak, Amber.
Leibniz, Wilhelm Gottfried (1995), *Monadologia*, przeł. S. Cichowicz, Wydawnictwo Comer.
Searle, John (1995), *Umysł, mózg i nauka*, przeł. J. Bobryk, PWN.
Searle, John (1999), *Umysł na nowo odkryty*, przeł. T. Baszniak, PIW.
Turing, Alan (1995), *Maszyna licząca a inteligencja*, przeł. M. Szczubiałka, (w:) B. Chwedeńczuk (red.), *Filozofia umysłu*, Fundacja Aletheia.

Streszczenie

Biorąc za dobrą monetę eksperymenty myślowe Leibniza i Searle'a, można argumentować, że stosując strategię kognitywistyczną nie uda się wyjaśnić intencjonalności i świadomości, w tym świadomości fenomenalnej. Oba eksperymenty mają podobną strukturę. Jednak ich konkluzywność jest dyskusyjna. Jak na tym tle przedstawia się stanowisko Chalmersa? Dość dziwnie. Chalmers z jednej strony stanowczo odpiera zarzuty Searle'a, z drugiej zaś zdaje się akceptować wniosek Leibniza o tym, że świadomości nie można wyjaśnić w mechanicystyczny sposób. Czy to, że przeżycia świadome są wyznaczane przez organizację funkcjonalną, choć nie redukują się do niej, jak twierdzi Chalmers, przekreśla nadzieje na funkcjonalistyczne wyjaśnienie roli świadomości fenomenalnej?