

JOHN R. SEARLE

CZY KOMPUTERY MOGĄ MYŚLEĆ?

(fragment z: *Umysł, mózg i nauka*, PWN, Warszawa 1995)

W poprzednim rozdziale przedstawiłem, przynajmniej w zarysie, rozwiązanie tzw. problemu psychofizycznego. Choć nie znamy szczegółów dotyczących zasad działania mózgu, wiemy jednak dostatecznie wiele, by mieć ogólny obraz stosunków pomiędzy procesami mózgowymi i procesami psychicznymi (umysłowymi). Procesy umysłowe są skutkiem działania elementów mózgu. Jednocześnie realizowane są w strukturze zbudowanej z tych właśnie elementów. Myślę, że to rozwiązanie problemu zgodne jest ze standardowym podejściem do zjawisk biologicznych. Niewątpliwie, oparty na naszej znajomości zjawisk w świecie, zdroworozsądkowy sposób rozwiązania problemu. Jednak jest to również punkt widzenia zdecydowanej mniejszości. W filozofii, psychologii i teorii sztucznej inteligencji przeważa stanowisko podkreślające analogie pomiędzy działaniem ludzkiego mózgu i cyfrowego komputera. Zgodnie ze skrajną wersją tego stanowiska, mózg jest rodzajem komputera, zaś umysł rodzajem komputerowego programu. Można ten pogląd streścić (określam go mianem "silnego podejścia w teorii sztucznej inteligencji" lub "silną SI") mówiąc, że **umysł jest tym dla mózgu, czym program dla komputera**.

Konsekwencją takiego poglądu jest uznanie, iż nie ma niczego zasadniczo biologicznego w ludzkim umyśle. Mózg staje się tu jedną z mnóstwa maszyn liczących, w których mogą być realizowane programy stwarzające ludzką inteligencję. Z tego punktu widzenia, każdy system fizyczny odpowiednio zaprogramowany, mający odpowiednie wejścia i wyjścia, będzie obdarzony umysłem dokładnie takim samym jak umysł Czytelnika i mój. Zatem jeśli zrobilibyśmy komputer ze starych puszek po piwie, napędzanych wiatrakami, jeśli zaprogramowalibyśmy go odpowiednio, byłby on obdarzony umysłem. Problem nie polega na tym, że w świetle naszej wiedzy taki system **mógłby** myśleć i odczuwać, raczej na tym, że **musiałby** on myśleć i czuć, gdyż **jedynym** warunkiem myślenia i odczuwania jest zainstalowanie odpowiedniego programu.

Większość zwolenników tego stanowiska sądzi, że nie stworzyliśmy jeszcze programów będących umysłami. Panuje jednak między nimi zgoda, że jest to tylko kwestia czasu, aż do chwili, gdy specjaliści od maszyn cyfrowych i sztucznej inteligencji stworzą odpowiednie urządzenia "hardware'owe" i programy, które będą ekwiwalentami ludzkich mózgów i umysłów. Wtedy powstaną sztuczne mózgi i umysły całkowicie porównywalne z ludzkimi mózgami i umysłami.

Ludzie nie związani z badaniami w dziedzinie sztucznej inteligencji są zdumieni, odkrywając, że ktokolwiek może reprezentować tego typu stanowisko. Przeto, zanim poddam je krytyce, pozwolę sobie przytoczyć kilka przykładów niedawnych wypowiedzi ludzi pracujących w tej dziedzinie. Herbert Simon z Uniwersytetu Carnegie-Mellon twierdzi, że maszyny cyfrowe już myślą w sensie dosłownym. Nie trzeba czekać na stworzenie w przyszłości myślących maszyn, gdyż istniejące komputery cyfrowe już obdarzone są, równoważną mojej lub Czytelnika, zdolnością myślenia. Nie do uwierzenia! Przez wieki filozofowie zastanawiali się czy

maszyny mogą, czy nie mogą myśleć i nagle odkrywamy, że mamy myślące maszyny w Carnegie-Mellon. Kolega Simona, Alan Newell, utrzymuje, że odkryliśmy właśnie (zwracam uwagę, że Newell mówi "odkryliśmy", a nie "hipotetycznie założyliśmy", czy "dopuszczamy możliwość"), iż inteligencja jest niczym innym jak manipulowaniem symbolami przez systemy fizyczne; nie ma żadnej konieczności jej związku z jakimś szczególnym rodzajem biologicznej lub fizycznej struktury. Każdy system, mogący w odpowiedni sposób manipulować fizycznie ucieleśnionymi symbolami, jest obdarzony zdolnościami intelektualnymi, będącymi tym samym, czym ludzka inteligencja. Zarówno Simon, jak Newell w dobrej wierze podkreślają, że nie ma nic metaforycznego w powyższym sformułowaniu, że trzeba rozumieć je dosłownie. Cytuje się Freemana Dysona, który miał stwierdzić, że komputery mają nad nami przewagę z ewolucyjnego punktu widzenia. Jeśli świadomość jest jedynie efektem działania pewnych programów, w przypadku komputerów programy te realizowane są w urządzeniach mających, w podlegającym entropii świecie, więcej szans na przetrwanie niż istoty, takie jak my, zbudowane z wilgotnych i nietrwałych substancji. Marvin Minsky z Massachusetts Institute of Technology twierdzi, że następna generacja komputerów ma być tak inteligentna, że będziemy "szczęśliwi, jeśli maszyny zechcą zatrzymać nas w domach w charakterze domowych zwierzątek". Ciągle moim ulubionym przykładem przesady w poglądach na temat maszyn cyfrowych pozostaje twierdzenie Johna McCarthy'ego, twórcy terminu "sztuczna inteligencja". McCarthy twierdzi, że nawet "o urządzeniach tak prostych, jak termostaty, możemy powiedzieć, że mają przekonania" i oczywiście, według niego, o każdej maszynie zdolnej rozwiązywać problemy można powiedzieć, że ma ona jakieś mniemania. Podziwiam śmiałość McCarthy'ego. Zapytałem go kiedyś: "jakie przekonania ma twój termostat?". Odpowiedział: "Mój termostat ma trzy sądy — tu jest zbyt gorąco, tu jest zbyt zimno, temperatura jest odpowiednia". Jako filozofowi, poglądy te podobają mi się z bardzo prostych powodów. Inaczej niż większość filozoficznych twierdzeń są one jasne i pozwalają na prostą i zdecydowaną ich falsyfikację. Rozdział ten poświęcony jest właśnie ich falsyfikacji.

Istota falsyfikacji tych poglądów jest niezależna od tego lub innego stanu rozwoju technologii maszyn cyfrowych. Należy to podkreślić, gdyż zwykle istnieje pokusa, by na rozwiązanie problemu poczekać, aż pojawi się jakiś nie stworzony dotąd technologiczny cud. Jednakże istotne punkty naszej falsyfikacji nie mają związku z jakimkolwiek etapem rozwoju technologicznego. Mają natomiast związek z definicją komputera cyfrowego oraz z tym, czym jest komputer.

W idei komputera cyfrowego istotne jest to, że jego operacje mogą być scharakteryzowane **czysto formalnie**, to znaczy, że możemy opisać jego działanie, krok po kroku, za pomocą abstrakcyjnych symboli, na przykład sekwencji zer i jedynek wydrukowanych na taśmie. "Zasady" działania maszyny cyfrowej decydują o tym, że jeśli maszyna jest w określonym stanie, a na taśmie zapisany jest określony symbol, wtedy maszyna wykona pewną operację, taką jak wymazanie symbolu lub wydrukowanie innego i wejdzie w stan następny, przesuając taśmę o jedno pole. Symbole tego typu nie mają znaczenia semantycznego, nie odnoszą się do niczego. Mogą one być scharakteryzowane czysto formalnie, ze względu na ich związki syntaktyczne. Zera i jedynki są po prostu cyframi i nie muszą nawet oznaczać liczb. Oczywiście dzięki maszynom cyfrowym nabierają swojej mocy. To samo elektroniczne urządzenie, jeśli jest odpowiednio skonstruowane, może być użyte do zrealizowania

nieokreślonej liczby programów. Zaś ten sam program może być wykonany przez wiele typów maszyn. Przyjęcie tezy, że programy można opisywać czysto formalnie lub syntaktycznie ma zgubne skutki dla poglądu, że proces umysłowy i proces wykonywania jakiegoś programu są czymś identycznym.

Łatwo można wykazać, dlaczego tak jest. Mieć umysł, to coś więcej niż realizować formalne czy syntaktyczne operacje. Nasze stany umysłowe, na mocy definicji, mają zawsze **jakąś treść**. Jeśli myślę o Kansas City, życzyłbym sobie wypić szklankę zimnego piwa, bądź zastanawiam się czy będzie spadek notowań giełdowych, to w każdym wypadku mój stan umysłowy, niezależnie od tego, jakie formalne właściwości mu przypiszemy, ma jakieś psychiczne treści. To znaczy, że nawet jeśli moje myśli 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 – ma semantykę**. Powód, dla którego komputerowy program nie może być umysłem, jest prosty: komputerowy program ma cechy syntaktyczne, zaś umysły ma ją coś więcej niż syntaktyka. Umysły są semantyczne w tym sensie, że poza strukturą formalną mają jeszcze jakieś treści.

By zilustrować tę tezę, proponuję rozważenie następującego obmyślonego przeze mnie eksperymentu myślowego. Wyobraźmy sobie, że zespół programistów napisał program, który jest w stanie symulować rozumienie języka chińskiego. Zatem, na przykład, jeśli komputerowi zadaje się pytanie w tym języku, może porównać je ze swoją pamięcią lub bazą danych i wyprodukować w języku chińskim odpowiedź na zadane pytanie. Załóżmy, dla celów dyskusji, że odpowiedzi te są tak dobre jak odpowiedzi prawdziwej osoby, której ojczystym językiem jest język chiński. Czy w takiej sytuacji komputer rozumie język chiński? Czy rozumie go dokładnie tak, jak użytkownicy języka chińskiego rozumieją swój ojczysty język? Wyobraźmy sobie teraz, że ktoś z nas jest zamknięty w pokoju, i że w pokoju tym jest szereg pudeł wypełnionych znakami z języka chińskiego. Załóżmy, że osoba ta, podobnie jak autor myślowego eksperymentu, nie zna języka chińskiego, otrzymała jednak napisaną w jej ojczystym języku książkę z regułami manipulowania znakami języka chińskiego. Reguły te opisują używanie symboli w sposób **czysto formalny**, opisują manipulowanie nimi w sposób syntaktyczny, nie semantyczny. Mogą mieć postać: "Wybierz ten znak z podwójnym zakrętasem z pudła numer jeden i umieść go za znakiem z dwoma zawijasami z pudła numer dwa". Przyjmijmy teraz, że w pokoju pojawiają się jakieś nowe symbole, a osoba w nim siedząca otrzymuje instrukcje, jakie chińskie symbole ma wysłać z pokoju w odpowiedzi na te, które się pojawiły. Załóżmy, że siedzący w pokoju nie wie, iż symbole przesyłane do pokoju przez ludzi z zewnątrz nazywane są przez nich pytaniami, zaś symbole, które siedzący w pokoju wysyła na zewnątrz, nazywane są odpowiedziami na te pytania. Przyjmijmy poza tym, że programiści napisali na tyle dobry program, zaś siedząca w pokoju osoba jest do tego stopnia sprawna w manipulowaniu symbolami, że jej odpowiedzi są **nieodróżnialne** od odpowiedzi osoby faktycznie znającej język chiński. Zatem jakaś osoba zamknięta jest w pokoju, w którym wybiera symbole chińskie i wysyła je w odpowiedzi na inne przychodzące do pokoju chińskie symbole. W sytuacji, jaką tu opisałem, **nie ma możliwości**, by w wyniku takiej manipulacji formalnie zdefiniowanymi symbolami naprawdę nauczyć się języka chińskiego.

Meritum mojego myślowego eksperymentu jest następujące: realizując taki formalny komputerowy program, z punktu widzenia obserwatora z zewnątrz, zachowujemy się dokładnie tak, **jakbyśmy** rozumieli język chiński, faktycznie jednak **nie rozumiemy ani jednego słowa** z tego języka. Jeśli wykonanie komputerowego programu symulującego rozumienie języka chińskiego nie jest wystarczające dla nas, byśmy ten język rozumieli, nie może być wystarczające także dla maszyny liczącej. Wyjaśnienie tego stanu rzeczy jest całkiem proste. Jeśli w takiej sytuacji człowiek nie rozumie chińskiego, to również inne "komputery" go nie rozumieją, gdyż żaden komputer cyfrowy nie może, dzięki temu że zrealizuje program, uzyskać zdolności, której nie uzyskuje człowiek. To wszystko, co ma komputer i co my mamy siedząc w chińskim pokoju, jest jedynie formalnym programem, pozwalającym na manipulowanie nie zinterpretowanymi chińskimi symbolami. Powtórzmy to: **komputer opanowuje tylko syntaktykę, nie zaś semantykę**. Cały morał naszej historii o chińskim pokoju przypomina nam to, co już wiemy. Rozumienie języka – a z pewnością i posiadanie wszelkich innych stanów umysłowych – wymaga czegoś więcej niż garści symboli formalnych. Konieczna tu jest jeszcze interpretacja, lub **znaczenie** związane z tymi symbolami.

Komputer cyfrowy zaś, na mocy definicji, nie może mieć niczego więcej poza formalnymi symbolami, gdyż operacje komputerowe, jak to już powiedziano, definiuje się w terminach maszynowej zdolności do wykonywania programów. Programy natomiast mają czysto formalną charakterystykę, to znaczy nie mają semantycznych treści. Możemy się przekonać o sile tego argumentu, jeśli porównamy sytuację, w której otrzymujemy pytania i udzielamy odpowiedzi w naszym ojczystym języku, z sytuacją, w której pytają nas i odpowiadamy w języku, w którym nie znamy znaczenia ani jednego słowa. Wyobraźmy sobie, że jesteśmy w chińskim pokoju, gdzie zadają nam w naszym ojczystym języku pytania dotyczące takich spraw, jak nasz wiek, czy fakty z naszej biografii, i że odpowiadamy na te pytania. Jaka jest różnica pomiędzy posługiwaniem się językiem chiński a posługiwaniem się językiem polskim? Otóż, jeśli ktoś nie zna języka chińskiego, a zna język polski, różnica jest oczywista. **Rozumiemy** pytania zadane po polsku, bo wyrażone są one symbolami, których **znaczenie** jest nam znane. Analogicznie, odpowiadając po polsku, produkujemy symbole, które mają dla nas znaczenie. W przypadku języka chińskiego zaś nic takiego nie występuje. W przypadku chińskiego posługujemy się po prostu formalnymi symbolami, zgodnie z komputerowym programem, jednak nie potrafimy przypisać tym elementom żadnego znaczenia.

Zarówno specjaliści od sztucznej inteligencji, jak i psychologowie oraz filozofowie rozmaicie odpowiadali na moje argumenty. Wszystkie te odpowiedzi miały wspólną cechę: wszystkie były chybione. Stanie się oczywiste, dlaczego były one takie, jeśli oprzemy nasze rozumowanie na prostej logicznej prawdzie, stwierdzającej, że **syntaktyka jako taka nie wystarcza dla semantyki**, i że komputery w związku z tym, że są komputerami, z definicji, mają tylko syntaktykę. Chciałbym to wyjaśnić, rozważając kilka kontrargumentów, jakie najczęściej mi przedstawiano.

W odpowiedzi na przykład chińskiego pokoju niektórzy ludzie stwierdzają, że cały system rozumie język chiński. Argument jest następujący: choć osoba manipulująca symbolami w pokoju nie rozumie języka chińskiego, to jest ona tylko centralnym procesorem całego komputerowego systemu. Twierdzi się tu, że cały system

(obejmujący pokój, pudła pełne symboli, księgę zawierająca reguły, oraz szereg innych elementów, jakie na system się składają) jest tym, co rozumie język chiński. Ten jednak punkt widzenia może być podważony w taki sam sposób, jak to już uczyniono. Także system w całości nie może przejść od syntaktyki do semantyki. Tak jak osoba w pokoju, jako centralny procesor, nie może wyobrazić sobie znaczenia jakiegokolwiek symbolu, tak również nie może tego uczynić system w całości.

Częsta jest także reakcja, proponująca, byśmy wyobrazili sobie, że wspomniany program rozumienia języka chińskiego znalazł się we wnętrzu robota. Gdyby robot mógł się poruszać, działać w świecie i podlegać jego działaniu, czyż nie mogłoby mu to zagwarantować zdolności rozumienia języka chińskiego? Jednakże nieubłagany charakter różnic pomiędzy syntaktyką i semantyką omija i ten intelektualny manewr. Jak długo, na mocy założenia, robot ma tylko komputer zamiast mózgu, to pomimo tego, że zachowuje się on tak jakby rozumiał język chiński, tak długo nie ma on możliwości przejścia od syntaktyki do semantyki języka chińskiego. Można to dostrzec wyobrażając sobie, że to ja zamknięty w chińskim pokoju jestem komputerem. W czasie robota umieszczono chiński pokój ze mną w środku, gdzie manipuluje chińskimi symbolami, nie zdając sobie sprawy z tego, że niektóre z nich poznają za pomocą telewizyjnej kamery przytwierdzonej do głowy robota, inne zaś wysyłam na zewnątrz poruszając jego kończynami. Jak długo używam tylko formalnego programu komputerowego, tak długo nie mam możliwości, by przyporządkować znaczenie któremukolwiek z chińskich symboli. A to, że robot jest tu zaangażowany w moje interakcje ze światem otaczającym, nie pomoże mi w odnalezieniu znaczenia chińskich symboli, jeśli nie znajdę innego sposobu na zrozumienie zachodzących wokół faktów. Przypuśćmy, że robot podnosi hamburger i to powoduje pojawienie się w pokoju ideogramu hamburgera. Jak długo mam tylko symbol, bez wiedzy o tym, co on powoduje lub jak się pojawia, tak długo nie mam sposobu poznania jego znaczenia. Interakcje pomiędzy robotem i resztą świata są tu bez znaczenia, jeśli te interakcje nie zostaną odzwierciedlone w tym, czy w innym umyśle. Nie mogą zostać jednak odzwierciedlone, jeśli "umysł" składa się z sieci czysto formalnych, syntaktycznych operacji.

Istotne jest to, byśmy rozumieli, co się twierdzi, czego zaś się nie twierdzi w przedstawionym tu przez mnie ciągu argumentów. Powiedzmy, że zadajemy postawione na wstępie pytanie: "Czy maszyna może myśleć?" Oczywiście, w jakimś sensie, wszyscy jesteśmy maszynami. To, co jest w naszych głowach, można traktować jako rodzaj maszyny organicznej. Z pewnością też wszyscy możemy myśleć. Przy pewnym sposobie rozumienia terminu "maszyna" – dokładnie wtedy, gdy "maszyna" rozumiana jest jako pewien fizyczny system zdolny do wykonywania określonych operacji – wszyscy jesteśmy maszynami i wszyscy myślimy. Zatem, banalizując, istnieją maszyny, które myślą. Nic tego typu pytanie nas jednak dręczy. Spróbujmy je zatem przeformułować. Czy artefakty mogą myśleć? Czy mogą myśleć maszyny stworzone przez człowieka? Przyjmijmy, że skonstruowaliśmy maszynę, która ani na molekułę nie różni się od istoty ludzkiej. Zatem, jeśli można skopiować przyczynę, można skopiować i skutek. Odpowiedź na pytanie jest w zasadzie banalna. Gdybyśmy mogli zbudować maszynę, która miałaby dokładnie taką samą strukturę jak istota ludzka, wtedy prawdopodobnie maszyna ta byłaby w stanie myśleć. Oczywiście byłaby wtedy surogatem człowieka. Spróbujmy inaczej postawić problem.

Nie stawiamy pytania: "Czy maszyna myśli?", ani "Czy artefakt myśli?" Pytanie brzmi: "**Czy komputer cyfrowy myśli?**" Podkreślamy po raz kolejny, musimy być ostrożni w interpretacji pytania. Z matematycznego punktu widzenia, wszystko, co może być opisane w taki sam sposób jak komputer, jest komputerem. A to dlatego, że to coś może być opisane jako urządzenie realizujące lub mające komputerowy program. Wulgaryzując można powiedzieć, że pióro leżące przede mną na stole da się opisać jako komputer cyfrowy. Tak się jednak składa, że komputer ten ma wyjątkowo nieciekawym program. Poleca on bowiem: "Pozostań na miejscu". Tak rozumując, wszystko jest komputerem, skoro może być opisane **jakby** posiadało program komputerowy. Zatem nasze pytanie, powtórzmy to po raz kolejny, zyskuje banalną odpowiedź. Oczywiście nasze mózgi są komputerami, jeżeli realizują pewną liczbę programów komputerowych. Jednocześnie nasze mózgi bez wątpienia myślą. Odpowiedź na pytanie jest banalna. Nie jest to jednak pytanie, jakie staraliśmy się postawić. Pytanie, jakie staraliśmy się postawić, brzmi: "Czy tak zdefiniowany komputer myśli?" Co można sformułować: "Czy zaopatrzenie w program lub wykonywanie odpowiedniego komputerowego programu, z odpowiednimi danymi na wejściach i wyjściach, **wystarcza** dla myślenia, lub **tworzy** myślenie?" Na to pytanie, inaczej niż na poprzednie, jasna odpowiedź brzmi: "nie". Udzielamy odpowiedzi "nie" z powodów już przedstawionych, czyli dlatego, że komputerowy program jest określony czysto syntaktycznie. Myślenie nie ogranicza się jednak do kwestii manipulowania pozbawionymi znaczenia symbolami, uwikłane są w nie treści semantyczne. Te semantyczne treści określamy mianem "znaczenia".

Warto jeszcze raz podkreślić, że nie mówimy tu o takiej lub innej fazie rozwoju technologii elektronicznej. Rozumowanie nasze jest niezależne od nadchodzących zdumiewających postępów w technice komputerowej. Nie ma też ono związku z rozróżnieniem procesorów działających sekwencyjnie i działających równolegle, ani z wielkością programu, ani z szybkością maszynowych operacji, ani z tym, czy komputery wchodzą w interakcje z otoczeniem, a także z wynalezieniem robotów. Zwykle przeceniamy aktualny postęp technologiczny, ale i bez tego rozwój techniki komputerowej jest całkiem zauważalny, a mamy zupełnie rozsądne powody, by oczekiwać, iż jeszcze większy postęp dokona się w przyszłości. Bez wątpienia będziemy mogli, znacznie lepiej niż możemy obecnie, i znacznie lepiej niż mogliśmy w przeszłości, symulować za pomocą maszyn ludzkie zachowania. Chodzi jednak oto, że wszystkie te **symulacje nie mają żadnego znaczenia**, gdy mówimy o posiadaniu stanów umysłowych, lub o posiadaniu umysłu; bez względu na to jak dobra jest technologia komputerowa i jak szybkie są obliczenia przez komputer wykonywane. Jeżeli coś faktycznie jest komputerem, jego operacje zdeterminowane są syntaktyką, podczas gdy świadomość, myślenie, doznania i emocje, oraz inne zjawiska tego typu **wymagają** czegoś poza samą syntaktyką. Są to własności, których komputer nic może skopiować ani powtórzyć, niezależnie od tego, jak wielkie byłyby jego zdolności w symulowaniu tych cech. Kluczowym rozróżnieniem jest tu **rozzróznienie symulacji robienia czegoś i autentycznego robienia tego samego**. Symulowanie czegoś nie jest robieniem ani powtarzaniem tego.

Wszystko co dotąd przedstawiłem, polega na stworzeniu podstaw do zrozumienia, że cytaty, jakimi się posłużyłem na wstępie snutyh tu rozważań, są faktycznie tak niedorzeczne, jak na to wyglądają. Niemniej jednak w naszych

rozważaniach pozostaje jeszcze jedno pytanie warte przemyślenia. Brzmi ono: "Dlaczego ludziom przychodzi czasem na myśl, że komputery mogłyby myśleć, mieć doznania i emocje oraz inne tego typu stany?" Zauważmy przede wszystkim, że możemy stworzyć symulację komputerową wszystkiego, co da się opisać formalnie. Możemy zatem, komputerowo symulować przepływ pieniędzy w ekonomii brytyjskiej lub strukturę podziału władzy w Partii Pracy. Możemy stworzyć symulację komputerową burz w kraju, w którym mieszkamy, lub pożarów magazynów we Wschodnim Londynie. W każdym z tych przypadków nikt jednak nie przypuszcza, że komputerowa symulacja stwarza realnie istniejące rzeczy; nikt nic zakłada, że komputerowa symulacja ulewy zmoczy nas, ani że komputerowa symulacja pożaru spali nam mieszkanie. Dlaczego wobec tego spotkać można osoby, które w dobrej wierze zakładają, iż symulacja procesów psychicznych zawiera autentyczne procesy psychiczne? W gruncie rzeczy nie znam odpowiedzi na to pytanie, gdyż tego typu idea, mówiąc szczerze, od początku wydaje mi się szalona. Mogę jednak przedstawić garść przypuszczeń.

Po pierwsze, rozważając naturę umysłu, możemy być kuszeni rozwiązaniami typu behawiorystycznego. Można sądzić, że skoro system zachowuje się **tak, jakby** rozumiał język chiński, to **faktycznie** rozumie on ten język. Jednakże przyjmując argumenty zawarte w idei chińskiego pokoju, odrzuciliśmy ten sposób behawiorystycznego myślenia. Inne przypuszczenie, nasuwające się wielu ludziom, polega na uznaniu, że umysł nie jest częścią biologicznej rzeczywistości, że nie należy do świata naturalnego. Stanowisko silnej wersji sztucznej inteligencji opiera się na przekonaniu, że umysł jest czymś czysto formalnym, że tak czy inaczej, nie może on być traktowany jak konkretny rezultat biologicznych procesów, podobny do innych rezultatów zjawisk biologicznych. Mówiąc krótko, jest to w naszej dyskusji przykład pozostałości myślenia dualistycznego. obrońcy sztucznej inteligencji są przekonani, że umysł jest czymś więcej niż częścią świata biologicznego, że może być scharakteryzowany wyłącznie formalnie. Paradoks polega tu na tym, że literatura poświęcona sztucznej inteligencji jest przepełniona gromami rzucanymi na "dualizm", w istocie rzeczy jednak teza silnej wersji sztucznej inteligencji oparta jest na myśleniu dualistycznym. Opiera się na odrzuceniu idei, że umysł jest tak naturalnym zjawiskiem jak inne zjawiska w świecie.

Chcę zakończyć rozdział, zestawiając ze sobą tezy tego i poprzedniego rozdziału. W obu wypadkach tezy mogą być wyrażone w sposób bardzo prosty. W istocie prawdopodobnie zamierzam je sformułować w sposób nazbyt surowy. Jeśli jednak zestawimy je razem, zyskamy, jak sądzę, solidną koncepcję relacji pomiędzy umysłami, mózgami i komputerami. Rozumowanie ma prostą strukturę, tak że każdy sam może ocenić jego wartość. Pierwsza przesłanka brzmi:

1. Mózgi są przyczynami umysłów.

Brzmi to faktycznie zbyt prosto. Rozumiem przez to, że procesy umysłowe, które, jak sądzimy, konstytuują umysł, są w całości skutkami procesów zachodzących wewnątrz mózgu. Pozwólmy sobie jednak na proste sformułowania, skróćmy to do czterech słów — mózgi są przyczynami umysłów. Jest to jednak faktyczna cecha zjawisk w świecie. Zapiszmy teraz twierdzenie drugie.

2. Syntaktyka nie wystarcza dla semantyki.

Twierdzenie jest prawdziwe na mocy treści pojęć. Wyraża ono po prostu pojęciowe rozróżnienie tego, co jest czysto formalne oraz tego, co ma treść. Teraz do twierdzenia, mówiącego, że mózgi są przyczynami umysłów i twierdzenia, że syntaktyka nie wystarcza dla semantyki dodajmy twierdzenia trzecie i czwarte.

3. Program komputerowy całkowicie określa jego formalna lub syntaktyczna struktura.

Przyjmuję to, co jest prawdą, na mocy definicji oraz jest częścią znaczenia tego, co rozumiemy przez program komputerowy.

4. Umysły zawierają treści psychiczne, mówiąc dokładniej – treści semantyczne.

Przyjmuję, że to także jest oczywistą, faktyczną cechą funkcjonowania umysłu. Moje myśli, sądy i pragnienia **odnoszą się do czegoś**, dotyczą stanów i wydarzeń w świecie; jest tak dlatego, że ich treści kierują je ku tym stanom i wydarzeniom w świecie.

Możemy teraz z powyższych przesłanek wyprowadzić pierwszy wniosek wynikający w sposób oczywisty z przesłanek 2, 3 i 4.

Wniosek 1. *Nic ma takiego programu komputerowego, który sam w sobie wyposażamy system w umysł. Mówiąc krótko, programy nie są umysłami, ani same w sobie nie wystarczą dla powstania umysłu.*

Jest to bardzo istotna konkluzja, oznacza ona, że projekt stworzeniu umysłu wyłącznie na podstawie programu jest z góry skazany na niepowodzenie. Ważne jest, bym jeszcze raz podkreślił, że nie ma to związku ani zdany stanem rozwoju komputerowej technologii, ani ze stopniem złożoności programu. Jest to formalny lub logiczny wniosek wypływający z założeń, na których przyjęcie zgodzili się wszyscy (bądź prawie wszyscy) zaangażowani w dyskusję. Nawet najgorętsi entuzjaści sztucznej inteligencji są zgodni co do tego, że jest faktem biologicznym to, iż procesy mózgowe powodują procesy umysłowe; są oni zgodni także w tym, że programy są określone czysto formalnie. Jednakże, gdy zestawimy te dwa twierdzenia z innymi znanymi nam rzeczami, nieuchronnie wyciągamy wniosek, że projekty zwolenników silnej wersji sztucznej inteligencji nie mogą być zrealizowane. Skoro już ustaliliśmy założenia, rozważmy, jakie jeszcze wnioski możemy wyciągnąć. Druga konkluzja brzmi:

Wniosek 2. *Czynności mózgu ograniczone tylko do realizowania programu komputerowego nie wystarczą, by funkcjonowanie mózgu doprowadziło do powstania umysłu.*

Ten drugi wniosek jest koniunkcją pierwszej przesłanki i pierwszego wniosku. Czyli z tego, że umysł jest efektem działania mózgu, i że wykonywanie programu nie daje tych efektów, wynika, że zrealizowanie programu komputerowego przez mózg samo w sobie nie wystarczy, by działanie mózgu doprowadziło do powstania umysłu. Jest to, jak myślę, ważny wniosek, gdyż w konsekwencji wynika z niego, że mózg nie jest komputerem cyfrowym, a może raczej jest czymś więcej. Przekonaliśmy się wcześniej, że wszystko co może być opisane jak komputer cyfrowy jest, upraszczając, komputerem, zaś mózg nie jest tu wyjątkiem. W naszej konkluzji ważne jest jednak to,

że obliczeniowe możliwości mózgu nie wystarczą do wyjaśnienia, jak działanie mózgu powoduje stany umysłowe. Tak czy inaczej winno to na nas sprawić wrażenie oczywistego naukowego wniosku, gdyż przypomina nam tylko znany fakt, że mózgi są biologicznymi maszynami; znaczenie ma tu ich biologiczność. Jak stwierdza wiele osób w dziedzinie sztucznej inteligencji, nie jest przypadkową cechą umysłu, że jest on realizowany przez ludzkie mózgi. Z pierwszego założenia możemy teraz wyciągnąć trzeci wniosek:

Wniosek 3. *Cokolwiek, co mogłoby być przyczyną umysłu, musiałoby mieć moc oddziaływania przyczynowego porównywalną z możliwościami mózgu.*

Trzeci wniosek jest banalną konsekwencją pierwszego założenia. Jest to mniej więcej tak, jak byśmy powiedzieli, że jeśli benzynowy silnik mojego samochodu pozwala mu poruszać się z prędkością 120 km/h, to każdy silnik dieslowski mogący uczynić to samo, musiałby na wyjściu mieć moc porównywalną z mocą mojego silnika benzynowego. Oczywiście, jakieś inne jeszcze systemy, używając innych chemicznych i biochemicznych cech swojego mózgu, mogą powodować istnienie umysłu. Może się okazać, że są istoty na innej planecie lub w innym systemie gwiazdowym, które mają stany umysłowe oparte na całkiem innej niż nasza biochemii. Przypuśćmy, że pojawili się na Ziemi Marsjanie a my wywnioskowaliśmy, że mają oni stany umysłowe. Załóżmy jednak, że kiedy otwarto ich głowy, okazało się, że wszyscy mają w środku zielony śluz. Czyli, jeśli zielony śluz działając powodowałby świadomość i całą resztę psychicznego życia, miałby moc przyczynową porównywalną z mocą ludzkiego mózgu. Teraz z naszego pierwszego wniosku, stwierdzającego kiedy program jest czymś niewystarczającym i z trzeciego wniosku, mówiącego o systemach mających moce przyczynowe porównywalne z mózgowymi, wynika wprost czwarta konkluzja:

Wniosek 4. *Wyposażenie jakiegoś zbudowanego przez nas artefaktu w program komputerowy nie wystarcza, by miał on stany umysłowe porównywalne z ludzkimi. Artefakt taki powinien oczywiście mieć zdolność przyczynowego oddziaływania porównywalną z możliwością ludzkiego mózgu.*

Sądzę, że ostateczny wynik dyskusji przypominał nam coś, co już od dawna wiemy: że stany umysłowe są zjawiskiem biologicznym. Świadomość, intencjonalność, subiektywność, moc przyczynowego oddziaływania umysłu, wszystko to należy do dziejów biologicznego życia, razem ze wzrostem, rozmnażaniem, wydzielaniem żółci czy trawieniem.