

Michał Sachnowski

Zaskakująca kariera pewnego wynalazku.

Kilka uwag w kwestii myślenia maszyn

Spór o tzw. sztuczną inteligencję, czyli myślenie maszyn, pojawił się wraz z pierwszymi maszynami liczącymi; w skali historii filozofii jest on niewątpliwie stosunkowo świeżej daty. Pojawił się jako odprysk dyskutowanego od zawsze problemu psychofizycznego: niektóre teorie, wypracowane odnośnie do umysłu ludzkiego, nadają się idealnie do roli metodologicznego zaplecza konstruowanych w naszym stuleciu cyfrowych artefaktów. Część przynajmniej teoretyków i konstruktorów tak utrzymuje. Uzyskane tą drogą uogólnienia znajdują z kolei wtórne zastosowania filozoficzne. Jak to się odbywa? Wystarczy powiedzieć, że zwolennicy dowolnej współczesnej teorii zaangażowanej w problem psychofizyczny chcą automatycznie rozstrzygać spór o sztuczną inteligencję, ba — są w gruncie rzeczy przekonani, że dzieje się to mimochodem. I odwrotnie: kontrowersja dotycząca sztucznej inteligencji przeradza się w dysputę na temat inteligencji w ogóle, na temat „znaczenia”, „rozumienia” i kryteriów dystrybucji tych pojęć. Jeżeli posuwamy się w tym kierunku, problem psychofizyczny spada do rangi jednej z wielu kwestii pomocniczych, a daniem głównym staje się teoria poznania.

Zasygnalizowaliśmy w ten sposób wiele rzeczy na raz. Przyjrzyjmy im się teraz w kolejnych przybliżeniach. Zwolennicy eliminacyjnych teorii umysłu, przekonani o intersubiektywności i komunikowalności wszelkiej wartościowej wiedzy, przychylnie na ogół odnoszą się do programów badawczych oraz pojęciowych propozycji znaczeniowych zgłaszanych przez twórców maszyn cyfrowych. Idea maszyny i programu obliczeniowego umożliwia włączenie umysłu do naukowego obrazu świata. Wystarczy zdefiniować umysł w kategoriach przetwarzania danych, predyspozycji funkcjonalnych albo wręcz obserwowalnych reakcji zachowaniowych, aby — niczym za dotknięciem czarodziejskiej różdżki — przełamać uporczywy prymat, jaki na tym obszarze utrzymywały zarówno byty psychiczne, jak i związane z nimi biologiczne układy nerwowe. „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 umysł.”¹ Oto bezpośrednia korzyść, jaką odnosi nauka: przy takim rozumieniu zjawisk umysłowych (w szczególności inteligencji) najlepszym sposobem ich badania, a nawet kryterium istnienia jest behawioralny test Turinga.

Przeciwników takiego stawiania sprawy ochrzczijmy umownie mianem „mentalistów”. W tej grupie — przekonanej o realnym istnieniu i nieredukowalności tradycyjnych, introspekcyjnych stanów psychicznych — dominuje nastawienie zgoła odwrotne. Wyraża się przeważnie w lekceważeniu deklaracji filozofujących informatyków oraz negowaniu całej behawioralno-funkcjonalistycznej metodologii. Jak przebiega taka negacja? Poważne traktowanie Turingowskiej teorii inteligencji — powiada się — stosowalnej rzekomo na równi do człowieka i jego cyfrowych artefaktów, a w konsekwencji także do nieograniczonej ilości rozmaitych sztucznych wytworów i nawet procesów naturalnych, prowadzi do kompletnego rozmycia używanych pojęć i jednej przynajmniej poważnej luki w opisie świata. Rozmycie polega na wymieszaniu tego, co „umysłowe”, „inteligentne” czy „świadome” z całą resztą zjawisk, dotąd za takie nie uznawanych. Luka w opisie jest z kolei efektem usunięcia świadomego podmiotu, dysponującego znaczącymi coś dlań doświadczeniami i przeżywającego jakoś własny system wiedzy.

Powyższa różnica zdań upewni Czytelnika, że kłótnie filozoficzne idą najczęściej o lepsze z kontrowersjami metafizycznymi. Nasz temat jest pod tym względem wręcz wzorcowy: redukcjoniści i zwolennicy pojęciowej tolerancji, niechętni idei „dochodzenia prawdy” i rozstrzygania sporów w jakimś bezwzględnym, nierelatywnym trybie, uważają kwestie inteligencji, świadomości, myślenia czy umysłu w ogóle za ściśle konwencjonalne, pozbawione jednego określonego sensu. Dlaczego? Ponieważ w grę wchodzi pojęcia zasadniczo nieostre, pozbawione kryteriów empirycznych, których znaczenie każdy użytkownik języka może próbować doprecyzować na własną rękę. Możemy z powodzeniem uważać maszyny za inteligentne i obdarzone umysłem, jeśli tylko służy to teoretycznym celom — na przykład pozwala trafnie przewidywać i spójnie opisać zachowania maszyn, a następnie włączyć ten opis w szerszy kontekst pewnej hipotezy lub teorii. Z drugiej strony możemy bez mrugnienia okiem powyższemu przeczyć, o ile stoimy w obliczu konieczności trwałego odróżnienia ludzi od ich wytworów — jak choćby w przypadkach nakładania odpowiedzialności cywilnej albo przyznawania praw obywatelskich. Nie ma wobec tego sensu pytać, które stanowisko jest prawdziwe, pytać trzeba tylko o to, które lepiej służy konkretnym założeniom i spełnia określone wymagania.

¹ J.R. Searle, *Umysł, mózg i nauka*, przeł. J. Bobryk, Warszawa 1995, str. 26.

Stanowisko przeciwne odwołuje się do fundamentalnych zasad naturalizmu i empiryzmu — konwencje można i należy rozpatrywać od strony faktów, odbierając im tym samym przywilej dowolności. Trzeba określić na podstawie stanu faktycznego, kto posiada umysł i świadomość, a co ich nie posiada: nie jest to żadna ponadempiryczna metafizyka ani też kwestia językowa, lecz zwykłe dociekanie „umeblowania świata”. Obecność umysłu i inteligencji przyjmujemy albo odrzucamy zawsze z ważnych powodów rzeczowych. Świat nie kryje więc przed nami nie znanego dotąd potencjału intelektualnego i nie jest tak, że nowe interpretacje pojęć filozoficznych wykażą kiedyś niezbitcie, iż maszyny jednak są krypto-ludźmi, obdarzonymi świadomymi stanami umysłowymi, a my wydamy im wówczas potulnie dowody osobiste i świadectwa udziałowe PPP. Z punktu widzenia naturalizmu konwencjonalizm metodologiczny i pojęciowy wydaje się przekonywający jeszcze w odniesieniu do abstraktów czy syntaktycznie zdefiniowanych symboli, ale wątpliwy już w takich przypadkach, jak sprawa myślenia maszyn, gdzie chodzi o badanie fragmentów świata zewnętrznego. Odnosnie zaś do nas samych jakiegokolwiek konwencji — może poza czysto nazewniczymi — są całkowicie nie do przyjęcia. Uznawanie wielu „równoprawnych” (w rzeczywistości zaś nie do pogodzenia, na przykład nawzajem sprzecznych) rozstrzygnięć jednego problemu wolno nam rozumieć jako wycofanie się z dyskusji.

Idźmy jednak dalej. Niezależnie od debaty filozoficznej obserwujemy burzliwy proces postępu technologicznego. Stosunkowo prymitywne początkowo konstrukcje i programy weszły już w fazę bezprecedensowych, działających na wyobraźnię możliwości. Pod wieloma względami przekroczyły przy tym „nienaruszalne” granice, ustanawiane wcześniej dekretemi sceptyków. Komputery wykonują coraz bardziej skomplikowane zadania, pełnią też odpowiedzialne funkcje, wymagające nie tylko obliczeniowej precyzji — co wystarczało jeszcze w przypadku szachowej strategii — ale także umiejętnego segregowania danych oraz wyboru określonych dróg postępowania. Maszyny potrafią uczyć się na popełnionych błędach, syntetyzować mowę, naśladować ludzki sposób myślenia; co tu zresztą wymieniać — widzi to każdy, kogo interesuje choć trochę stan współczesnej cywilizacji.

Filozofia, by tak rzec, nie znosi próżni. Na zapleczu techniki powstały natychmiast optymistyczne teorie Krzemowej Doliny. Że, po pierwsze — jeśli nawet konstruowane obecnie artefakty nie są doskonałe (w szczególności mają kłopoty z zadowalającym zdaniem niektórych wersji kluczowego testu Turinga), to nic nie stoi na przeszkodzie, by w niedalekiej przyszłości nie miały prześcignąć człowieka, z natury niedokładnego i mocno ograniczonego w swych rozwojowych możliwościach. W czym prześcignąć? Ano, we wszystkich dziedzinach umysłowej aktywności. Że, po wtóre — jeżeli tylko teoretycznie istnieje taka możliwość, to jest to wyłącznie kwestia czasu. I że, na koniec — maszyny cyfrowe nie tylko opanują stopniowo wszelkie odmiany ludzkich zachowań: stanowią one już teraz wyjaśnienie i ostateczny model funkcjonowania człowieka. Myślą, postrzegają i odczuwają w sensie zupełnie dosłownym. „Większość zwolenników tego stano-

wiska 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.”² Trudno przecenić doniosłość tej prognozy. „W mocnej wersji SI [sztucznej inteligencji — przyp. M.S.], skoro zaprogramowane komputery mają stany poznawcze, programy komputerowe nie są ledwie narzędziami pozwalającymi nam sprawdzać wyjaśnienia psychologiczne; one same są raczej wyjaśnieniami.”³

Wiemy już sporo na temat perspektyw maszyny Turinga, pozostaje pytanie, co z nami — z Czytelnikiem i piszącym te słowa? Umysł ludzki okazuje się specyficzną odmianą informatycznego software'u, przypadkowo realizowaną na takim akurat, a nie innym podłożu biologicznym. Mózg, jako się rzekło, jest rodzajem maszyny przetwarzającej dane: tak się składa, że przetwarza z reguły tylko jeden umysłowy program (chyba, że jest niesprawny), ale tak bynajmniej być nie musi. Uprzywilejowana pozycja umysłu to przesąd. W tym miejscu sygnalizujemy, co następuje: po pierwsze, wszystkie cytowane pomysły ogłaszane są całkiem serio, bez dystansu, bez cienia przerośnię. Na takich założeniach poważne zespoły opierają poważne projekty badawcze. Po drugie zaś, jeżeli powyższa teoria wygląda Czytelnikowi na jakąś zmodyfikowaną wersję dualizmu, ma on świętą rację. Funkcjonalizm, jakkolwiek oparty na nowoczesnym w formie teście Turinga, jest w gruncie rzeczy dualistyczny niczym filozofia Kartezjusza; na miejscu szyszynki mamy procesor, ot i cała różnica.

Przyjrzyjmy się teraz tak pomyślanej koncepcji. Maszynę cyfrową Turing definiuje jako odpowiednik „komputera ludzkiego”: jest to „maszyna przeznaczona do wykonywania wszelkich takich operacji, jakich dokonać mógłby ludzki komputer. Przez „ludzki komputer” rozumiemy tu człowieka, którego sposób postępowania wyznaczony jest przez z góry przypisane reguły”⁴. Ale człowiek i maszyna to tylko pierwsze z brzegu przykłady systemów przetwarzających informacje. Każdy system mający uporządkowane wejścia i wyjścia jest równoważny innym takim systemom, choćby składał się z makulatury oraz sznurka do pakowania. W skrajnych przypadkach jest to oczywiście ucieleśnienie inteligencji dość ograniczonej; termostat ma, przykładowo, zawsze trzy przekonania, których nigdy nie zmienia: „tu jest za ciepło”, „tu jest za zimno” i „tu jest w sam raz”. Każdy system można jednak wzbogacić o dowolne właściwości, funkcje i predyspozycje umysłowe. W jaki sposób? Zainstalujmy na przykład mechanizm losowy

² J.R. Searle, *Umysły, mózgi i programy*, przeł. B. Chwedeńczuk, w: B. Chwedeńczuk (red.), *Filozofia umysłu*, Warszawa 1995, s. 301.

³ J.R. Searle, *Umysł, mózg i nauka*, dz. cyt., s. 26.

⁴ A.M. Turing, *Maszyna licząca a inteligencja*, przeł. Michał Szczubiałka, w: *Filozofia umysłu*, dz. cyt., s. 274.

w naszej pocziwej lodówce, a zaraz zacznie kierować się wolną wolą przy podejmowaniu decyzji, na przemian chłodząc i rozmrażając bez żadnego uchwytanego sensu.

Trzeba pamiętać, że Turing przygotowuje maszyny do behawioralnych testów, do gry w udawanie. Nie interesuje go abstrakcyjne pytanie, czy mogą myśleć: wraz z nadejściem ery wyrafinowanego i precyzyjnego „udawania”, czyli symulacji zachowań naszego gatunku, nikomu nie będzie przychodziło do głowy wątpić w myślenie maszyn. W tym przypadku, niczym w przysłowiu, nie tyle prawda zwycięży, co wykruszą się jej przeciwnicy. Nie wiemy tylko jeszcze, jak odróżniać stwory obdarzone inteligencją od nieinteligentnych. W tym celu Turing proponuje prościutki test, znany pod jego imieniem. Każdy może sobie łatwo wyobrazić rozmowę, prowadzoną przez jakiś rodzaj zasłony — niech to będzie rozmowa telefoniczna; zdarza nam się w ten sposób komunikować nie znając nawet wyglądu rozmówcy. Reagujemy wtedy jedynie na treść komunikatów językowych. Otóż, gdy na podstawie dowolnie długiej wymiany zdań nie udaje się nam ustalić tożsamości rozmówcy — w szczególności odróżnić maszyny od człowieka — musimy uznać jego inteligencję za równoważną naszej; i nie wolno nam się z tego później wycofać, choćby się miało okazać, że dyskutowaliśmy z szalenie sprawną obliczeniowo stertą żelaznego złomu. Niechaj nikomu nie wygląda to na gołosłowną fantastykę! Już teraz powstają maszyny, zdolne z powodzeniem zdać tak pomyślany egzamin. Daleko im co prawda do ludzkiej wszechstronności — realizują zadania bardzo ściśle, jak gry strategiczne albo tzw. programy ekspertowe. Niemniej, ewolucja sztucznej inteligencji nigdy nie stoi w miejscu, przeciwnie — skokowo nabiera tempa. Już dziś możemy wyspekulować algorytmiczną imitację rozwoju mózgu (od poziomu embrionalnego aż do stadium pełnego ukształtowania). Niemożliwe? Istnieją przecież sztuczne sieci neuronalne, które rozwijają się samoczynnie, uczą, doskonalą, mają własną dynamikę: kiedy zaś proces edukacji da się przełożyć na język programowania, myślenie maszyn stanie się faktem. Maszyny zaczną uczyć się i „udawać” już nie dla potrzeb testów, ale na własny rachunek, wyciągając wnioski z popełnianych błędów i mnożąc sukcesy w społecznym dyskursie.

Powstaje pytanie, czy słabe punkty filozoficznej — *sit venia verbo* — doktryny sztucznej inteligencji są poważnie dyskutowane? I czy są w ogóle dostrzegane? Przyjrzyjmy się teraz, jak mogą wyglądać ewentualne zastrzeżenia. Sam Turing wymienia ich kilka; jak powiada — najczęściej występujących⁵. Po pierwsze: rozwój maszyn cyfrowych napotka liczne a nieprzekraczalne ograniczenia o charakterze teoretycznym — jak twierdzenie Gödla — praktycznym, technicznym lub zgoła finansowym. Po wtóre: maszyny nie posiadają świadomości ani uczuć; potrafią zachowywać się tak, jakby je miały, jednakże imitacja pozostaje imitacją.

⁵ Tamże.

Poza tym maszyny nie mogą myśleć o sobie samych, brakuje im więc tego, co określa się mianem samoświadomości — to po trzecie. Są to oczywiście zaledwie skromne próbki: cała humanistyka roi się od mniej lub bardziej prawdopodobnych sugestii, co właściwie różni człowieka od maszyny — w szczególności cyfrowej. Weźmy chociażby żartobliwą opowieść o ziemskiej podróży Diabła i Pana Boga (*vel* Starego Człowieka): „W tym kraju” — komentuje Diabeł tarapaty, w jakich się znaleźli — „każde wykroczenie, choćby nieszkodliwe, każde najmniejsze naruszenie prawa, każdą anomalię, która osobistym poglądom nadaje niekonwencjonalny charakter, wprowadza się do komputera wyposażonego w pamięć [...]”.

— Do komputera? — niepewnie zapytał Stary Człowiek.

— Do wykonanej rękami człowieka imitacji ludzkiego umysłu.

— Czy jest udana? — Starego Człowieka nagle chwyciło drżenie.

— Są dwie zasadnicze różnice. Komputer nie ma i nie może mieć wyobraźni z tej prostej przyczyny, że gdyby ją kiedyś posiadał, aby dorównać ludziom, stałby się tak niesprawny jak ludzie, a przez to mało dla nich przydatny. Druga różnica polega na tym, że podczas gdy ludzie kiedyś zapominają, komputer nigdy tego nie robi.⁶ Albo inny przykład, z dziedziny, zdawałoby się, przez komputery opanowanej niemal bez reszty — szachów: „styl ich gry jest nadal tak samo nieefektywny i polega na coraz szybszym i szybszym wyliczaniu, co może się zdarzyć z każdą figurą, gdy każdemu jej ruchowi będzie towarzyszył każdy możliwy ruch figur przeciwnika, po którym możliwe jest znowu każde poruszenie własnych, potem przeciwnika i tak dalej. Tym sposobem można wygrać partię, ale takiego gracza trudno nazwać mądrym. Maszyny wciąż pozostają tylko szybkimi idiotami.”⁷ Takie obiekcje można mnożyć w nieskończoność. Problem polega na tym, że większość z nich sformułowana jest nie dość precyzyjnie. Tymczasem nieprecyzyjny zarzut, czegokolwiek by dotyczył, aż prosi się o odprawienie z kwitkiem. Jak to wygląda w praktyce? Jeśli zarzucamy maszynom — przykładowo — brak wolnej woli, Turing proponuje nam egzemplarze zaopatrzone w mechanizm losowy (czymże innym jest wolna wola?). W sprawie samoświadomości argumentuje, że nie jest maszynom obca: te ostatnie dokonują obliczeń dotyczących własnych stanów funkcjonalnych i mogą na tej podstawie prognozować na swój temat (bo czym innym jest samoświadomość?). Należy nam się więc bliższy, krytyczny wgląd w całą sprawę.

Zacznijmy od początku. U podłoża Turingowskiej gry w udawanie leżą pojęcia „przetwarzania informacji” i „przestrzegania reguł”. Przetwarzanie informacji to każdy rodzaj myślenia, czytaj: celowego reagowania na bodźce, jakimi są dane wejściowe. Umysł to nic innego, jak zestaw reguł rządzących przepływem informacji — dlatego droga do jego poznania zawiera się w formalnej, systematycznej strukturze tych reguł: w programie. Cóż to konkretnie znaczy? Między neurofiz-

⁶ P. Ustinov, *Stary Człowiek i Pan Smith*, Warszawa 1990, s. 141

⁷ A. Smith, *Umysł*, tłum. B. Kamiński, Warszawa 1989, s. 316.

jologicznym poziomem opisu a poziomem psychologicznym mieszczą się zjawiska charakteryzowane algorytmicznie, obliczeniowo. One właśnie konstytuują umysł: tylko one są wspólne człowiekowi i maszynie cyfrowej. Zarówno człowiek, jak i komputer stosują pewne reguły obliczeniowe. Te ostatnie nie są, ma się rozumieć, dowolne: istnieje pewien zasób uniwersalnych praw, których zgłębianie ma na celu tzw. nauka kognitywna. Ale skąd się bierze — mamy prawo spytać — przekonanie o istnieniu takich wszechoblicznych reguł? Jego źródeł można doszukać się w filozofii, ale także w socjologii, cybernetyce, informatyce czy wreszcie lingwistyce; historycznie obejmują one okres od Platona do Chomsky'ego — w zależności od tego, jak głęboko chcemy szukać. Każda umysłowa umiejętność — powtórzmy — musi mieć za podstawę nie uświadamianą teorię obliczeniową, „na przykład, skoro nasza zdolność do nauczania się języka, czy zdolność do rozpoznawania twarzy są władzami poznawczymi, którym można przypisać znaczenie — to musi tym samym istnieć jakaś teoria, uwewnętrzniona w naszym mózgu, będąca podłożem tej zdolności”⁸. Jeżeli jednak nie gustujemy w platońskich ideach (uniwersalne a ukryte prawa algorytmiczne), ani też w teoriach dualistycznych (wędrujące umysły — programy komputerowe), nie przyjmujemy na wiarę takiego założenia. Co więcej, śmiało możemy uznać, że komputery i inne maszyny w ogóle nie przestrzegają w swym działaniu reguł w tym sensie, w jakim robią to ludzie. Maszyna funkcjonuje zgodnie ze ściśle określonymi formalnymi procedurami i to wszystko. Program determinuje jej kolejne kroki. Można powiedzieć metaforycznie, że komputer czegoś „przestrzega”, jednakże literalnie rzecz biorąc działa co najwyżej tak, jakby owego czegoś przestrzegał.

Na czym polega różnica? W przypadku człowieka przestrzeganie reguł zasadza się na fakcie, że ich treść — semantyczna zawartość — odgrywa przyczynową rolę w przedsięwziętym działaniu. Komputer nie może wykonać nawet najprostszej operacji nie stosując reguł zawartych w programie; człowiek nie przestrzega żadnej reguły, o ile jej nie zna, choćby przechodził kolejno przez wszystkie fizyczne stany, które odnośna reguła przewiduje. „Zbliżaj się do krawężnika, skręcając maksymalnie w prawo kierownicę, ustaw samochód tak, by jego przednie koła zrownały się z kołami samochodu stojącego przed tobą. Następnie skręcaj kierownicę cały czas maksymalnie w lewo”⁹: oto przykładowa reguła, w tym przypadku dotycząca parkowania przy ograniczonej ilości miejsca. Czytelnik zna ją zapewne doskonale; chodzi jednak o to, że można prowadzić samochód przez wiele lat i nie znać ani tej, ani też żadnej innej reguły parkowania: można parkować na oko albo metodą prób i błędów. I łatwo może się zdarzyć, że dla obserwatora z zewnątrz przypadki te będą nieodróżnialne. Weźmy inny przykład: „znaną cechą ludzkiego widzenia jest to, iż nie widzimy promieni podczerwonych

⁸ J.R. Searle, *Umysł, mózg i nauka*, dz. cyt., s. 43.

⁹ Tamże, s. 41.

i ultrafioletowych. Czyż jest tak dlatego, że mamy uniwersalną gramatykę widzenia, która mówi: «nie dostrzegaj podczerwieni i ultrafioletu»? Oczywiście, że nie, jest tak dlatego, że nasz aparat widzenia nie jest wrażliwy na te właśnie pasma promieniowania elektromagnetycznego. Oczywiście możemy siebie samych opisać tak, **jak gdybyśmy** przestrzegali reguł uniwersalnej gramatyki widzenia, w rzeczywistości jednak nie kierujemy się żadnymi regułami”¹⁰. Zwróćmy uwagę na konkluzję: „formalne cechy zachowania nie wystarczają, by wykazać, że reguła jest przestrzegana. Aby była ona przestrzegana, jej znaczenie musi być jedną z przyczyn zachowania”¹¹.

Wróćmy teraz do wyjściowego pojęcia „przetwarzania informacji”. Metaforycznie i tylko metaforycznie możemy powiedzieć, że człowiek i jego podręczny kalkulator na równi przetwarzają informacje; robią to w zupełnie różnych sensach tego słowa. My angażujemy intencjonalne bądź nieświadome procesy umysłowe, maszyna zaś przechodzi szereg zdefiniowanych kroków. Działania kalkulatora są czysto powierzchowną imitacją procesów umysłowych: wykonując trywialne dodawanie kalkulator nie ma pojęcia, co oznaczają wyświetlane cyfry ani jaki jest sens symbolu „+”. Dlaczego? Ponieważ on w ogóle nic nie wie. „Z tego powodu używamy kalkulatorów. Liczą szybciej i dokładniej niż my, nie musząc przechodzić przez psychiczne wysiłki wykonania tej operacji. W tym znaczeniu, w jakim my przechodzimy przez proces przetwarzania informacji, one tego nie czynią.”¹²

W kontekście teorii sztucznej inteligencji na paradoks zakrawa fakt, że dowolne zjawisko może być opisane tak, jak gdyby zachodziło przetwarzanie informacji. Da się tak ująć proces spływania wody z dachu, reakcje termostatu, fotokomórki, a przy odrobinie wyobraźni funkcjonowanie dużego supermarketu. I analogicznie, wszystko — włącznie z biologicznym układem nerwowym — może zostać przedstawione jako szczególny przypadek maszyny cyfrowej. Są to daleko idące metafory literackie. Funkcjonalizm stał się źródłem ich filozoficznej kariery, uruchomił łańcuchową reakcję nieporozumień; z jakiego powodu? Za pomieszczenie pojęć odpowiada z pewnością założenie, że konstytutywny dla umysłu jest algorytmiczny poziom reguł, w oparciu o które działa empirycznie znana nam psychologia. Komputerowa symulacja zjawisk naturalnych zakłada każdorazowo obliczeniowe przetwarzanie danych; z tego faktu wyprowadzono zgoła nieoczywisty wniosek, iż same naturalne zjawiska wiążą się z takim przetwarzaniem. Od strony formalnej jest to zaledwie analogia. Wszystkie twierdzenia na niej oparte — możemy stwierdzić — mają logiczną moc co najwyżej rozumowania przez analogię; to niewiele. Dlatego na gruncie naturalizmu i empiryzmu „nie ma potrzeby zakładać, że na szczycie neurofizjologicznych struktur znajdują się

¹⁰ Tamże, s. 47.

¹¹ Tamże, s. 43.

¹² Tamże, s. 44.

jeszcze jakieś reguły [...], że poza poziomem stanów umysłowych i poziomem neurofizjologii, zachodzą jeszcze jakieś nieświadome obliczenia”¹³. W praktyce zaś sprawy przedstawiają się jeszcze klarowniej. Umiejętność chodzenia, przykładowo, czyli zdolność utrzymywania równowagi ciała, nie wymaga żadnej teorii geometrycznej, a tylko obecności płynu w uchu środkowym. Płyn, jak wiadomo, nie dokonuje żadnych obliczeń; interpretacja jego położenia wpływa w prostej linii z fizjologii mózgu. Podobnie jest, dajmy na to, w przypadku umiejętności rozpoznawania twarzy: ludzie dokonują tego zupełnie bez wysiłku, zgola nieświadomie. W mózgu istnieje bowiem obszar wyspecjalizowany w tym konkretnym zadaniu, który je realizuje. Neuronalny *hardware* wykonuje po prostu wszystko, co wynika z jego konstrukcyjnych założeń. Gdy tymczasem próbujemy dokonać cyfrowej symulacji tych (a także jakichkolwiek innych) funkcji, stajemy w obliczu monstrualnych kłopotów informatycznych.

Jaki stąd wniosek? Behawioralna „gra w udawanie” jest tylko udawaniem, zabawą w symulację. Przypomina nieco popisy sceniczne: nikt nie przypuszcza, że grający przeistaczają się realnie w Hamleta bądź chromego króla Ryszarda. Cały kłopot w tym, by taki wniosek precyzyjnie uzasadnić. Dokonaliśmy przybliżenia problemu, ale wracamy tu na chwilę do poruszonego wcześniej wątku: jasne uzasadnianie swego stanowiska, w oparciu o konkretne, najlepiej empiryczne przesłanki — oto sprawa kluczowa. Dlaczego? Ponieważ tylko takie stanowisko trudno jest zlekceważyć, a jeszcze trudniej przeinterpretować. Przeciwnicy sztucznej inteligencji zdążyli już zgłosić wobec niej najróżniejsze zastrzeżenia. Cóż z tego, skoro większość z nich miała charakter luźnych pomysłów, przepowiedni i porzekadeł. Ogłoszono na przykład, że maszyny nigdy nie nauczą się rozumieć opowieści, nawet na poziomie czteroletnich dzieci. Specyfika świadomego śledzenia toku opowiadania polega na tym, że słuchacz wie — a właściwie dowiadyuje się — znacznie więcej, niż zostało powiedziane *expressis verbis*. Rozumienie nie jest prostym sumowaniem uzyskanych informacji, ale raczej włączaniem ich w szerszy kontekst ogólnych struktur semantycznych. Przykładem takiej struktury jest tzw. światopogląd. I to jest właśnie dla maszyny nieosiągalne. W odpowiedzi na ten i temu podobne zarzuty powstały, co było do przewidzenia, odpowiednie projekty badawcze. Cel był jasny: przystosowanie maszyn do funkcjonowania w kłopotliwych dla nich sytuacjach, jak choćby wówczas — niech będzie — kiedy ktoś im coś opowiada. Metody okazały się różne: jedne projekty imitować miały dynamikę procesów mózgowych, inne — symulować ewolucyjną drogę organizmów żywych, jaką te ostatnie przeszły od najprostszych tropizmów motorycznych, aż po obecny poziom funkcjonalnej komplikacji; znaleźli się także badacze, którzy postanowili oprzeć sztuczną inteligencję na naśladownictwie ludzkiego trybu nabywania doświadczeń. Komputerowi dostarczyć trzeba najrozmaitszych informacji w ilościach porównywalnych z tymi, jakimi sami nieświadomie dysponuje-

¹³ Tamże.

my. Człowiek podlega programowaniu przez wiele lat; magazynuje dane, utrwała własne reakcje i kształtuje zdolności kojarzenia. W przypadku komputera nie mamy tyle czasu, ale jest on bardzo pojętny i sprawny obliczeniowo.

Zgodnie z tym założeniem umieszczono w elektronicznych rejestrach gigantyczne zasoby danych, miliony symboli, wykresów i opisów, odnotowując nawet fakty dla nas zgoła oczywiste, że na przykład „kiedy Lincoln był w Waszyngtonie, była tam również jego lewa stopa”¹⁴. No i koniec końców osiągnięto oczekiwany stan, w którym maszyny zaczęły udzielać odpowiedzi na wszelkie pytania, dotyczące opowieści. Nie poprzestały nawet na tym, ale samorzutnie podawały informacje i uwagi, jakich nikt od nich nie oczekiwał (co oczywiście dało sceptykom asumpt do twierdzenia, że komputery dalej nic nie rozumieją, a tylko automatycznie demonstrują wszelkie zgromadzone dane). Sprawa wróciła do punktu wyjścia: zwolennicy sztucznej inteligencji uznali, że dowiedziony został dogmatyzm i bezzasadność powyższej krytyki; sceptycy przeciwnie — ogłosili, iż ludzie rozumieją opowiadania mimo wszystko na innych niż komputery zasadach. I że rozumienie jest czymś diametralnie różnym od drukowania gotowych odpowiedzi na wszelkie możliwe pytania.

Kto tu zawinił? Niejasny charakter kontrowersji. Spróbujmy zatem dokonać jeszcze jednego przybliżenia problemu, mając na oku jego precyzację i ewentualne rozwiązanie zarazem. Czytelnik zna już kluczowe pytanie: co jest fundamentalną cechą odróżniającą maszynę cyfrową od człowieka? Charakterystyczny dla komputerów jest fakt zdefiniowania całego ich oprogramowania — czyli umysłu — w postaci sekwencji symboli formalnych. Zasada ta nie zmieniła się od czasów powstania maszyny liczącej, leży bowiem u podstaw samej idei komputera: zależności syntaktyczne wystarczą, by w całości opisać dowolny program. A konsekwencje? „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.”¹⁵ Stany umysłowe posiadają bowiem określoną treść i znaczenie, muszą więc być opisywane w kategoriach semantycznych. Oto zasadnicza różnica. Dla lepszej ilustracji tej tezy przypomnijmy przykład „chińskiego pokoju”, wyjątkowo przejrzysty i sugestywny, a przy tym trafiający w samo sedno sporu. Wyobraźmy sobie człowieka, umieszczonego w izolowanym pomieszczeniu: jego zadaniem jest przejść jak najlepiej test Turinga, w zakresie posługiwania się językiem chińskim. W tym celu zaopatrzony został w zredagowany po polsku (przyjmijmy, że mowa o naszym rodaku) podręcznik, zawierający zasady manipulowania orientalnymi symbolami, których stosy piętrzą się w pokoju. W miarę, jak do pokoju zaczynają napływać chińskie komunikaty, człowiek nasz, posługując się podręcznikiem, selekcjonuje

¹⁴ Cyt. z wywiadu R. Schanka. Por.: J.R. Searle, *Umysł, mózgi i programy*, w: *Filozofia umysłu*, dz. cyt., przyp. 1, s. 301.

¹⁵ J.R. Searle, *Umysł, mózg i nauka*, dz. cyt., s. 28.

i wyrzuca na zewnątrz odpowiednie ideogramy. Co właściwie zachodzi w tym pokoju — zapytajmy — i ile wie na ten temat siedzący wewnątrz? W pokoju, jako się rzekło, realizowany jest test Turinga. Ludzie na zewnątrz — przyjmujemy tym razem, że są to rodowici Chińczycy — interpretują znaki wkładane do środka jako „pytania”, a znaki otrzymywane stamtąd jako „odpowiedzi”. Nasz rodak w pokoju na tyle sprawnie operuje znakami, że Chińczycy uważają go za kogoś co najmniej dwujęzycznego, a może nawet tej samej, co oni narodowości. Tymczasem człowiek w pokoju nic o tym wszystkim nie wie: nie zna sytuacji na zewnątrz, przede wszystkim zaś nie rozumie ani słowa po chińsku, i to niezależnie od długości trwania eksperymentu. Jak to możliwe? Podręcznik, którym się posługuje, niczego mu nie wyjaśnia; ustala tylko syntaktyczne zależności między chińskimi symbolami, na zasadzie — „po zawijasio pionowym typu X, użyj zawijasio poziomego typu Y” (większość z nas nie zna chińskiego, nie będziemy więc zagłębiać się w szczegóły). Człowiek zamknięty w pokoju nie opanuje tą drogą języka chińskiego ani na biegłym, ani nawet na zupełnie podstawowym poziomie. Nie sposób nauczyć się obcego języka operując samymi znakami, czytając — powiedzmy — chińskie zwoje bez słownika ani tłumaczenia: „realizując taki formalny komputerowy program, z punktu widzenia obserwatora z zewnątrz, zachowujemy się dokładnie tak, jakbyśmy rozumieli język chiński, jednocześnie jednak nie znamy ani jednego słowa z tego języka naturalnego. 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”¹⁶.

Czytelnik zna niektóre spośród języków naturalnych; istnieją też takie języki, których kompletnie nie rozumie: łatwo więc może wypraktykować na sobie różnicę, zachodzącą między posługiwaniem się językami z pierwszej i drugiej grupy. Siedząc w „polskim pokoju” rozumielibyśmy wszystko, nie potrzebując instrukcji ani nie znając żadnych formalnie zdefiniowanych operacji, w rodzaju gramatycznych związków rzędu i zgody. W pokoju chińskim nie rozumiemy nic. Trudno przypuszczać, że zarówno w pierwszym, jak i drugim przypadku w grę wchodzi jakiś program formalny, tyle, że raz realizowany sprawnie, innym razem zaś z ubolewania godną opieszałością: „są wyraźne przypadki, do których odnosi się dosłownie termin «rozumienie», i wyraźne przypadki, do których się nie odnosi [...]. Rozumiem opowieści po angielsku, w mniejszym stopniu mogę rozumieć opowieści francuskie, w jeszcze mniejszym niemieckie, a chińskich w ogóle nie rozumiem. Mój samochód natomiast i mój kalkulator nie rozumieją niczego; w ogóle nie mieszczą się w obszarze tej działalności”¹⁷.

Tym razem mamy do czynienia z precyzyjną krytyką, która pod znakiem zapytania stawia sztuczną inteligencję, wraz z całą teorią „komputera ludzkiego”.

¹⁶ Tamże, s. 29.

¹⁷ J.R. Searle, *Umysły, mózgi i programy*, dz. cyt., s. 306.

Wspominaliśmy już o sile funkcjonalistycznych metafor, czas i tę sprawę postawić jasno: przypisujemy maszynom rozumienie i świadomość na tej samej zasadzie, na jakiej imputujemy zwierzętom ludzkie intencje. Jako metafora funkcjonuje to bardzo dobrze, jednakże „biorąc dosłownie, zaprogramowany komputer rozumie to, co rozumie samochód i kalkulator, a mianowicie, dokładnie nic nie rozumie. Rozumienie komputerowe nie jest tylko (tak jak moje rozumienie języka niemieckiego) częściowe czy niepełne; jest zerowe”¹⁸. Nie jest tak, że mózg robi zasadniczo to samo co komputer, tylko sprawniej, na wyższym niż sekwencyjny poziomie komplikacji. Mózg robi zupełnie co innego i dlatego maszyna stanowi słaby model umysłu — podobnie nieadekwatna byłaby w tej roli encyklopedia zaopatrzona w czytnik lub biblioteka z systemem hydraulicznym, otwierającym książki na żądanej stronie. Cóż takiego mózg właściwie robi? Odpowiedź na to pytanie przynależy bardziej do neurofizjologii niż filozofii, ale możemy coś zasugerować: mózg wytwarza znaczenie. Symbole języka polskiego różnią się dla Polaka od chińskich tym, że posiadają znaczenie. Nie jest to kwestia poziomu komplikacji lub stopnia przyswojenia odnośnych informacji. „Znaczenia” nie opisują też definicje formalne, typu: „znaczenie to ogół logicznych konsekwencji”; chodzi tu raczej o intencjonalność i świadomość, o całość umysłowych doświadczeń pozajęzykowych, o indywidualną praktykę życiową, odruchy, wyobrażenia *etc.* Wszystko to jest wynikiem działania mózgu — świadomości nie wytworzy formalny model funkcji umysłowych. I dlatego zaprogramowane maszyny nie są świadome, nie są inteligentne — inteligentni (bo że świadomi, to pewne) bywają co najwyżej ich twórcy.

Jeżeli chcemy podać inne rozwiązanie problemu sztucznej inteligencji, musimy najpierw odrzucić powyższy argument. Co da się w nim zakwestionować? Z punktu widzenia behawioryzmu — wszystko. Ale eksperyment „chińskiego pokoju” pokazuje empirycznie, że test Turinga i cała behawioralna metodologia nie sprawdza się w roli kryterium inteligencji, świadomości i jakichkolwiek innych zjawisk umysłowych. Podobnie z funkcjonalizmem: wyobraźmy sobie na moment, że chiński pokój znajduje się we wnętrzu robota. Robot porusza się swobodnie i wchodzi w normalne, „ludzkie” relacje ze światem zewnętrznym. Człowieka zamkniętego w tym przenośnym pokoju moglibyśmy nazwać „homunkulusem Turinga” — w odróżnieniu od tradycyjnego homunkulusa nie miałby on bowiem pojęcia, co dzieje się w otaczającej pokój rzeczywistości. Przypomnijmy — dysponowałby wyłącznie chińskimi, nie zinterpretowanymi danymi na ten temat. Robotowi takiemu przypisałibyśmy może świadomość, o ile byłby odpowiednio przekonywającym humanoidem, na zasadzie czysto zewnętrznego podobieństwa do nas samych. Gdyby jednak w końcu wyszło na jaw, że jego zachowaniem sterował formalny program — czyli podręcznik, którego używa sterujący robotem homunkulus — upadłaby tym samym hipoteza rozumności całej maszyny.

¹⁸ Tamże, s. 307.

Możemy więc pokusić się o rodzaj podsumowania: „wszystkie te symulacje nie mają żadnego znaczenia, gdy mówimy o posiadaniu stanów umysłowych lub o posiadaniu umysłu [...]. Jeżeli coś faktycznie jest komputerem, jego operacje zdeterminowane są syntaktyką, podczas gdy świadomość, myśli, doznania i emocje, oraz inne zjawiska tego typu wymagają czegoś poza samą syntaktyką. Są to własności, których komputer nie może skopiować ani powtórzyć, niezależnie od tego, jak wielkie byłyby jego zdolności w symulowaniu tych cech”¹⁹. Imitacji — nawet doskonałej, nieodróżnialnej z zewnątrz od oryginału — nie nazwiemy nigdy nową wersją pierwowzoru. Większej części młodzieży, przykładowo, nieobce jest uzyskiwanie zwolnień lekarskich metodą symulacji objawów chorobowych; skuteczne udawanie choroby nie jest jednak chorobą (na równi zresztą z udawaniem nieskutecznym). Naśladowanie reakcji lękowych nie równa się odczuwaniu strachu. Kreacja aktorska nie jest ucieleśnianiem sobowtóra. Po co zresztą tak daleko szukać — komputerowa symulacja trawienia nikomu nie kojarzy się nawet z własnym żołądkiem; nie przypuszczamy też, by audiowizualny model burzy z piorunami naprawdę mógł nas zmoczyć.

Dlaczego więc, gdy przychodzi do symulowania stanów umysłowych, nagle zmieniamy zdanie? I dlaczego rzeczywistą popularność zyskał sobie pomysł, że cyfrowa symulacja umysłu zawiera rzeczywiste stany umysłowe? Brak tu jednoznacznej odpowiedzi. Zapewne jest to wynik oddziaływania behawioryzmu, przede wszystkim zaś efekt myślenia dualistycznego, dominującego w historii filozofii i potężnego również dzisiaj. Różne wersje dualizmu mają tyle wspólnego, że próbują opisać umysł jako w gruncie rzeczy niezależny od biologii mózgu i biologii w ogóle; głównym dogmatem dualizmu pozostaje pogląd, iż funkcje umysłowe różnią się diametralnie od innych zjawisk naturalnych. Samo sformułowanie teorii sztucznej inteligencji było możliwe tylko dzięki założonemu z góry rozdziałowi umysłu od jego przyrodniczego podłoża. Wiele wskazuje na to, że nie jest to założenie wiarygodne. „Nie ma takiego programu komputerowego, który sam w sobie wyposażałby system w umysł. Mówiąc krótko, programy nie są umysłami ani same w sobie nie wystarczą dla powstania umysłu”²⁰. Nieprzypadkowo bowiem umysły realizowane są przez biologiczne agregaty o określonej strukturze, systemowej złożoności, oraz takiej a nie innej ewolucyjnej historii filogenetycznej. Oczywiście nie wiadomo jeszcze, jakie czynniki przyrodnicze wpłynęły na pojawienie się umysłu i ukształtowały jego ostateczną postać — być może nigdy nie będziemy tego wiedzieć dokładnie. Ale też nie próbowaliśmy tu niczego szczegółowo przesądzać — w tej dziedzinie wyrezytowało nas wielu znakomitych autorów²¹. Odpowiedź na większość pytań leży na obszarze szeroko rozumianych empirycznych badań nad mózgiem; w tym sensie filozof nie ma tu już nic ogólnego do roboty.

¹⁹ J.R. Searle, *Umysł, mózg i nauka*, dz. cyt., s. 33.

²⁰ Tamże, s. 35.

²¹ Por. H. von Ditfurth, *Duch nie spadł z nieba*, Warszawa 1989.

Na koniec jeszcze jedno: czy możemy żyć bez dualistycznych poglądów na temat nas samych? Oczywiście, że tak. Tymi, bez których nie możemy żyć, są filozofia i nauka: filozofia dodaje nam skrzydeł, nauka pozwala stąpać po ziemi. Z naturalistycznego punktu widzenia nie jest żadnym zaskoczeniem konkluzja, 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 i trawieniem”²². Który to fakt, należy sobie życzyć, już wkrótce przestanie być odbierany — przez filozofów i nie tylko — jako zamach na ludzką godność.

²² J.R. Searle, *Umysł, mózg i nauka*, dz. cyt., s. 37.