# MÓZGI W NACZYNIU[1]

Mrówka pełznie po piasku. Jej ślad wije się i wielokrotnie przecina z sobą tak, że w końcu, zupełnie przypadkowo, wyraźnie wygląda jak karykatura Winstona Churchilla. Czy mrówka nakreśliła w ten sposób podobiznę Winstona Churchilla, rysunek, który *przedstawia* Churchilla?

Większość ludzi powiedziałaby, po chwili zastanowienia, że nie. Mrówka nigdy przecież nie widziała Churchilla, ani nawet jego podobizny, i nie miała wcale zamiaru sporządzenia jego portretu. Po prostu nakreśliła pewien ślad (a i *to* nieintencjonalnie), ślad, który *my* możemy „postrzegać jako" podobiznę Churchilla.

Można tę myśl wyrazić mówiąc, że ślad nakreślony przez mrówkę nie jest „sam w sobie" reprezentacją[2] żadnej rzeczy

[1] W oryginale: *Brains in a vat*, co dosłownie znaczy: „mózgi w kadzi". Wyraz „kadź" uznałem jednak za nieodpowiedni w tym kontekście. „Mózgi w kadzi" brzmi nieładnie i ma niestosowne w kontekście eseju konotacje piwowarskie. Laboratoryjna „probówka" również wywołuje niewłaściwe skojarzenia (objętość). Dlatego zdecydowałem się na neutralne, ogólne „naczynie". – Przyp. tłum.

[2] W tym artykule terminy „reprezentowanie"/„reprezentacja" (*representation*) i „odnoszenie się"/„odniesienie przedmiotowe" (*reference*) każdorazowo odnoszą się do relacji między słowem (lub innego rodzaju znakiem, symbolem lub przedstawieniem) a czymś, co rzeczywiście istnieje (tzn. nie jest tylko „przedmiotem myślenia"). Można mówić o „odnoszeniu się" w takim sensie, w którym moje słowa mogą „odnosić się" do czegoś, co nie istnieje;

wyróżnionej spośród innych rzeczy. Podobieństwo (nader złożone) do rysów Winstona Churchilla nie wystarcza do uznania śladu za reprezentację Churchilla, za coś, co się jakoś do niego odnosi. Nie jest również do tego niezbędne: w naszej wspólnocie kulturowej wydrukowany napis „Winston Churchill", wypowiedziane słowa „Winston Churchill" i wiele innych rzeczy służy do reprezentowania Churchilla (aczkolwiek nie obrazowo), chociaż żadna z nich nie jest podobna do Churchilla w taki sposób, jak portret, ani nawet jak schematyczny szkic. Jeżeli *podobieństwo* nie jest ani konieczne, ani wystarczające do tego, aby coś reprezentowało coś innego, to jak *cokolwiek* może być do tego niezbędne lub wystarczające? Jak, u licha, jedna rzecz może reprezentować (lub „oznaczać" itd.) jakąś inną rzecz?

Odpowiedź może wydawać się łatwa. Przypuśćmy, że mrówka widziała kiedyś Winstona Churchilla, i że ma dość inteligencji i umiejętności, aby naszkicować jego podobiznę. Przypuśćmy, że sporządziła jego karykaturę *intencjonalnie*. Wówczas jej ślad będzie reprezentował Churchilla.

Z drugiej strony, przypuśćmy, że ślad ma kształt: WINSTON CHURCHILL. Przypuśćmy również (pomijając kwestię prawdopodobieństwa), że złożyło się tak zupełnie przypadkowo.

tutaj jednak termin „odnosić się" nie będzie w takim sensie występował. Starszym określeniem na to, co nazywam „reprezentowaniem" lub „odnoszeniem się", jest *denotowanie*.

Ponadto, wzorem nowoczesnych logików, używam wyrazu „istnieje" w znaczeniu „istnieje w przeszłości, obecnie lub w przyszłości". Tym samym Winston Churchill „istnieje" i możemy „odnosić się do" Winstona Churchilla lub tworzyć jego „reprezentacje", mimo że on już nie żyje.

[Zgodnie z tradycją, termin *representation* w zastosowaniu do wyobrażeń i pojęć tłumaczę jako „przedstawienie". Zwracam Czytelnikowi uwagę na to, że homonimia słów „reprezentacja" i „przedstawienie" w języku oryginału odgrywa ważną rolę w argumentacji tego eseju. – Przyp. tłum.]



Wówczas „drukowany napis" WINSTON CHURCHILL *nie* będzie reprezentował Churchilla, mimo że drukowany napis tego kształtu reprezentuje Churchilla we wszystkich niemal książkach, w których dziś się znajduje.

Wydaje się więc, że warunkiem koniecznym, ażeby coś było reprezentacją, lub najważniejszym warunkiem koniecznym, ażeby coś było reprezentacją, jest *intencja* reprezentowania.

Ażebym mógł jednak mieć intencję, by *cokolwiek reprezentowało* Churchilla, nawet w języku prywatnym (nawet gdyby słowa „Winston Churchill" były wypowiedziane w myśli, a nie głośno), muszę przede wszystkim być zdolny *pomyśleć* o Churchillu. Jeżeli ślady na piasku, dźwięki itd. nie mogą niczego reprezentować czy przedstawiać „same przez się", to jakim sposobem zdolność reprezentowania lub przedstawiania może przysługiwać formom myślenia „samym przez się"? Czy rzeczywiście może? Jak myśl może sięgnąć poza siebie i „uchwycić" coś zewnętrznego?

W przeszłości niektórzy filozofowie skwapliwie przechodzili od tego rodzaju rozważań do dowodzenia, że umysł jest *zasadniczo innej natury niż rzeczy fizyczne*. Argument jest prosty; to, co zostało powiedziane o krzywej nakreślonej przez mrówkę, dotyczy dowolnego obiektu fizycznego. Żaden obiekt fizyczny nie może, sam przez się, odnosić się do określonej rzeczy wyróżnionej spośród wszystkich innych rzeczy; natomiast *myśli w umyśle* niewątpliwie z powodzeniem pełnią tę funkcję. Myśli (a tym samym i umysł) są więc zasadniczo innej natury niż przedmioty fizyczne. Myśli cechuje *intencjonalność*: mogą one odnosić się do czegoś innego; nic, co fizyczne, nie posiada cechy „intencjonalności", poza intencjonalnością będącą pochodną posługiwania się danym przedmiotem fizycznym przez jakiś umysł. Tak się przynajmniej twierdzi. Przytoczony wywód jest wszakże cokolwiek pospieszny: proste

postulowanie tajemniczych władz umysłu niczego nie rozwiązuje. A problem jest bardzo poważny. Jak możliwa jest intencjonalność? Jak możliwe jest odnoszenie się?

## *Magiczne teorie odnoszenia się*

Przekonaliśmy się, że „obrazu" nakreślonego przez mrówkę nie łączy żaden związek konieczny z Winstonem Churchillem. Sam fakt, że ów „obraz" odznacza się niejakim „podobieństwem" do Churchilla, nie czyni zeń autentycznego portretu, czy reprezentacji Churchilla. Dopóki mrówka nie jest inteligentną mrówką (a nie jest) i nie wie nic na temat Churchilla (a nie wie), nakreślona przez nią krzywa nie jest żadnym obrazem, ani nawet reprezentacją czegokolwiek. Niektóre ludy pierwotne uważają, że pewne reprezentacje (w szczególności *imiona*) łączy jakiś związek konieczny z ich przedmiotami (nosicielami); że znajomość czyjegoś „prawdziwego imienia" lub „prawdziwej nazwy" czegoś daje władzę nad tym kimś lub czymś. Źródłem tej władzy jest *magiczny związek* między nazwą a jej nosicielem; odkąd wiadomo, że nazwę łączy z jej nosicielem związek *jedynie* kontekstowy, przygodny i konwencjonalny, trudno zrozumieć, dlaczego znajomość imienia miałaby mieć jakiekolwiek mistyczne konsekwencje.

Należy sobie jednak uświadomić, że to, co dotyczy obrazów fizycznych, dotyczy również wyobrażeń i przedstawień umysłu w ogóle; przedstawienia umysłu nie różnią się od reprezentacji fizycznych, jeśli chodzi o konieczny lub przygodny charakter związku między nimi a ich przedmiotem. Założenie przeciwne jest przeżytkiem myślenia magicznego.

Przypuszczalnie najłatwiej można będzie tę kwestię zrozumieć na przykładzie *wyobrażeń*. (Bodaj pierwszym filozofem, który dostrzegł niezwykłą doniosłość tej kwestii, nawet



jeśli to nie on pierwszy ją podniósł, był Wittgenstein.) Przypuśćmy, że na jakiejś odległej planecie wyewoluował gatunek istot ludzkich (lub został tam osadzony przez jakichś kosmitów, lub cokolwiek w tym rodzaju). Przypuśćmy, że owi ludzie, choć pod innymi względami do nas podobni, nigdy nie widzieli *drzew*. Przypuśćmy, że nigdy nie wyobrażali sobie drzew (być może jedyną formą życia roślinnego na ich planecie jest pleśń). Przypuśćmy, że pewnego dnia przypadkowo, ze statku kosmicznego, który przeleciał obok, nie nawiązując z nimi żadnego kontaktu, na ich planetę spadł wizerunek drzewa. Cóż to takiego jest? Najrozmaitsze myśli przychodzą im do głowy: jakaś budowla, baldachim, może jakieś zwierzę. Przypuśćmy jednak, że nigdy nie doszli dostatecznie blisko do prawdy.

Dla *nas* obraz, o którym mowa, reprezentuje drzewo. Dla ludzi z tamtej planety reprezentuje on jedynie jakiś dziwny przedmiot o nieznanej naturze i przeznaczeniu. Przypuśćmy, że ktoś z nich, w rezultacie zaznajomienia się z wizerunkiem drzewa, wytworzył sobie w umyśle wyobrażenie dokładnie takie samo, jak moje wyobrażenie drzewa. Jego wyobrażenie nie jest *przedstawieniem* drzewa. Jest tylko przedstawieniem jakiegoś dziwnego przedmiotu (czymkolwiek on jest), przedstawionego na zagadkowym obrazku.

Niemniej można upierać się, że owo wyobrażenie jest *naprawdę* przedstawieniem drzewa, choćby dlatego, że wizerunek, który je wywołał, sam jest przecież reprezentacją drzewa. Istnieje więc łańcuch przyczynowy, aczkolwiek nader osobliwy, który łączy wyobrażenie mieszkańca obcej planety z rzeczywistymi drzewami.

Można jednak sobie wyobrazić sytuację, w której nie ma nawet takiego łańcucha przyczynowego. Przypuśćmy, że upuszczony ze statku kosmicznego „wizerunek drzewa" nie był naprawdę wizerunkiem drzewa, lecz malowidłem powstałym przez przypadkowe chlapnięcia farbami. Nawet jeżeli wy-

glądałoby ono na wizerunek drzewa, w rzeczywistości byłoby nim w takim samym stopniu, jak wykonana przez mrówkę „karykatura" Churchilla jest wizerunkiem tego męża stanu. Można nawet wyobrazić sobie, że statek kosmiczny, który upuścił „wizerunek", pochodzi z planety nie znającej drzew. Wówczas bohaterowie naszej historyjki mogliby mieć nadal wyobrażenia jakościowo identyczne z moim wyobrażeniem drzewa, lecz byłyby one równie dobrze przedstawieniami drzewa, jak czegokolwiek innego.

To samo można powiedzieć o *słowach*. Rozprawa na papierze mogłaby wydawać się dokładnym opisem drzew, gdyby jednak została sporządzona przez małpy przypadkowo uderzające w klawiaturę maszyny do pisania przez milion lat, słowa przez nie wystukane nie odnosiłyby się do niczego. Gdyby ktoś je zapamiętał i powtórzył w myśli bez zrozumienia, słowa te, choć pomyślane, również nie odnosiłyby się do niczego.

Wyobraźmy sobie, że osoba, która w myśli powtarza te słowa, znajduje się pod hipnozą. Przypuśćmy, że te słowa są wyrazami języka japońskiego i zahipnotyzowanej osobie powiedziano, że zna japoński. Przypuśćmy, że powtarzając w myśli te słowa ma ona „wrażenie, że je rozumie". (Aczkolwiek gdyby ktoś przerwał tok jej myślenia i zadał pytanie o *znaczenie* słów przez nią pomyślanych, stwierdziłaby, że nie potrafi odpowiedzieć.) Złudzenie mogłoby nawet być tak doskonałe, że osoba poddana hipnozie zdołałaby zmylić japońskiego telepatę! Jeżeli jednak nie potrafiłaby używać tych słów we właściwych kontekstach, powiedzieć, o czym „myślała" itd., znaczyłoby to, że ich nie rozumie.

Łącząc ze sobą opowiedziane przeze mnie historyjki science fiction, można skonstruować przypadek, w którym ktoś wypowiada w myśli słowa opisujące drzewa w jakimś języku *i* zarazem ma odpowiednie wyobrażenia, lecz *ani* nie rozumie tych słów, *ani* nie wie, co to jest drzewo. Możemy sobie nawet



wyobrazić, że owe wyobrażenia zostały wywołane przez przypadkowe chlapnięcia farbami (aczkolwiek osobę, o której mowa, zahinotyzowano tak, aby sądziła, że jej wyobrażenia dotyczą rzeczy odpowiednich do jej myśli – tylko że kiedy ją spytać, nie będzie umiała powiedzieć, czego one dotyczą). Możemy sobie też wyobrazić, że język, w którym owa osoba myśli w transie hipnotycznym, jest nieznany ani hipnotyzerowi, ani jego medium – może zwykłym zbiegiem okoliczności „nonsensowne zdania", za które uważa je hipnotyzer, składają się na opis drzew w języku japońskim. Krótko mówiąc, wszystko, co jawi się w umyśle osoby, o której mowa, mogłoby być jakościowo identyczne z tym, co jawi się w umyśle Japończyka, który *naprawdę* myśli o drzewach, a mimo to żadna jej myśl nie odnosiłaby się do drzew.

Opisany ciąg zdarzeń jest właściwie niemożliwy, rzecz jasna, podobnie jak jest właściwie niemożliwe, ażeby małpy przypadkowo wystukały na klawiaturze tekst *Hamleta*. To znaczy prawdopodobieństwo opisanych zdarzeń jest tak nikłe, że nigdy się one nie zrealizują (jak sądzimy). Nie są one jednak logicznie ani nawet fizycznie niemożliwe. *Mogłyby* się zdarzyć (zgodnie z prawami fizyki i, być może, z rzeczywistym stanem kosmosu, jeżeli na wielu innych planetach mieszkają istoty inteligentne). I gdyby się zdarzyły, byłyby wymownym dowodem ważnej prawdy pojęciowej, w myśl której nawet rozległy i złożony system reprezentacji, tak werbalnej, jak i wizualnej, nie ma żadnego *istotowego*, wpisanego w rzeczywistość, magicznego związku z reprezentowanymi przedmiotami; związku niezależnego od tego, jak doszło do utworzenia reprezentacji czy przedstawienia, i niezależnego od dyspozycji osoby mówiącej czy myślącej. Nie ma przy tym znaczenia, czy system reprezentacji (słowa i wyobrażenia, w przytoczonym przykładzie) jest zrealizowany fizycznie, słowa zostały napisane lub wypowiedziane, a wizerunki są wizerunkami fizycznymi, czy też

pozostaje zrealizowany jedynie w myśli. Nie jest tak, że słowa pomyślane i obrazy w umyśle są z samej ich *wewnętrznej natury* przedstawieniem tego, czego dotyczą.

### *Przypowieść o mózgach w naczyniu*

Oto dyskutowana przez filozofów możliwość rodem z science fiction: wyobraźmy sobie, że pewien człowiek (Czytelnik może wyobrazić sobie siebie w tej roli) poddał się operacji wykonanej przez niegodziwego uczonego. Jego (twój) mózg został usunięty z ciała i umieszczony w naczyniu wypełnionym pożywką, która podtrzymuje mózg przy życiu. Zakończenia nerwowe zostały podłączone do superkomputera, który powoduje, że osoba, której mózg wyjęto, doświadcza iluzji, iż wszystko jest w najlepszym porządku. Ma ona złudzenie istnienia osób, przedmiotów, nieboskłonu itd.; natomiast w rzeczywistości wszystko, czego ów człowiek doznaje (ty doznajesz), jest następstwem impulsów elektronicznych, płynących od komputera do zakończeń nerwowych. Komputer jest tak sprytnie zaprogramowany, że kiedy ofiara eksperymentu usiłuje podnieść rękę w górę, dzięki sprzężeniu zwrotnemu „widzi" i „czuje", że ręka podnosi się w górę. Odpowiednio modyfikując program, niegodziwy uczony może spowodować, że jego ofiara „doświadczy" (lub dozna złudzenia) uczestnictwa w dowolnie zaaranżowanej przez niegodziwca sytuacji lub obecności w dowolnie zaprojektowanym otoczeniu. Może on również wymazywać pamięć mózgu zamkniętego w naczyniu tak, że nieszczęsnej ofierze będzie się wydawało, że zawsze przebywała w otoczeniu przeznaczonym jej przez eksperymentatora. Ofierze może nawet wydawać się, że siedzi i czyta te słowa o zabawnym, lecz zgoła absurdalnym domniemaniu, iż pewien niegodziwy uczony usuwa ludziom mózgi i umieszcza



je w naczyniu z pożywką, która podtrzymuje je przy życiu. Zakończenia nerwowe są jakoby podłączone do superkomputera, który powoduje, że osoba, której mózg wyjęto, doświadcza iluzji, iż ...

Celem takich opowiastek w toku wykładu z teorii poznania jest oczywiście postawienie w nowoczesnej formie klasycznego problemu sceptycyzmu wobec świata zewnętrznego. (*Skąd wiesz, że nie znajdujesz się w takim żałosnym położeniu*?) Niemniej opisane położenie nadaje się również doskonale do zilustrowania kwestii stosunku umysłu do rzeczywistości.

Zamiast rozważać tylko jeden mózg w naczyniu, moglibyśmy wyobrazić sobie, że wszyscy ludzie (a może wszystkie istoty zmysłowe) są mózgami w naczyniu (lub układami nerwowymi w naczyniu, w przypadku istot o najprostszym układzie nerwowym, jaki pozwala zaliczyć je do istot „zmysłowych"). Niegodziwy uczony musi oczywiście znajdować się na zewnątrz – ale czy aby na pewno? Może nie ma żadnego niegodziwego uczonego, może (choć to absurd) kosmos po prostu przypadkiem składa się z urządzeń automatycznych służących zaopatrzeniu naczynia wypełnionego mózgami i układami nerwowymi.

Tym razem przypuśćmy, że owe automatyczne urządzenia są zaprogramowane tak, aby wywoływać u nas pewną *zbiorową* halucynację, a nie odrębne, pozbawione wzajemnych związków, rozmaite halucynacje. Kiedy więc wydaje mi się, że rozmawiam z tobą, tobie wydaje się, że słyszysz moje słowa. Nie znaczy to, rzecz jasna, że moje słowa faktycznie docierają do twoich uszu – ponieważ (w rzeczywistości) nie masz uszu, a ja w rzeczywistości nie mam ust ani języka. Zamiast tego, kiedy wypowiadam me słowa, bodźce wychodzące z mojego mózgu płyną do komputera, który powoduje, że „słyszę", jak mój własny głos je wymawia, i „czuję", jak porusza się mój język, itd., oraz powoduje, że ty „słyszysz" moje słowa,

„widzisz" mnie, jak mówię itd. Skoro tak, to w pewnym sensie autentycznie ze sobą rozmawiamy. Nie mylę się co do twojego rzeczywistego istnienia (tylko co do istnienia twojego ciała i „świata zewnętrznego" poza mózgami). Z pewnego punktu widzenia fakt, że „cały świat" jest zbiorową halucynacją, nie jest nawet szczególnie istotny; w końcu słyszysz przecież moje słowa, kiedy mówię do ciebie, mimo że mechanizm tego zjawiska jest inny, niż przypuszczamy. (Gdybyśmy byli parą kochanków oddających się miłosnym uściskom, a nie tylko dwojgiem ludzi zajętych rozmową, wówczas, rzecz jasna, sugestia, że jesteśmy tylko mózgami w naczyniu, mogłaby być nieco kłopotliwa.)

Pragnę teraz zadać pytanie, na pozór głupkowate i oczywiste (przynajmniej dla niektórych, w tym również paru nader wyrafinowanych filozofów), które jednak dość szybko doprowadzi nas do autentycznych głębi myśli filozoficznej. Przypuśćmy, że cała nasza historyjka jest rzeczywiście prawdziwa. Czy moglibyśmy, będąc takimi mózgami w naczyniu, *powiedzieć* lub *pomyśleć*, że nimi jesteśmy?

Zamierzam udowodnić, że odpowiedź brzmi: „Nie, nie moglibyśmy". A właściwie zamierzam udowodnić, że domniemanie, iż naprawdę jesteśmy mózgami w naczyniu, choć nie gwałci żadnego prawa fizyki i w żaden sposób nie kłóci się z całym naszym doświadczeniem, nie może być prawdziwe. *Nie może być prawdziwe*, ponieważ w pewnym sensie obala samo siebie.

Argument, który zamierzam przedstawić, jest dość niezwykły i trzeba mi było kilku lat, aby przekonać się, czy nie ma w nim żadnego błędu. Okazał się jednak poprawny. Jego osobliwość polega na tym, że ma on powiązania z pewnymi nader głębokimi zagadnieniami filozoficznymi. (Po raz pierwszy zdałem sobie z tego sprawę, kiedy rozmyślając nad pewnym twierdzeniem nowoczesnej logiki, twierdzeniem Skolema–Löwenheima, nagle dostrzegłem związek między tym twierdzeniem a niektórymi argumentami z *Dociekań filozoficznych* Wittgensteina.)



Domniemanie samo siebie obala, jeżeli z założenia o jego prawdziwości wynika, że jest fałszywe. Weźmy, na przykład, pod uwagę tezę, że *wszystkie zdania ogólne są fałszywe*. Jest ona zdaniem ogólnym. Jeżeli więc jest prawdziwa, musi być fałszywa. A zatem jest fałszywa. Czasami mówi się, że ta czy inna teza jest fałszywa, jeżeli *z założenia, że ktoś ją wyznaje lub głosi*, wynika, iż jest fałszywa. Na przykład teza „Ja nie istnieję" sama się obala, jeżeli została pomyślana przeze mnie (dla dowolnego „*mnie*"). Można zatem mieć pewność, że się istnieje, ilekroć się pomyśli o własnym istnieniu (jak dowodził Kartezjusz).

W dalszym ciągu wywodu wykażę, iż domniemanie, że jesteśmy mózgami w naczyniu, ma tę właśnie właściwość. Jeżeli mamy możność zastanawiać się nad tym, czy jest prawdziwe, czy fałszywe, wówczas nie jest prawdziwe (co jest do wykazania). A zatem nie jest prawdziwe.

Zanim przedstawię dowód, zastanówmy się nad tym, dlaczego wydaje się rzeczą osobliwą, iż taki argument jest w ogóle możliwy (przynajmniej filozofom wyznającym koncepcję prawdy jako „odbicia"). Zgodziliśmy się, że nie kłóciłoby się z prawami fizyki to, że istniałby świat, w którym wszystkie istoty zmysłowe są mózgami w naczyniu. Jak powiadają niektórzy filozofowie, istnieją „możliwe światy", w których wszystkie istoty zmysłowe są mózgami w naczyniu. (Mówienie o „możliwych światach" sugeruje, że jest jakieś *miejsce*, w którym dowolne absurdalne domniemanie jest prawdziwe – przez co może prowadzić na filozoficzne manowce.) Istoty ludzkie w owym możliwym świecie mają dokładnie takie same doświadczenia, jak *my*. Mają takie same myśli (a przynajmniej wypowiadają takie same słowa, mają takie same wyobrażenia,

formy myślenia itd.). Twierdzę jednak, że można podać argument, który dowodzi, iż nie jesteśmy mózgami w naczyniu. W jaki sposób? I dlaczego mieszkańcy możliwego świata, którzy naprawdę *są* mózgami w naczyniu, nie mogą podać tego argumentu?

Odpowiedź będzie brzmiała (w zasadzie) następująco: ludzie w owym możliwym świecie mogą wprawdzie pomyśleć i „wypowiedzieć" każde słowo, które my potrafimy pomyśleć lub wypowiedzieć, lecz ich słowa nie mogą (jak twierdzę) *odnosić się* do tego samego, co nasze słowa. W szczególności nie mogą oni pomyśleć ani powiedzieć, że są mózgami w naczyniu (*nawet myśląc: „jesteśmy mózgami w naczyniu"*).

## *Test Turinga*

Przypuśćmy, że udało się skonstruować komputer zdolny do prowadzenia inteligentnej rozmowy (na każdy temat, na który może rozmawiać inteligentna osoba). Jak można rozstrzygnąć kwestię, czy ów komputer „myśli"?

Brytyjski logik Alan Turing zaproponował następujący test[3]: posadzić kogoś do rozmowy z komputerem i do rozmowy z nie znaną temu komuś osobą. Jeżeli nie udaje się odróżnić, kiedy partnerem rozmowy jest komputer, a kiedy człowiek, to (przy założeniu, że test jest powtarzany wystarczająco wiele razy ze zmieniającymi się rozmówcami) komputer myśli. Krótko mówiąc, maszyna myśli, jeżeli potrafi przejść „test Turinga". (Rozmowy nie mają odbywać się twarzą w twarz, rzecz jasna, ponieważ eksperymentator nie może wiedzieć, jak wyglądają dwaj pozostali uczestnicy rozmowy. Nie należy też posługiwać

[3] A. M. Turing, *Maszyny liczące a inteligencja*, tłum. D. Gajkowicz, w: *Maszyny matematyczne a myślenie*, A. Feigenbaum, J. Feldman (red.), PWN, Warszawa 1972, ss. 24-47.



się dźwiękiem, ponieważ dźwięk mechaniczny może mieć po prostu inne brzmienie od ludzkiego głosu. Wyobraźmy sobie raczej, że cała rozmowa toczy się przy użyciu elektronicznej maszyny do pisania. Eksperymentator wystukuje swoje stwierdzenia, zapytania itp., a uczestnicy eksperymentu – maszyna i człowiek – odpowiadają za pośrednictwem elektronicznej klawiatury. Maszyna może też kłamać: na pytanie: „Czy jesteś maszyną?" może odrzec: „Nie, jestem laborantem w tutejszym laboratorium".)

Myśl, że taki test jest rzeczywiście ostatecznym sprawdzianem myślenia, stała się przedmiotem krytyki ze strony licznych autorów (którzy w zasadzie nie byli wcale wrogo nastawieni do idei, że maszyna może myśleć). Tym razem jednak chodzi o coś innego. Pragnę posłużyć się ideą testu Turinga, ogólną ideą *dialogicznego testu kompetencji*, do innego celu, do celu zbadania pojęcia *odnoszenia się*.

Wyobraźmy sobie sytuację, gdy problemem nie jest rozstrzygnięcie, czy uczestnik rozmowy jest człowiekiem czy maszyną, lecz czy stosuje słowa do tych samych rzeczy, co my. Oczywistym sprawdzianem jest, po raz wtóry, rozmowa, i jeżeli nie powstają żadne problemy, jeżeli rozmówca „przechodzi test" w tym sensie, że okazuje się nieodróżnialny od kogoś, o kim wcześniej stwierdzono, iż mówi tym samym językiem, iż jego słowa mają zwyczajowe odniesienie przedmiotowe itd., należy wnosić, że stosuje słowa do tych samych rzeczy, co my. Ilekroć test Turinga będzie miał na celu ustalenie, czy istnieje (wspólne) odniesienie przedmiotowe, będę nazywał go *testem Turinga na odniesienie przedmiotowe*. I tak samo, jak różni filozofowie spierali się o to, czy oryginalny test Turinga jest *ostatecznym* sprawdzianem myślenia, tj. o to, czy maszynę, która „przechodzi" ten test nie tylko raz, lecz regularnie, należy *koniecznie* uznać za myślącą, tak też pragnę przedyskutować kwestię, czy zaproponowany przed chwilą test Turinga na

odniesienie przedmiotowe jest ostatecznym sprawdzianem odnoszenia się słów do tych samych rzeczy.

Odpowiedź okaże się brzmieć: „Nie". Test Turinga na odniesienie przedmiotowe nie jest ostateczny. Jest oczywiście znakomitym sprawdzianem praktycznym, nie jest jednak logicznie niemożliwe (aczkolwiek na pewno wysoce nieprawdopodobne), że ktoś może przejść test Turinga na odniesienie przedmiotowe, chociaż nie odnosi swoich słów do niczego. Wynika stąd, jak stwierdzimy, że można uogólnić nasze spostrzeżenie, iż słów (i całych tekstów oraz dyskursów) nie łączy żaden związek konieczny z ich odniesieniem przedmiotowym. Nawet gdy weźmiemy pod uwagę nie słowa, jako takie, lecz reguły właściwego zastosowania słów w określonych kontekstach – nawet gdy weźmiemy pod uwagę, ujmując to w żargonie komputerowym, *programy posługiwania się słowami* – dopóki te programy jako takie nie będą *odnosiły się do czegoś pozajęzykowego*, słowa nie będą miały ustalonego odniesienia przedmiotowego. Oto decydujący krok w procesie dochodzenia do konkluzji, że mieszkańcy świata mózgów zamkniętych w naczyniu nie mogą w ogóle swoich słów odnosić do niczego zewnętrznego (i dlatego nie mogą powiedzieć, *że* są mieszkańcami świata mózgów zamkniętych w naczyniu).

Przypuśćmy, na przykład, że znajduję się w sytuacji opisanej przez Turinga (gram w „grę w naśladownictwo", w terminologii Turinga) i moim partnerem faktycznie jest maszyna. Przypuśćmy, że maszyna jest zdolna wygrać („przejść" test). Wyobraźmy sobie, że maszyna jest tak zaprogramowana, by pięknie po polsku odpowiadać na polskie stwierdzenia, zapytania, uwagi itd., lecz nie ma żadnych organów zmysłowych (poza podłączeniem do mojej elektronicznej maszyny do pisania). (Jeśli dobrze rozumiem, Turing nie zakłada, że posiadanie organów zmysłowych lub organów ruchu jest



niezbędnym warunkiem myślenia lub inteligencji.) Załóżmy, że maszynie brakuje nie tylko elektronicznych oczu, uszu itd., lecz że i oprogramowanie maszyny, program gry w naśladownictwo, nie przewiduje możliwości odbioru sygnałów dostarczanych przez takie narządy zmysłowe ani kierowania ruchami ciała. Co należałoby powiedzieć o takiej maszynie?

Wydaje mi się oczywiste, że nie możemy, i nie powinniśmy, przypisywać takiemu urządzeniu zdolności odnoszenia się do czegokolwiek zewnętrznego. Co prawda, maszyna może przepięknie rozmawiać o, powiedzmy, krajobrazach polskich. Nie potrafiłaby jednak rozpoznać drzewa, ani jabłka, góry, ani krowy, pola, ani wieży, gdyby postawić ją przed którąś z tych rzeczy.

Mamy do czynienia z urządzeniem do produkowania zdań w odpowiedzi na zdania. Natomiast żadne z tych zdań nijak nie dotyczy rzeczywistego świata. *Jeżeli postawić obok siebie dwie maszyny i kazać im rozgrywać ze sobą grę w naśladownictwo, nie zaprzestaną nigdy „nabierać" się wzajem, nawet gdyby reszta świata przestała istnieć!* Nie ma powodu sądzić, że gdy maszyna mówi o jabłkach, jej słowa odnoszą się do rzeczywistych jabłek, tak samo jak nie ma powodu sądzić, że „szkic" mrówki odnosi się do Winstona Churchilla.

Złudzenie posiadania odniesienia przedmiotowego, znaczenia, inteligencji itd. wynika stąd, że na mocy *naszych* konwencji reprezentowania mowa maszyny odnosi się do jabłek, wież, Polski itd. Z podobnych powodów powstaje *złudzenie*, że mrówka narysowała karykaturę Churchilla. My jednak potrafimy postrzegać jabłka i pola, oraz podejmować różne zabiegi z nimi związane. Nasze mówienie o jabłkach i polach pozostaje w ścisłych związkach z *pozawerbalnymi* czynnościami dotyczącymi jabłek i pól. Istnieją „językowe reguły wejścia", które łączą doznawanie obecności jabłek z wypowiedziami typu „Widzę jabłko", oraz „językowe reguły wyjścia",

które łączą postanowienia wyrażone w formie językowej („Idę kupić jabłka") z działaniami pozajęzykowymi. W sytuacji, gdy brak językowych reguł wejścia i wyjścia, nie ma powodu uważać rozmowy z maszyną (lub między dwiema maszynami, we wspomnianym przez nas przypadku, w którym dwie maszyny grają ze sobą w naśladownictwo) za coś więcej niż grę syntaktyczną. Gra syntaktyczna na pewno *przypomina* inteligentną rozmowę; lecz tylko tak (ani trochę bardziej) jak krzywa nakreślona przez mrówkę przypomina kąśliwą karykaturę.

W przypadku mrówki moglibyśmy twierdzić, że mrówka nakreśliłaby tę samą krzywą, nawet gdyby Winston Churchill nigdy nie istniał. W przypadku maszyny nie możemy wysunąć analogicznego argumentu; gdyby nie istniały jabłka, pola lub wieże, programiści przypuszczalnie nie ułożyliby takiego programu. Maszyna nie *postrzega* jabłek, pól ani wież, lecz jej twórcy i projektanci je postrzegali. Istnieje *pewnego rodzaju* związek przyczynowy między maszyną a jabłkami z rzeczywistego świata itd., zapośredniczony przez doświadczenie postrzeżeniowe i wiedzę twórców i projektantów. Niemniej tak słaby związek nie wystarcza, aby uznać, że mowa maszyny ma odniesienie przedmiotowe. Nie tylko jest logicznie możliwe, choć fantastycznie nieprawdopodobne, że ta sama maszyna *mogłaby* istnieć, nawet gdyby jabłka, pola i wieże nie istniały; lecz, co ważniejsze, maszynie jest zupełnie obojętna kwestia *kontynuacji* istnienia jabłek, pól, wież itd. Nawet gdyby to wszystko *przestało* istnieć, maszyna prowadziłaby swoje rozmowy równie swobodnie, jak przedtem. Oto dlaczego nie można uważać, że mowa maszyny do czegokolwiek się odnosi.

Dla naszej dyskusji istotne znaczenie ma to, że test Turinga w żaden sposób nie pozwala wyeliminować maszyn zaprogramowanych *wyłącznie* do gry w naśladownictwo, i że maszyna, która nie potrafi nic *poza* uprawianiem gry w naśladownictwo, *zdecydowanie* do niczego nie odnosi swoich zagrań, podobnie jak magnetofon.



## *Mózgi w naczyniu (po raz wtóry)*

Porównajmy hipotetyczne „mózgi w naczyniu" z opisanymi przed chwilą maszynami. Są między nimi oczywiście poważne różnice. Mózgi w naczyniu nie mają narządów zmysłowych, lecz mają *namiastkę* narządów zmysłowych; to jest mają zakończenia nerwów dośrodkowych, odbierają sygnały płynące z zakończeń nerwów dośrodkowych i te sygnały stanowią istotne dane dla „programu" mózgów w naczyniu, tak samo jak dla programu naszych mózgów. Mózgi w naczyniu są *mózgami*; są ponadto *działającymi* mózgami, które funkcjonują na tych samych zasadach, co mózgi w świecie rzeczywistym. Dlatego wydaje się absurdem odmawiać im świadomości lub inteligencji. Niemniej ich świadomość i inteligencja nie świadczy wcale o tym, że słowa, którymi się posługują, odnoszą się do tych samych rzeczy, co nasze słowa. Pytanie, które nas interesuje, brzmi następująco: czy ich sformułowania zawierające, powiedzmy, słowo „drzewo" rzeczywiście odnoszą się do *drzew*? Ogólniej: czy ich słowa mogą w ogóle odnosić się do przedmiotów *zewnętrznych*? (A nie, na przykład, przedmiotów w świecie pozoru, wytworzonym przez urządzenia automatyczne.)

Aby usprawnić tok rozumowania, przyjmijmy, że urządzenia automatyczne, o których mowa, powstały na skutek swego rodzaju kosmicznego zbiegu okoliczności (lub, ewentualnie, istnieją od zawsze). Same te automatyczne urządzenia w naszym hipotetycznym świecie nie muszą mieć żadnych inteligentnych twórców czy projektantów. Jak powiedziałem na

początku, możemy sobie wyobrazić, że wszystkie istoty zmysłowe (o najmniejszej nawet wrażliwości) znajdują się wewnątrz naczynia.

Powyższe założenie nic nie pomaga. Nie ma bowiem żadnego związku między *słowem* „drzewo", którym posługują się mózgi z naszej opowieści, a rzeczywistymi drzewami. Posługiwałyby się słowem „drzewo" dokładnie tak samo, myślałyby dokładnie tak samo, miałyby dokładnie takie same wyobrażenia, nawet gdyby w rzeczywistości nie istniały żadne drzewa. Ich wyobrażenia, słowa itd. są jakościowo identyczne z wyobrażeniami, słowami itd., które rzeczywiście reprezentują drzewa w *naszym* świecie, przekonaliśmy się jednak (znów mrówka!), że jakościowe podobieństwo do czegoś, co reprezentuje pewien przedmiot (Winstona Churchilla lub drzewo), nie czyni jeszcze reprezentacji z samej tej rzeczy wykazującej owo podobieństwo. Krótko mówiąc, mózgi w naczyniu nie mają na myśli rzeczywistych drzew, gdy myślą sobie: „naprzeciw mnie stoi drzewo", ponieważ nic nie wskazuje na to, że ich myśl „drzewo" jest przedstawieniem rzeczywistego drzewa.

Jeżeli taki wniosek wydaje się pochopny, weźmy pod uwagę rzecz następującą: stwierdziliśmy, że słowa niekoniecznie odnoszą się do drzew, nawet jeśli ułożone są w ciąg identyczny z wypowiedzią, która (gdyby powstała w umyśle kogoś z nas) niewątpliwie mówiłaby *o drzewach* w świecie rzeczywistym. Również „program", w sensie reguł, zwyczajów i dyspozycji mózgu do określonych zachowań werbalnych, niekoniecznie odnosi się do drzew lub ustanawia odniesienie do drzew poprzez ustalenie związków między wyrazami, lub między sygnałami *językowymi, a językowymi* odzewami na nie. Owe mózgi mogą myśleć o drzewach, odnosić się do drzew, przedstawiać sobie drzewa (rzeczywiste drzewa, poza naczyniem) tylko wtedy, gdy „program" w szczególny sposób łączy



system językowy z *pozawerbalnymi* sygnałami wejścia i wyjścia. W świecie Mózgów w Naczyniu faktycznie istnieją takie pozawerbalne sygnały wejścia i wyjścia (znów zakończenia nerwów dośrodkowych i odśrodkowych!), stwierdziliśmy jednak, że „dane zmysłowe" wytworzone przez urządzenia automatyczne nie przedstawiają drzew (ani niczego zewnętrznego), nawet gdy są dokładnymi podobiznami naszych wyobrażeń drzew. Tak jak chlapnięcie farbą może przypominać wizerunek drzewa nie *będąc* wizerunkiem drzewa, tak i „dana zmysłowa", jak stwierdziliśmy, może być jakościowo identyczna z „wyobrażeniem drzewa" nie będąc wyobrażeniem drzewa. Jeśli idzie o mózgi w naczyniu – jakim sposobem fakt, że język ma ustalone przez program związki z odbieranymi bodźcami zmysłowymi, które ani ze swej istoty, ani na żadnej konwencjonalnej zasadzie nie reprezentują drzew (ani niczego zewnętrznego), może sprawić, by cały system reprezentacji, język w jego użyciu, *rzeczywiście* odnosił się do drzew, lub reprezentował drzewa, bądź cokolwiek zewnętrznego?

Odpowiedź brzmi: nie może. Cały system danych zmysłowych, sygnałów motorycznych do zakończeń nerwów odśrodkowych oraz werbalnie lub pojęciowo zapośredniczonych myśli, połączonych przez „językowe reguły wejścia" z danymi zmysłowymi (lub czymkolwiek w tym rodzaju) jako sygnałami wejścia i przez „językowe reguły wyjścia" z bodźcami motorycznymi jako sygnałami wyjścia, ma nie większy związek z *drzewami*, niż krzywa nakreślona przez mrówkę – z Winstonem Churchillem. Skoro stwierdzamy, że *podobieństwo jakościowe* (czy nawet, jeśli chcecie, jakościowa identyczność) myśli mózgów w naczyniu i myśli kogoś w świecie rzeczywistym w żaden sposób nie implikuje tożsamości odniesienia przedmiotowego, nietrudno zrozumieć, że nie ma żadnych podstaw, aby sądzić, iż słowa mózgu w naczyniu odnoszą się do rzeczy zewnętrznych.

## *Przesłanki argumentu*

W ten sposób przedstawiłem obiecany argument na dowód tezy, że mózgi w naczyniu nie mogą pomyśleć ani powiedzieć, iż są mózgami w naczyniu. Teraz pozostaje tylko nadać mu bardziej przejrzystą formę i zbadać jego strukturę.

Zgodnie z tym, co zostało powiedziane, kiedy mózg w naczyniu (w świecie, w którym każda istota zmysłowa jest i zawsze była mózgiem w naczyniu) myśli sobie: „Przede mną stoi drzewo", jego myśl nie odnosi się do rzeczywistych drzew. Na gruncie pewnych teorii, które będziemy omawiać później, jego myśl może odnosić się do drzew w świecie pozoru, lub do impulsów elektronicznych wywołujących wrażenie drzewa, lub do właściwości programu odpowiedzialnych za te impulsy. Teorie te nie są sprzeczne z tym, co zostało powiedziane przed chwilą, ponieważ zachodzi ścisły związek przyczynowy między użyciem słowa „drzewo" w polszczyźnie mózgów w naczyniu a występowaniem drzew w świecie pozoru – określonego rodzaju impulsami elektronicznymi oraz określonymi właściwościami programu maszyny. W myśl tych teorii, mózg ma *rację*, nie *myli się*, gdy myśli: „Przede mną stoi drzewo". Biorąc pod uwagę to, do czego w polszczyźnie mózgów w naczyniu odnosi się „drzewo", i do czego odnosi się „przede mną", przy założeniu, że jedna ze wspomnianych teorii jest trafna, warunki prawdziwości zdania „Przede mną stoi drzewo" wypowiedzianego w polszczyźnie mózgów w naczyniu stwierdzają po prostu, że drzewo ze świata pozoru jest „przede" tym „mną", o którym mowa – w świecie pozoru – lub że urządzenia automatyczne wysłały impuls elektroniczny, który normalnie wywołuje takie wrażenie, lub że zostały uruchomione funkcje urządzenia, które mają wywoływać wrażenie „drzewa przede mną". Takie warunki prawdziwości są oczywiście spełnione.



Na mocy tej samej argumentacji, w polszczyźnie mózgów w naczyniu słowo „naczynie" odnosi się do naczyń ze świata pozoru, lub czegoś z tym związanego (impulsów elektronicznych lub właściwości programu), lecz na pewno nie do rzeczywistych naczyń, ponieważ nie ma żadnego związku przyczynowego między użyciem słowa „naczynie" w polszczyźnie mózgów w naczyniu a rzeczywistymi naczyniami (poza tym, że mózgi w naczyniu nie byłyby zdolne posługiwać się słowem „naczynie", gdyby nie było jednego szczególnego naczynia – tego, w którym zostały zamknięte; ten związek jednak zachodzi między użyciem *dowolnego* słowa w polszczyźnie mózgów w naczyniu a tym konkretnym naczyniem; nie jest to związek między *szczególnym* wyrazem „naczynie" a naczyniami). Podobnie „pożywka" odnosi się w polszczyźnie mózgów w naczyniu do pożywki ze świata pozoru, lub czegoś z tym związanego (impulsów elektronicznych lub funkcji programu). Wynika stąd, że jeżeli ich „możliwy świat" jest naprawdę światem rzeczywistym, a my naprawdę jesteśmy mózgami w naczyniu, to mówiąc: „jesteśmy mózgami w naczyniu" mamy na myśli to, że *jesteśmy mózgami w naczyniu w świecie pozoru*, lub coś w tym rodzaju (jeżeli w ogóle mamy coś na myśli). Niemniej w skład hipotezy, według której jesteśmy mózgami w naczyniu, wchodzi domniemanie, że nie jesteśmy mózgami w naczyniu w świecie pozoru (tj. że nie jesteśmy mózgami w naczyniu w naszej „halucynacji"). Zatem, jeżeli jesteśmy mózgami w naczyniu, zdanie „Jesteśmy mózgami w naczyniu" stwierdza pewien fałsz (jeżeli stwierdza cokolwiek). Krótko mówiąc, jeżeli jesteśmy mózgami w naczyniu, zdanie „Jesteśmy mózgami w naczyniu" jest fałszywe. Jest zatem (z konieczności) fałszywe.

Przypuszczenie, że taka możliwość ma w ogóle sens, wynika z połączenia dwóch błędów: (1) nazbyt poważnego potraktowania *możliwości fizycznej*; oraz (2) nieświadomego po-

służenia się magiczną teorią odniesienia przedmiotowego, na mocy której określone przedstawienia w umyśle koniecznie odnoszą się do określonych rzeczy i rodzajów rzeczy.

Istnieje „fizycznie możliwy świat", w którym jesteśmy mózgami w naczyniu – cóż to znaczy poza tym, że istnieje *opis* takiego właśnie stanu rzeczy, który nie jest sprzeczny z prawami fizyki? Tak jak w naszej kulturze występuje tendencja (począwszy od siedemnastego wieku) do traktowania *fizyki* jako metafizyki, to jest do upatrywania w naukach ścisłych długo poszukiwanego opisu „prawdziwego i ostatecznego umeblowania wszechświata", tak też występuje, jako jej bezpośrednie następstwo, tendencja do traktowania „fizycznej możliwości" jako kryterium decydującego o tym, co faktycznie może mieć miejsce. Prawda jest prawdą fizyczną; możliwość możliwością fizyczną; a konieczność – koniecznością fizyczną, na gruncie tego poglądu. Przed chwilą jednak stwierdziliśmy, na razie choćby tylko na przytoczonym wymyślnym przykładzie, że ów pogląd jest błędny. Istnienie „fizycznie możliwego świata", w którym jesteśmy mózgami w naczyniu (i zawsze nimi byliśmy oraz zawsze będziemy), nie oznacza, że moglibyśmy rzeczywiście być mózgami w naczyniu. Tę możliwość wyklucza nie fizyka, lecz *filozofia*.

Niektórzy filozofowie, pragnący zdecydowanie głosić tezy własnej dyscypliny, a jednocześnie pomniejszać ich znaczenie (typowy stan umysłowy anglo-amerykańskiej filozofii dwudziestego wieku), powiedzieliby: „Jasne. Udowodniłeś, że coś, co wydaje się fizycznie możliwe, jest naprawdę niemożliwe *pojęciowo*. I co w tym dziwnego?"

Cóż, mój argument na pewno można określić jako rozumowanie „pojęciowej" natury. Atoli ujmowanie działalności filozoficznej jako poszukiwania prawd „pojęciowych" sprowadza ją do *dociekania znaczenia słów*. A przecież zajmujemy się czymś innym zgoła.



Zajmujemy się mianowicie rozpatrywaniem *koniecznych warunków myślenia o czymś, reprezentowania, odnoszenia się* itd. Badamy te warunki nie za pomocą dociekań znaczenia tych słów i wyrażeń (jak mogliby to czynić np. językoznawcy), lecz za pomocą *rozumowania a priori*. Nie w starym, „absolutnym" sensie (ponieważ nie twierdzimy, że magiczne teorie odniesienia przedmiotowego są *a priori* błędne), lecz w sensie dociekań tego, co jest *racjonalnie* możliwe, *przyjmując* w tych dociekaniach pewne naczelne przesłanki, lub pewne bardzo ogólne założenia teoretyczne. Tego rodzaju postępowanie nie jest „empiryczne", ani też całkowicie „*a priori*", lecz zawiera pierwiastki obu sposobów badania. Mimo że moja metoda jest zawodna, i że zależy od założeń, które można określić jako „empiryczne" (np. założenie, w myśl którego umysł ma dostęp do przedmiotów zewnętrznych wyłącznie za pomocą zmysłów), ma ona bliskie związki z tym, co Kant nazywał dedukcją „transcendentalną"; polega ona bowiem, powtarzam, na poszukiwaniu *koniecznych warunków* odnoszenia się, a zatem myślenia – warunków wpisanych w naturę naszych umysłów, choć nie całkiem (jak Kant się spodziewał) niezależnych od założeń empirycznych.

Jedna z przesłanek rozumowania jest oczywista: magiczne teorie odnoszenia się są błędne, tak jeśli chodzi o przedstawienia umysłu, jak i reprezentacje fizyczne. Kolejna przesłanka mówi, że nie można odnosić się do rzeczy tego czy innego rodzaju, np. do *drzew*, jeżeli w ogóle nie wchodzi się w interakcje przyczynowe z nimi[4], ani z żadnymi innymi

[4] Jeżeli mózgi w naczyniu wejdą *w przyszłości* w związki przyczynowe, powiedzmy, z drzewami, to być może mogą *teraz* odnosić swoje wypowiedzi do drzew za pomocą opisu: „rzeczy, które będę nazywał drzewami wtedy i wtedy w przyszłości". Mamy jednak wyobrazić sobie przypadek, w którym mózgi w naczyniu *nigdy* nie opuszczają naczynia, a więc *nigdy* nie wchodzą w związki przyczynowe z drzewami itd.

rzeczami, za pomocą których można je opisywać. Lecz dlaczego mielibyśmy te przesłanki zaakceptować? Skoro to one tworzą ogólne ramy dyskusji, najwyższy czas przyjrzeć się im dokładniej.

### *Powody odrzucenia związku koniecznego między reprezentacjami i przedstawieniami a ich przedmiotami*

Wspomniałem uprzednio, że niektórzy filozofowie przypisywali umysłowi (najbardziej zasłynął z tego Brentano) szczególną władzę, „intencjonalność", która to władza umożliwia *odnoszenie się* jego czynności do określonych przedmiotów. Odrzuciłem ten pogląd stanowczo, ponieważ według mnie nie stanowi on żadnego rozwiązania. Cóż jednak upoważnia mnie do tego? Może podjąłem decyzję nazbyt pochopną?

Filozofowie, o których mowa, nie twierdzili, że możemy myśleć o zewnętrznych przedmiotach lub ich własnościach bez posługiwania się przedstawieniami. A podany przeze mnie argument porównujący wzrokowe dane zmysłowe z „wizerunkiem" nakreślonym przez mrówkę (argument nawiązujący do opowieści science fiction o „wizerunku" drzewa powstałym wskutek chlapnięcia farbą, który wywołuje dane zmysłowe jakościowo podobne do naszych „wzrokowych wyobrażeń drzew", lecz nie związane z żadnych *pojęciem* drzewa) zostałby przez nich zaakceptowany jako dowód na to, że *wyobrażenia* niekoniecznie odnoszą się do czegoś. Jeżeli istnieją przedstawienia umysłu, które z konieczności odnoszą się do rzeczy zewnętrznych, muszą one mieć naturę *pojęć*, a nie wyobrażeń. Lecz czym są *pojęcia*?

W introspekcji nie postrzegamy „pojęć" żeglujących jako takie przez umysł. Zatrzymując w dowolnym momencie potok



myśli, chwytamy słowa, wyobrażenia, doznania, odczucia. Wypowiadając swoje myśli na głos nie wymyślam ich po raz drugi. Słyszę moje słowa tak samo, jak ty. Na pewno mam inne odczucia, gdy wierzę w to, co mówię, niż wtedy, gdy nie wierzę (choć czasem, gdy jestem zdenerwowany, lub zwracam się do nieprzychylnych słuchaczy, czuję się tak, jakbym kłamał, nawet jeżeli wiem, że mówię prawdę); i mam inne odczucia, gdy rozumiem to, co mówię, niż wtedy, gdy nie rozumiem. Mogę sobie jednak bez trudu wyobrazić kogoś, kto myśli tymi samymi słowami (w tym sensie, że wypowiada je w myśli), i tak samo jak ja czuje, że je rozumie, potwierdza itd. i w chwilę później uświadamia sobie (np. po przebudzeniu przez hipnotyzera), że wcale nie rozumiał swoich myśli, a nawet nie rozumiał języka, w którym były sformułowane. Nie twierdzę, że jest to wysoce prawdopodobne; twierdzę tylko, że wcale nie jest to niewyobrażalne. Wynika stąd nie to, że pojęcia są słowami (lub wyobrażeniami, doznaniami itd.), lecz to, że przypisywać komuś „pojęcie" lub „myśl" to coś zgoła innego niż przypisywanie mu jakiegokolwiek „przedstawienia" umysłu, jakiegokolwiek wykrywalnego w introspekcji jestestwa czy zdarzenia. Pojęcia nie są przedstawieniami umysłu, które ze swej istoty odnoszą się do przedmiotów zewnętrznych, z tego prostego powodu, że w ogóle nie są przedstawieniami umysłu. Pojęcia są znakami stosowanymi w szczególny sposób; są znaki publiczne i prywatne, jestestwa myślne i fizyczne, jednak nawet wtedy, gdy znaki są „myślne" i „prywatne", znak sam przez się, w oderwaniu od jego użycia, nie jest pojęciem. Znaki same przez się, ze swej wewnętrznej natury, nie odnoszą się do niczego.

Można to wyraźnie stwierdzić wykonując bardzo prosty eksperyment myślowy. Przypuśćmy, że tak samo jak ja, nie potrafisz odróżnić wiązu od buka. Mimo to mówimy, że odniesienie przedmiotowe słowa „wiąz" wypowiedzianego

przeze mnie jest takie samo, jak słowa „wiąz" wypowiedzianego przez kogokolwiek innego, mianowicie są nim drzewa tego gatunku. Natomiast ekstensją słowa „buk" wypowiedzianego zarówno przeze mnie jak i przez ciebie jest zbiór wszystkich buków (tj. zbiór przedmiotów, o których można prawdziwie orzec: „jest bukiem"). Czy można uznać za wiarygodny pogląd, że różnica między odniesieniem przedmiotowym słów „wiąz" i „buk" wynika z różnicy między naszymi *pojęciami*? Moje pojęcie wiązu jest dokładnie takie samo, jak moje pojęcie buka (wstyd się przyznać). (Nawiasem mówiąc, dowodzi to, że ustalenie odniesienia przedmiotowego jest kwestią społeczną, a nie indywidualną; ty i ja polegamy na ekspertach, którzy *potrafią* odróżniać wiązy od buków.) Kto bohatersko usiłuje twierdzić, że wyjaśnieniem różnicy odniesienia przedmiotowego słów „wiąz" i „buk" w *moim* języku jest różnica między stanami psychicznymi, towarzyszącymi ich wypowiadaniu, niech sobie wyobrazi Ziemię Bliźniaczą, gdzie te dwa słowa mają przestawione znaczenia. Ziemia Bliźniacza jest bardzo podobna do Ziemi; można nawet założyć, że poza przestawieniem znaczenia wyrazów „wiąz" i „buk" wszystko inne na Ziemi Bliźniaczej ma się dokładnie tak samo, jak na Ziemi. Przypuśćmy, że mam sobowtóra na Ziemi Bliźniaczej, który jest molekuła w molekułę identyczny ze mną (w takim samym sensie, w jakim dwa krawaty mogą być „identyczne"). Jeżeli jesteś dualistą, możesz również założyć, że myśli mego sobowtóra są zwerbalizowane identycznie jak moje i ma on takie same dane zmysłowe, takie same dyspozycje itd. Byłoby absurdem uważać, że jego stan psychiczny różni się czymkolwiek od mojego: niemniej jego słowo „wiąz" reprezentuje *buki*, a moje – wiązy. (Podobnie, jeżeli „woda" na Ziemi Bliźniaczej jest jakimś innym płynem – powiedzmy, *XYZ* zamiast $H_2O$ – wówczas słowo „woda" reprezentuje inny płyn, jeżeli zostało wypowiedziane na Ziemi Bliźniaczej, a inny – jeżeli zostało



wypowiedziane na Ziemi, itd.) Wbrew doktrynie, która nam towarzyszy od siedemnastego wieku, *znaczenia nie znajdują się w głowie*.

Stwierdziliśmy, że posiadanie pojęcia nie jest kwestią posiadania wyobrażeń (powiedzmy, wyobrażeń drzew – lub nawet wyobrażeń, „wzrokowych" lub „słuchowych", zdań lub dłuższych wypowiedzi, skoro o tym mowa), ponieważ można posiadać dowolnie bogaty system wyobrażeń nie posiadając *zdolności* posługiwania się zdaniami w sposób odpowiedni do sytuacji (odpowiedni jeśli chodzi zarówno o czynniki językowe – o czym była mowa poprzednio – jak i pozajęzykowe, decydujące o „stosowności w danej sytuacji"). Człowiek może mieć wszelkie możliwe wyobrażenia i mimo to być zupełnie bezradny, gdy mu się powie: „pokaż mi drzewo", nawet jeśli drzew jest pod dostatkiem. Może nawet mieć wyobrażenie o tym, co ma robić, i mimo to nie wiedzieć, co ma robić. Wyobrażenie bowiem, jeżeli nie towarzyszy mu zdolność do określonego typu działania, jest tylko *wyobrażeniem*, a umiejętność działania wedle wyobrażenia jest umiejętnością, którą można posiadać lub nie. (Człowiek może wyobrazić sobie, że pokazuje palcem drzewo, po to tylko, by rozważyć pewną logiczną możliwość: możliwość wskazania palcem na drzewo po tym, jak ktoś wypowie – niezrozumiały dlań – ciąg dźwięków „proszę pokazać palcem drzewo"). Nadal nie wiedziałby, że ma pokazać palcem drzewo, i nie *rozumiałby* rozkazu „pokaż palcem drzewo".

Zastanawiałem się nad uznaniem zdolności do używania określonych zdań za kryterium posiadania w pełni wykształconego pojęcia. Kryterium to można jednak zliberalizować. Moglibyśmy na przykład dopuścić symbolikę, która nie zawiera wyrazów języka naturalnego, i moglibyśmy dopuścić takie zjawiska mentalne, jak wyobrażenia i inne zdarzenia wewnętrzne. Istotne jest, aby były one tak samo skomplikowane, jak

zdania języka naturalnego; by mogły być składane ze sobą, jak one, itd. Bo choć takie czy inne przedstawienie – powiedzmy, niebieski błysk – mogłoby posłużyć jakiemuś matematykowi jako wewnętrzne sformułowanie pełnego dowodu twierdzenia o liczbach pierwszych, to przecież trudno byłoby tak twierdzić (i byłoby fałszem tak twierdzić), gdyby ów matematyk nie potrafił rozłożyć swojego „niebieskiego błysku" na poszczególne kroki dowodowe i prześledzić zachodzące między nimi związki logiczne. Niemniej, jakiekolwiek zjawiska wewnętrzne uznamy za możliwe *wyrażenia* myśli, można przedstawić argumenty analogiczne do poprzedniego na dowód, że to nie same te zjawiska składają się na rozumienie, lecz zdolność podmiotu myślącego do *posługiwania się* nimi, do wywoływania właściwych zjawisk we właściwych okolicznościach.

Przytoczone rozumowanie jest bardzo skróconą wersją argumentu Wittgensteina z *Dociekań filozoficznych*. Jeżeli jest ono poprawne, to próba zrozumienia myślenia na drodze badań zwanych „fenomenologicznymi" jest z gruntu chybiona; fenomenologowie nie przyjmują bowiem do wiadomości tego, że opisują wewnętrzne *wyrażanie* myśli, *rozumienie* zaś tego wyrażania – rozumienie własnych myśli – nie jest *zajściem* tego czy innego zdarzenia, lecz *zdolnością* wywoływania określonych zdarzeń mentalnych. Przykład człowieka udającego, że myśli po japońsku (i oszukującego w ten sposób japońskiego telepatę), wystarczy, aby wykazać bezskuteczność fenomenologicznego podejścia do problemu *rozumienia*. Jeśli bowiem nawet istnieje wykrywalna w introspekcji szczególna jakość, która występuje wtedy i tylko wtedy, gdy podmiot *rzeczywiście* rozumie swoje myśli (co wydaje się, gdy idzie o introspekcję, przypuszczeniem fałszywym), owa jakość jest przecież tylko *skorelowana* z rozumieniem, i, ponadto, nie jest wykluczone, że człowiek oszukujący japońskiego telepatę również wykazuje tę jakość, a *mimo to* nie rozumie ani słowa po japońsku.



Z drugiej strony, wyobraźmy sobie człowieka, który – co jest całkiem możliwe – nie prowadzi w ogóle „monologów wewnętrznych". Mówi nienaganną polszczyzną i ilekroć zapytać go o zdanie na jakiś temat, odpowiada dokładnie i szczegółowo. Nigdy jednak nie myśli (słowami, obrazami itd.), kiedy nie mówi głośno; nic też nie „przychodzi mu do głowy", wyjąwszy przypadki (rzecz jasna), w których słyszy swój własny głos oraz doznaje zwykłych wrażeń zmysłowych z otoczenia i przeżywa przy tym ogólne „uczucie rozumienia". (Może ma zwyczaj mówić na głos do siebie.) Kiedy pisze list lub idzie po zakupy itd., nie doświadcza wewnętrznego „strumienia świadomości"; niemniej działa inteligentnie i celowo, a jeśli ktoś podejdzie i zapyta „Co robisz?", udzieli w pełni sensownej odpowiedzi.

Można doskonale wyobrazić sobie kogoś takiego. Nikt nie zawaha się przed stwierdzeniem, że człowiek ów ma świadomość, że nie znosi rock and rolla (jeżeli często wyrażał zdecydowaną awersję do tego rodzaju muzyki) itd., tylko dlatego, że nie myśli świadomie wyjąwszy przypadki, w których mówi na głos.

Z dotychczasowych rozważań wynika, że (a) żaden zbiór zdarzeń mentalnych – wyobrażeń czy bardziej „abstrakcyjnych" czynności i jakości umysłowych – nie *stanowi* rozumienia; oraz (b) żaden zbiór zdarzeń mentalnych nie jest *niezbędny* dla rozumienia. W szczególności, *pojęcia nie mogą być identyczne z żadnymi przedmiotami w umyśle*. Stwierdziliśmy bowiem, że – założywszy, iż przedmioty w umyśle mają być wykrywalne w introspekcji – czymkolwiek są te przedmioty, nie muszą występować u człowieka, który rozumie stosowne słowa (a zatem ma w pełni wykształcone pojęcia), mogą natomiast występować u człowieka w ogóle pojęć pozbawionego.

Wracając do krytyki magicznych teorii odniesienia (temat, który interesował również Wittgensteina): stwierdziliśmy, z je-

dnej strony, że te „przedmioty w umyśle", które *dadzą się* wykryć w introspekcji – słowa, wyobrażenia, odczucia itd. – nie odnoszą się do żadnych określonych przedmiotów ze swej natury, tak samo (i z takich samych powodów) jak wizerunek nakreślony przez mrówkę nie odnosi się do Churchilla; z drugiej zaś – że próby postulowania szczególnych przedmiotów w umyśle, „pojęć", które *są* połączone związkiem koniecznym ze swoimi przedmiotami, a które potrafią wykryć jedynie wykwalifikowani fenomenologowie, są pod względem *logicznym* niestosowne; pojęcia są bowiem (przynajmniej po części) *zdolnością* do wywoływania określonych zdarzeń mentalnych, a nie samymi tymi zdarzeniami. Doktryna, w myśl której przedstawienia umysłowe z konieczności odnoszą się do rzeczy zewnętrznych, jest nie tylko błędną teorią przyrodniczą; jest również błędną fenomenologią, a także pomieszaniem pojęć.

# WIELE TWARZY REALIZMU

## Wykład I

## CZY JEST JESZCZE COŚ DO POWIEDZENIA NA TEMAT RZECZYWISTOŚCI I PRAWDY?

Eddington napomina nas, że człowiekowi z ulicy stół jawi się jako „ciało stałe", to znaczy, jako rzecz *przede wszystkim* masywna, pełna. Jednak fizyka odkryła, że stół składa się głównie z wolnych przestrzeni: że odległości międzycząsteczkowe są ogromne w porównaniu z promieniem elektronu lub jądra któregokolwiek z atomów, z których stół jest zbudowany. W odpowiedzi na to Wilfrid Sellars[1] zaprzeczył istnieniu stołów w postaci, w jakiej je sobie zwykle wyobrażamy (aczkolwiek jako przykładu użył kostki lodu, zamiast stołu). Potoczne wyobrażenie zwykłych, średniowymiarowych przedmiotów materialnych, jak stoły i kostki lodu (jawny wizerunek świata"), jest według Sellarsa po prostu *fałszywe* (aczkolwiek nie pozbawione pewnych walorów poznawczych: pod postacią „stołów" i „kostek lodu" jawny wizerunek świata przedstawia przedmioty rzeczywiście istniejące, nawet jeśli przedmioty te różnią się od stołów i kostek lodu profanów). Nie zgadzam się z Sellarsem, lecz mam nadzieję, że mi wybaczy, iż posłużę się jego poglądem, czy też samym faktem pojawienia się jego poglądu na filozoficznej scenie, do naświetlenia pewnych aspektów filozoficznego sporu o „realizm".

Przede wszystkim, pogląd Sellarsa ilustruje fakt, że Realizm

[1] *Science, Perception, and Reality*, Humanities Press, Atlantic Highlands NJ 1963.