

Tadeusz Skalski

Działanie neuronami

Filozofia Nauki 12/3/4, 87-106

2004

Artykuł został zdigitalizowany i opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

Tadeusz Skalski

Działanie neuronami

Wyrażenie *działanie neuronami* brzmi egzotycznie i — o ile mi wiadomo — nie było dotąd używane. Postaram się zatem wyjaśnić, jak należy je rozumieć i dlaczego — jak sądzę — jest potrzebne. Najpierw jednakże wprowadzę pewne rozróżnienie, które w moich wywodach odgrywać będzie ważną rolę.

1. ZACHOWANIE FIZYCZNE I ZACHOWANIE PODMIOTOWE

Pojęcie zachowania uchodzi za względnie jasne. Filozofowie spierają się w kwestii oceny i opisu tego, co dzieje się pomiędzy tzw. *wejściem* i *wyjściem*, natomiast bodźce i reakcje (czyli zachowania) nie wywołują emocji. Jedyny znany nam rodzaj niejasności czy też dwuznaczności dotyczący zachowania jest następujący. Otóż, zwykło się wprowadzać rozróżnienie na tzw. *zachowanie fizyczne* i *zachowanie podmiotowe* (Byrne 1994). Pierwsze obejmuje wszystko, co przydarza się ciału lub jego częściom; drugie dotyczy tylko tych zachowań, w odniesieniu do których zwykło się mówić o *działaniu*, *robieniu lub wykonywaniu* pewnych czynności i *sprawstwie*. Prosty przykład wyjaśni, jaka jest podstawa tego rozróżnienia.

Załóżmy, że ktoś w ciągu godziny co chwilę zadaje mi pytanie „Co robisz?“, a ja staram się na nie odpowiedzieć. Oto kilka przykładów moich odpowiedzi: *Włączam telewizor, Zamykam drzwi, Odbijam twój obraz na siatkówce, Czytam gazetę, Wydzielam kwasy żołądkowe, Wychodzę na balkon, Spadam na dół z przyspieszeniem 9, 81 m/s²*. Otóż, sądzę, że nikt lub prawie nikt nie miałby kłopotów z oddzieleniem odpowiedzi sensownych od jawnie absurdalnych. Mogę powiedzieć, że to ja włączam telewizor, ale nie mogę powiedzieć, że to ja wydzielam kwasy żołądkowe. Kwasy żołądkowe są po prostu wydzielane. Jest to proces, którego nie mogę kontrolować i na który nie mam żadnego wpływu. Rozróżnienie, o które nam idzie, uchodzi —

wszędzie tam, gdzie w grę wchodzi kontakty pomiędzy ludźmi — za rozumiałe samo przez się i niewymagające żadnych dalszych wyjaśnień. Opiera się na nim np. wymiar sprawiedliwości. Można ukarać Iksa za to, że tupnął on nogą i spowodował obsunięcie się lawiny, ale najsurowszy nawet prokurator nie domagałby się ukarania Iksa, gdyby wykazano, że podmuch wiatru przewrócił Iksa, a upadek spowodował obsunięcie lawiny. Ludzie miewają do siebie pretensje wyłącznie o to, co zrobili, i rozróżnienie to (w praktyce) nie następuje żadnych trudności. Pan Iks może mieć żal do pana Igreka, że ten nie oddał mu książki; nie może mieć jednak pretensji o to, że Igrekowi zrobiły się kamienie nerkowe (aczkolwiek może mu wyrzucać, że pił zbyt dużo herbaty). Rozróżnienie, o które nam idzie, jest całkowicie czytelne dla małych dzieci. Dzięki temu proces wychowywania dzieci jest w ogóle możliwy. Okrzyk matki *Co robisz!*, gdy Jaś opluwa właśnie Małgosię, jest sensowny. Inny egzemplarz tego samego okrzyku w sytuacji, gdy Jaś ma biegunkę lub gorączkę, byłby całkowicie absurdalny.

Rozróżnienie, o które nam idzie, bywa jednak (i to coraz częściej) kwestionowane. Bywa ono kwestionowane nie dlatego, że ludzie nagle przestali odróżniać to, co się wydarza, od tego, co oni sami robią, i nie dlatego, że odkryto jakiś krąg kulturowy, w obrębie którego nie jest ono czytelne, ani nawet dlatego, że pojawiła się jakaś mniejszość, której przedstawiciele mają pod tym względem inną orientację. Tym razem nic z tych rzeczy. Powód jest inny. Rozróżnienie, o które nam tu idzie, „pachnie” sprawstwem, a więc podmiotem, wolną wolą i w ogóle kartezyjizmem. Nam to jednak nie przeszkadza. Celem tej pracy nie jest naturalizacja umysłu. Dlatego będziemy odróżniać zachowanie fizyczne od podmiotowego i tylko w tym drugim przypadku używać określeń „działanie” lub „robienie”.

2. DZIAŁANIE NEURONAMI

Sądźmy, że najlepiej i najszybciej wyjaśnimy co mamy na myśli, odwołując się do dwóch konkretnych przykładów. Pierwszy dotyczy szczurów, drugi ludzi.

Przykład pierwszy. Do mózgu szczura wszczepiono 16 elektrod, dzięki którym rejestrowano czynności 30 neuronów (jednoczesny zapis jest ważny, gdyż aktywność jednego neuronu nie stanowi pełnej instrukcji umożliwiającej wykonanie odpowiedniego ruchu). Następnie naukowcy nauczyli szczura naciskania łapą dźwigni w celu otrzymania pokarmu, który był podawany za pomocą robota. Wreszcie (kluczowy moment!) skonstruowano sztuczną sieć neuronową, którą wykorzystano do rozpoznawania wzorców czynności neuronów występujących podczas naciskania dźwigni. Gdy sieć rozpoznawała odpowiedni wzorzec, uruchamiała (za pośrednictwem robota) dźwignię i zwierzę otrzymywało pokarm. Ramię automatu reagowało szybciej niż mięśnie kończyny zwierzęcia, więc dostawało ono pokarm przed naciśnięciem łapą dźwigni. Po pewnym czasie szczur zauważył, że naciskanie łapą dźwigni jest zbędne

i zaprzestał wykonywania tej czynności; niekiedy poruszał minimalnie łapą, a niekiedy nie wykonywał żadnego ruchu (Zucker 2000).

Założmy teraz, że nasz szczur uległ całkowitemu i nieuleczalnemu paraliżowi, ale, że mózg zwierzęcia pracuje bez zarzutu, a w każdym razie, że bez zarzutu (czyli tak jak przedtem) działa 30 neuronów, do których podłączono 16 elektrod. Nie ulega wątpliwości, że szczur mógłby w dalszym ciągu uruchamiać dźwignię (za pośrednictwem odpowiednich artefaktów) i tą drogą uzyskiwać pokarm. Jeżeli w ogóle chcemy mówić o czymś takim jak zachowanie szczura, to musimy uznać, że uruchamianie dźwigni jest częścią jego wyuczonego zachowania.

Przykład drugi. Naukowcy z Emory University wszczepili sparaliżowanej ofierze udaru mózgu układ scalony pozwalający jej za pomocą fal mózgowych sterować kursorem na ekranie komputera (Kurzweil 1999).

Zauważmy, że sterowanie kursorem na ekranie komputera jest, jak się wydaje, typowym zachowaniem podmiotowym. Zauważmy ponadto że ktoś, kto potrafi sterować zachowaniem komputera, potrafi też robić wiele innych rzeczy. Osoba taka może np. może kontrolować zachowanie wielu różnych urządzeń i obiektów, na przykład, w tzw. *inteligentnym domu*. Osoba taka może — *via* odpowiednie artefakty — gasić i zapalać światło, korzystać z telewizora, uruchamiać i zatrzymywać windę, otwierać i zamykać okna itd., itd. We wszystkich tych przypadkach mówienie o robieniu, działaniu lub zachowaniu podmiotowym wydaje się całkowicie uzasadnione. Nie można przecież pozbawiać ludzi podmiotowości tylko na tej podstawie, że nie są w stanie poruszać kończynami!

Rozprawianie o zachowaniu ludzi całkowicie sparaliżowanych może się jednak wydać dziwactwem i to nie tylko dlatego, że nie jesteśmy do tego przyzwyczajeni. Ilekroć mówimy o zachowaniu podmiotowym, o robieniu czegoś, automatycznie szukamy pośrednictwa ręki, nogi, głowy, języka itd. Jeżeli Jaś zgasił światło, to musiał posłużyć się jakąś częścią swego ciała — ręką, nogą, głową itd. Jakże więc całkowicie sparaliżowany Jaś mógł coś zrobić, skoro jego ręce, nogi, głowa są nieruchome?! Odczuwamy dyskomfort — brakuje nam pośrednictwa stosownego organu. I właśnie dlatego proponujemy nowe określenie. Od tej pory będziemy mówili o *działaniu neuronami* lub po prostu o *robieniu czegoś neuronami*. Przyjmujemy zatem, że Jaś może zgasić światło ręką lub nogą lub — niekiedy — neuronami.

Robienie czegoś neuronami lub przy pomocy neuronów chcemy rozumieć analogicznie do robienia innych rzeczy, np. do robienia różnych rzeczy przy pomocy rąk. Ktoś może, posługując się rękami, zapalić telewizor lub zgasić światło; ktoś inny może zrobić to samo, używając neuronów.

Oczywiście, w „świecie neuronów” dzieje się wiele różnych rzeczy, o których z całą pewnością nie mogę powiedzieć, że to ja je robię. Zapewne nie reguluję przy pomocy neuronów rytmu serca, nie zmieniam poziomu cukru we krwi itd. Wszystkie te rzeczy po prostu zachodzą lub wydarzają się, a ja nie mam na nie żadnego wpływu. W świecie dających się obserwować zachowań mego ciała też jednak dzieje się wiele różnych rzeczy, których nie kontroluję i na które nie mam żadnego wpływu, np. chra-

panie lub kichanie. W jednym i drugim przypadku mamy do czynienia zarówno z zachowaniami podmiotowymi, jak i fizycznymi.

2.1. Działanie neuronami i środowisko

Założmy, że 2-letni Jaś nauczył się po długim treningu wypowiadać poprawnie zdanie języka polskiego *Ala ma kota*. Aby wypowiadać to zdanie, Jaś musi wyemitować określoną falę głosową, wprawić powietrze w odpowiednie drgania. Na pytanie *Czym Jaś wprawił powietrze w odpowiednie drgania?*, odpowiadamy bez wahania — aparatem artykulacyjnym. Założmy dalej, że aparat artykulacyjny dziecka (np. w wyniku jakiejś choroby) uległ paraliżowi i Jaś jest teraz całkowicie niemy. Pytanie nasze brzmi: **Czy sparaliżowany Jaś umie (potrafi) artykułować „Ala ma kota”?** Otóż, naszym zdaniem odpowiedź jest pozytywna i powinna brzmieć: **Tak, Jaś w dalszym ciągu umie (potrafi) wyartykułować „Ala ma kota”**, pod warunkiem, oczywiście, że mózg dziecka nie uległ uszkodzeniu i działa tak jak przedtem. Wytłumaczmy się dokładniej.

Nauka artykułowania wypowiedzi *Ala ma kota* trwała przez wiele dni. Czego jednak uczył się Jaś. Proponujemy następującą interpretację. **Jaś w tym czasie opanowywał trudną sztukę posługiwania się skomplikowanym urządzeniem zwanym aparatem artykulacyjnym.** Wreszcie, nauczył się. Ilekroć tego zapragnie, artykułuje *Ala ma kota*. Teraz następuje paraliż aparatu artykulacyjnego. Czy Jaś w dalszym ciągu umie posługiwać się tym aparatem? Oczywiście, że tak. Jaś w dalszym ciągu umie zrobić (neuronami) wszystko to, co jest potrzebne (co do tej pory wystarczało) do tego, aby *Ala ma kota* zostało wyartykułowane. Przecież się tego nauczył. Cóż z tego, skoro urządzenie, którym się posługiwał jest niesprawne, „zepsuło się”. Sytuacja Jasia jest w gruncie rzeczy analogiczna do sytuacji jakiegokolwiek człowieka, który nauczył się obsługi skomplikowanego urządzenia. Założmy, że Joanna nauczyła się grać na harfie. A teraz wyobraźmy sobie, że struny harfy nagle sztywnieją, „ulegają paraliżowi”. Czy to znaczy, że Joanna zatraciła umiejętność gry na harfie i nie potrafi zagrać dobrze znanej melodii? Jaś nauczył się wygrywać na strunach głosowych *Ala ma kota*. Teraz, struny głosowe sztywnieją. Czy to znaczy, że Jaś nie umie już „zagrać” *Ala ma kota*? Oczywiście, że w dalszym ciągu potrafi to robić. Gdyby tylko dano mu sprawne struny głosowe, natychmiast wyartykułowałby *Ala ma kota*. Wydarzenia, które dochodzą do skutku zależą nie tylko od neuronów Jasia (nie tylko od tego, co on nimi robi), lecz także od środowiska, w którym Jaś nimi działa.

Spróbujmy wyrazić myśl z poprzedniego akapitu w nieco inny sposób. Weźmy pod uwagę trzech ludzi: Jana, Piotra i Michała. Jan i Piotr są normalnymi użytkownikami języka polskiego, Michał natomiast nigdy nie nauczył się żadnego języka. Przyjmijmy, że aparaty głosowe Jana i Michała uległy w pewnym momencie całkowitemu paraliżowi, podczas gdy aparat artykulacyjny Piotra jest w dalszym ciągu sprawny. Zastanówmy się, jak należałoby opisać tę sytuację z interesującego nas

punktu widzenia. Otóż, naszym zdaniem należałoby powiedzieć, że Jan (podobnie jak Piotr) w dalszym ciągu umie posługiwać się językiem polskim, podczas gdy Michał w dalszym ciągu tego nie potrafi. Jan w dalszym ciągu umie robić neuronami wszystko, co jest potrzebne do tego aby stosowne zdania j. polskiego były artykułowane. Podejście, którego bronimy jest następujące. Twierdzimy, że Jan i Piotr potrafią robić neuronami takie same rzeczy, dlatego, że potrafią robić takie same rzeczy w takich samych *środowiskach*. Faktyczną przyczyną różnicy jaka istnieje pomiędzy przymusowym milczeniem Jana i wymownością Piotra nie są ich różne umiejętności, ale odmiennosc środowisk, w których wypadło działać ich neuronom.. Gdyby neuronom Jana dano sprawny aparat artykulacyjny (innymi słowy, gdyby znalazły się one w środowisku, w którym znajdują się neurony Piotra) ten przemówiłby natychmiast; gdyby z kolei neurony Piotra znalazły się w środowisku zbliżonym do tego, w którym znajdują się neurony Jana (czyli, gdyby z kolei aparat artykulacyjny Piotra doznał paraliżu), Piotr natychmiast stałby się niemy. Jan i Piotr robią neuronami rzeczy bardzo podobne, a to, że efekty tych działań są tak odmienne jest następstwem odmiennosci środowisk.

Zaznaczmy, że środowisko, w którym działają neurony, nie musi być środowiskiem biologicznym. W opisanych wcześniej przypadkach — szczeniaka poruszającego neuronami ramieniem dźwigni, czy człowieka sterującego kursorem za pomocą fal mózgowych — środowiskiem, w którym działają neurony są artefakty.

2.2. Działanie neuronami i zachowanie

Jeżeli chcemy mówić o działaniu neuronami, to musimy uporać się z pewną niejasnością. Otóż, nasuwa się pytanie, co w tych okolicznościach jest zachowaniem. W normalnych warunkach poprzez zachowanie rozumie się zarówno ruchy ciała jak i skutki tych ruchów w odpowiednim środowisku. Jedne i drugie dają się obserwować i wcale często łącznie figurują w typowym opisie zachowania. Jeżeli skutki są takie same, ale poprzedzające je ruchy ciała różne, to i zachowania są różne. Jeżeli *Jaś wybił szybę*, to zupełnie inaczej opisujemy i oceniamy zachowanie Jasia, w zależności od tego czy *wybił ją ręką, nogą, czy głową*. Są to trzy różne zachowania. Jeżeli całkowicie sparaliżowany Jaś działałby neuronami, to obserwowalibyśmy tylko skutki tych działań, ich końcowy efekt. Nasuwa się w związku pytanie, czy w tym przypadku poprzez zachowanie Jasia będziemy rozumieć tylko końcowy efekt, dające się obserwować skutki jakie powoduje działalność Jasia w określonym środowisku, czy też poprzez zachowanie Jasia będziemy rozumieć łącznie, zarówno skutki, jak i ukryte przed obserwatorem i będące ich przyczyną „ruchy neuronów”. Proponujemy od tej pory poprzez zachowanie rozumieć tylko skutki, dające się obserwować zmiany w środowisku powodowane działalnością neuronów. Ograniczenie takie jest naturalne i zgodne z tradycyjnym rozumieniem pojęcia zachowania. Podążamy tu szlakiem przetartym przez *behawiorystów*, którzy zawsze twierdzili, że nie interesuje ich *układ*

nerwowy i że zachowaniem jest tylko to, co dzieje się w powszechnie dostępnym i dającym się obserwować świecie. Tak więc do zachowania kompletnie sparaliżowanego człowieka sterującego kursorem na ekranie komputera zaliczylibyśmy ruchy kursora, a zachowaniem szczura uruchamiającego neuronami dźwignię automatu byłoby zmiany w położeniu dźwigni. Nasuwają się jednak w związku z tym poważne wątpliwości.

Załóżmy, że Jan i Piotr nauczyli się — *via* odpowiednie i takie same artefakty — sterować neuronami kursorem na ekranie komputera. Załóżmy dalej, że Jan i Piotr, aby kliknąć tę samą ikonkę tego samego programu, umieszczoną w takim samym miejscu takiego samego ekranu robią swymi neuronami takie same rzeczy. Przyjmijmy, teraz, że pewnego dnia ktoś bez wiedzy i zgody Jana wymienił artefakty podłączone do jego neuronów na inne, sprzężone z telewizorem. Efekt jest taki, że Jan i Piotr w dalszym ciągu robią neuronami takie same rzeczy, z tym, że Piotr klika odpowiednią ikonkę odpowiedniego programu, natomiast Jan (ku swemu zdziwieniu) zapala telewizor. Ponieważ wcześniej uznaliśmy, że poprzez zachowanie będziemy rozumieć końcowy, dający się obserwować rezultat, to teraz powinniśmy uznać, że Jan i Piotr robią (neuronami) takie same rzeczy ale zachowują się inaczej — pierwszy zapala telewizor natomiast drugi uruchamia odpowiedni program. Odpowiadamy w sposób następujący. Jan i Piotr robią takie same rzeczy ale robią je **w innych środowiskach**. Środowisko, w którym działają neurony Jana uległo zmianie i teraz są to dwa różne środowiska. Gdyby ktoś bez mej wiedzy i zgody pozmieniał funkcje klawiszy na klawiaturze mego komputera, to ja robiłbym (rękami) to samo co zazwyczaj, ale efekty tych działań byłyby całkowicie odmienne. Ryba znajdująca się w wodzie i ryba wyrzucona na piasek robią takie same rzeczy, ale efekty ich działań są całkowicie odmienne. Nie ma w tym nic dziwnego.

3. DZIAŁANIE NEURONAMI A FILOZOFIA UMYŚLU

Staraliśmy się wyjaśnić jak należy rozumieć działanie neuronami i przekonywaliśmy, że mówienie o działaniu lub robieniu neuronami jest sensowne. Obecnie przechodzimy do filozoficznej części tej pracy. Zastanowimy się jak przedstawiają się niektóre filozoficzne koncepcje umysłu w kontekście rozważanego tutaj działania neuronami, a na koniec zajmiemy się osobą sprawcy.

3.1. Działanie neuronami i behawioryzm

3.1.1. UWAGI WSTĘPNE

Nie jesteśmy w stanie przedstawić wyczerpującej charakterystyki behawioryzmu ani przeprowadzić klasyfikacji różnych jego odmian. Wymagałoby to osobnego artykułu. Krótkie wprowadzenie jest jednak niezbędne. Otóż, zwykło się odróżniać dwa

rodzaje behawioryzmu — *psychologiczny* oraz *filozoficzny*. Ten drugi zwany bywa również *behawioryzmem analitycznym* lub *logicznym* (w dalszym ciągu na jego oznaczenie będziemy używać tych określeń wymiennie). Pierwszy był dziełem psychologów, drugi — filozofów. Behawioryzm psychologiczny rodził się i dojrzewał w laboratoriach, filozoficzny w fotelach zacisznych gabinetów; pierwszy skupiał się raczej na szczurach i gołębiach, drugi — najczęściej na ludziach. Nie były to jedyne różnice. Dla psychologów behawioryzm był w pierwszym rzędzie *programem badawczym* dążącym do odkrycia empirycznych praw łączących zachowanie żywych organizmów z tzw. *bodźcami* oraz *historią warunkowania*. Dla filozofów, behawioryzm oznaczał nade wszystko logiczną analizę pojęć codziennej psychologii. W dalszym ciągu interesować nas będzie tylko behawioryzm filozoficzny. To, co najbardziej charakterystyczne dla tej odmiany behawioryzmu, to przeświadczenie, że każde zdanie psychologiczne można w zasadzie przetłumaczyć (zachowując jego właściwy sens) na zdanie (lub zdania) o zachowaniu lub dyspozycjach do odpowiednich zachowań. Dyspozycje, dodajmy, muszą ujawniać się w dających się obserwować zachowaniach. Pomyłka kartezjańskiej psychologii, w tym ujęciu, polegała na błędzie logicznym. Wszystkie tzw. terminy mentalne lub psychologiczne opisują nie stany ducha, ale w gruncie rzeczy zachowania lub dyspozycje do odpowiednich zachowań. Tam, gdzie takich zachowań nie ma, psychologia traci przedmiot. Ideę tę w sposób niesłychanie obrazowy wyraził Ludwig Wittgenstein.

Spójrz na kamień i przedstaw sobie, że ma on doznania! — Mówimy sobie: jak można było w ogóle wpaść na pomysł, by *rzeczom* przypisywać *doznania*? Równie dobrze można by je przypisywać liczbom! — A teraz spójrz na trzepoczącą się muchę i trudność znika natychmiast; ból ma tu o co *zahaczyć*, podczas gdy przedtem wszystko — by tak rzec — było dla niego *gładkie* (Wittgenstein 1972, str. 142).

Przypisywanie doznań kamieniom to, wedle Wittgensteina, nie tyle szalona fantazja nie mająca żadnego oparcia w naszej wiedzy o świecie, co — raczej — pogwałcenie semantycznych reguł języka psychologii. Reguły te dotyczą zachowań. Tam, gdzie nie ma zachowania, język psychologii traci przedmiot. Jeżeli ktoś twierdzi, że *świnie fruwać*, popełnia błąd rzeczowy; jeżeli utrzymuje, że *kamienie cierpią*, popełnia błąd logiczny.

3.1.2. DZIAŁANIE NEURONAMI I FAŁSZYWOŚĆ BEHAVIORYZMU

Jeżeli prezentowane tutaj podejście jest trafne, to behawioryzm jest stanowiskiem z gruntu fałszywym. Przy czym, to co godne uwagi, to nie tyle fałszywość behawioryzmu (o tym i tak jest przekonana większość współczesnych filozofów umysłu), co raczej łatwość z jaką można to pokazać.

Weźmy pod uwagę człowieka dotkniętego całkowitym i nieuleczalnym paraliżem. Osobnik taki myśli, pragnie, cierpi..., ale jego życie wewnętrzne nie posiada żadnego zewnętrznego wyrazu. Co się tyczy zachowania jest on podobny kamieniom. Predykaty psychologiczne nie mają się „o co zahaczyć”. Konsekwentny behawiorysta —

można argumentować — winien uznać, że przypisywanie takiemu człowiekowi doznań jest równie absurdalne, jak przypisywanie ich kamieniom. Nie ulega jednak wątpliwości, że wniosek taki byłby fałszywy. I to nie dlatego, że nauczyliśmy się leczyć wszystkie nieuleczalne paraliże, ale dlatego, że — *via* odpowiednie artefakty — tzw. życie wewnętrzne ludzi dotkniętych nieuleczalnym paraliżem może znajdować i znajduje swój zewnętrzny wyraz.

Czy jednak behawiorysta nie mógłby się obronić? Zgoda, mógłby powiedzieć, myliłem się. Była to jednak pomyłka rzeczowa, a nie filozoficzna. Świat się zmienił, a wraz z nim zmienił się sens filozoficznych argumentów. Teraz twierdzą, że człowiek dotknięty całkowitym paraliżem może myśleć i cierpieć. Twierdzą tak jednak tylko dlatego, że zaczął on się zachowywać. Dyspozycje do odpowiednich zachowań (np. ruchy kursora na ekranie komputera) tkwiły w nim od dawna; teraz zostały ujawnione i odpowiednie predykaty psychologiczne mają się o co „zahaczyć”. Wbrew tej odpowiedzi behawiorysty sądzimy, że behawioryzm jest jednak nie do obrony i można go ostatecznie zagwoździć. Rzecz w tym, że w przypadku ludzi dotkniętych całkowitym paraliżem jakikolwiek związek pomiędzy zachowaniem i psychologią zostaje przecięty.

Wyobraźmy sobie człowieka całkowicie sparaliżowanego w inteligentnym domu. *Via* odpowiednie artefakty zmyka on okna, otwiera drzwi, zapala i gasi telewizor. Załóżmy, że umawia się on z opiekunem, że *ilekroć będzie odczuwał ból, to zapali i zgasi światło*. Zapalanie i gaszenie światła jest sygnałem bólu. Jest nim jednak tylko dla opiekuna i tylko na czas obowiązywania umowy. Równie dobrze sygnałem bólu mogłoby być zapalenie i zgaszenie telewizora lub zamknięcie i otwarcie okna; cokolwiek zresztą. Związek pomiędzy *bólem* i *zachowaniem bólowym* staje się dowolny. Określenie *typowe zachowanie bólowe* całkowicie traci sens. Próba definiowania bólu przy pomocy zachowania wedle recepty *X odczuwa ból wtedy i tylko wtedy, gdy X zachowuje się tak a tak* byłaby w tych okolicznościach czymś absurdalnym. Behawioryzm w tym przypadku byłby stanowiskiem bezsensownym. Ból może się „zahaczyć o cokolwiek”.

W jednej ze swych prac Hilary Putnam, chcąc wykazać, że związek pomiędzy *bólem* i *zachowaniem bólowym* nie ma charakteru definicyjnego, wprowadza fikcyjne postaci *superspartan* (Putnam 1965). Superspartanie wdrażani są (w sposób o wiele skuteczniejszy niż historyczni Spartanie) do nie ujawniania (przy pomocy zachowań) bólu. Jeżeli, powiada autor, przyłożymy superspartaninowi do gołych pleców rozpalony do czerwoności pogrzebacz, to ten w żaden sposób nie ujawni zachowaniem swego cierpienia. Morał płynący z tej historyjki jest między innymi taki, że zachowanie jest uzależnione od kręgu kulturowego. Sądzimy jednak, że nasz przykład jest bardziej przekonujący.

Jeżeli mówienie o zachowaniu ludzi całkowicie sparaliżowanych ma sens, to behawioryzm nie ma sensu.

3.2. Działanie neuronami i funkcjonalizm

3.2.1. UWAGI WSTĘPNE

Behawioryzm nie jest już modny. Pojawiło się kilka nowych koncepcji, a mianowicie tzw. *teoria identyczności* w dwóch co najmniej wariantach (Smart 1962; Davidson. 1970), *eliminacjonizm* (Churchland, P.M. 1981; Churchland, P.M. 1994) i wreszcie *funkcjonalizm*. Ostatnia wśród wymienionych recept na umysł — funkcjonalizm — zasługuje na szczególną uwagę. Większość analitycznych filozofów umysłu oraz adeptów nowopowstałej interdyscyplinarnej nauki — *nauki poznawczej (cognitive science)*, to właśnie funkcjonałiści. Funkcjonalizm to obecnie główny nurt, dominująca koncepcja, wspólna płaszczyzna porozumienia. Wydaje się, że pełni on taką rolę, jaką niegdyś pełnił behawioryzm. Pomijając etap przejściowy (teorię identyczności i eliminacjonizm), można byłoby rzec, iż behawioryzm został ostatecznie wyparty przez funkcjonalizm.

Główna idea funkcjonalizmu jest bardzo prosta. Cieleśna maszyna (*alias* człowiek, zwierzę, kosmita lub robot) powiada się, posiada różne stany wewnętrzne. Owe stany wewnętrzne powiązane są siecią związków przyczynowo skutkowych. Coś je wywołuje (zewnętrzne bodźce lub inne stany), a i one same coś powodują (powstawanie innych wewnętrznych stanów lub zachowań). Stany psychologiczne lub mentalne miałyby być identyfikowane w tym ujęciu nie poprzez to, czym są lub wydają się być, ale poprzez sieć przyczynowo skutkowych powiązań. Ulubionym przykładem funkcjonalistów jest *ból*. Analizując fenomen bólu nie należy ulegać tradycyjnym pokusom. Po pierwsze, nie powinno się utożsamiać bólu z jego aspektem wrażeniowocuciowym, czyli jego fenomenalną istotą, gdyż ta jest bez znaczenia. A poza tym prowadzi na manowce kartezjanizmu. Po drugie, nie należy utożsamiać bólu z jego fizjologicznym podłożem. Jest to błąd, który często popełniają tradycyjnie zorientowani materialiści (teoria identyczności!). Postępując w ten sposób dyskryminowalibyśmy istoty wykonane z innych materiałów, np. kosmitów i roboty. Materialne podłoże — mogą to być związki białka lub silikon lub cokolwiek innego — jest nieistotne. Po trzecie, analizując ból, nie należy utożsamiać go z zachowaniem. Zachowanie jest ważne, ale nie można się do niego ograniczać. Czymże jest więc ból? Otóż, ból to pewien stan wewnętrzny, który jest przez coś powodowany (na ogół przez zewnętrzne bodźce) i który sam z kolei coś powoduje. To co istotne, to nie czuciowo-wrażeniowa strona bólu, ale jego przyczyny i skutki. W zakres skutków — zaznaczmy — wchodzi nie tylko zachowanie (choć jest ono niezmiernie ważne), ale ponadto wpływ jaki ból wywiera (ewentualnie) na inne stany wewnętrzne, np. na różne tzw. *postawy propozycjonalne*. Ból, np. ten oto ból po ukłuciu tą oto szpilką, jest zatem dopiero w pełni określony poprzez ogół swych przyczynowo-skutkowych ról jakie pełni w całej tzw. *architekturze funkcjonalnej* organizmu.

Ból, jako się rzekło, to tylko przykład. Funkcjonalistyczna ideologia w mocnej wersji dotyczy, *mutatis mutandis* wszystkich tzw. stanów psychologicznych. Psychi-

ka, albo tzw. życie mentalne jest w pełni zdeterminowane poprzez sieć przyczynowo skutkowych powiązań. Dwa różne pod względem fizycznym obiekty, posiadające taką samą architekturę funkcjonalną (przyjmuje się tutaj jako rzecz samo-oczywistą, że dwa różne obiekty, np. człowiek i robot, mogą być nosicielami takiej samej architektury funkcjonalnej; sztandarowy przykład to różne obiekty tak samo oprogramowane) są zatem identyczne pod względem psychologicznym, a w każdym razie zostałyby identycznie opisane przez funkcjonalistę.

Główne przyczyny popularności i atrakcyjności prezentowanej ideologii leżą, jak się wydaje, poza filozofią. Funkcjonalizm jest, w znacznej mierze, filozoficzną reakcją na rewolucję naukowo techniczną, a w szczególności na rozwój badań nad *sztuczną inteligencją*. Wedle słów prekursora tego sposobu myślenia, Hilarego Putnama:

Funkcjonalizm głosi, że jesteśmy analogiczni do komputerów i że nasze psychologiczne stany są po prostu „funkcjonalnymi stanami”, tzn. są one stanami, które figurowałyby w idealnym opisie naszych programów (...) „Funkcjonalizm” widzi w nas automaty; tzn. komputery, którym przytrafiło się być wykonanym z krwi i kości (Putnam 1994, str. 507; tłumaczenie autora).

Aczkolwiek Putnamowska formuła na określenie funkcjonalizmu wydaje się zbyt restrykcyjna i zapewne nie wszyscy funkcjoniści byliby skłonni ją zaakceptować, to nie ulega wątpliwości, że rewolucja komputerowa odegrała kluczową rolę w powstaniu i rozwoju funkcjonalistycznej ideologii. Funkcjonalizm pod tą właśnie postacią przeniknął zresztą do świata filmu i literatury. Odpowiednio zaprogramowany robot, cyborg — istota zdolna do samodzielnego myślenia, okazywania i posiadania uczuć i wreszcie do odczuwania bólu — to jedna z ikon kultury masowej.

Wcześniejsze recepty filozofów analitycznych na umysł — behawioryzm lub różne wersje teorii identyczności znane były i dyskutowane w wąskich gronach fachowców. Funkcjonalizm w swej komputerowej wersji — umysł to program — to dla ludzi przed trzydziestką nieomal teoria zdrowego rozsądku.

3.2.2. DZIAŁANIE NEURONAMI I FAŁSZYWOŚĆ FUNKCJONALIZMU

Funkcjonalizm bywał wielokrotnie atakowany. Najczęściej wskazywano na to, że nie rozwiązuje on problemu tzw. *qualiów* oraz, że kiepsko radzi sobie z tzw. *intencjonalnością*. (Ilość filozofów, którzy zajmowali się tymi sprawami jest tak ogromna, że nie jesteśmy ich w stanie omawiać ani nawet wymieniać.) Natomiast, o ile wiemy, jak dotąd nie próbowano kwestionować funkcjonalizmu odwołując się do zachowania. To, że obiekty wykonane z różnych materiałów (np. człowiek i robot) mogą zachowywać się podobnie, a nawet identycznie uchodzi — mamy takie wrażenie — za rzecz samo-oczywistą. Wielu filozofów zastanawia się i ma wątpliwości co do tego, czy odpowiednio spreparowany artefakt (tzn. odpowiednio oprogramowany lub o odpowiedniej architekturze funkcjonalnej), będzie istotą świadomą (np. czy będzie miał stosowne *qualia*, a więc czy, na przykład będzie odczuwał ból). Nikt natomiast nie

kwestionuje tego, że będzie się zachowywał w odpowiedni sposób. Mówiąc krótko, co się tyczy zachowania, funkcjonalizmowi nie można nic zarzucić!

Jednym z najbardziej nieprzejednanych i wpływowych krytyków funkcjonalizmu jest John Searle. Od kilkunastu lat toczy się spór pomiędzy tym filozofem z jednej strony i zwolennikami tzw. *mocnej wersji sztucznej inteligencji* z drugiej. Nawet Searle zdaje się jednak sądzić, że, co się tyczy zachowania, funkcjoniści nie mają sobie nic do zarzucenia. Charakterystyczny jest pod tym względem fragment jednej z ostatnich prac tego autora. W ostatnim rozdziale książki *The Mystery of Consciousness* filozof ten omawia i odpiera różne obiekcje dotyczące bronionej przez niego koncepcji świadomości, zwanej *biologicznym materializmem*. W pewnym momencie wyimaginowany polemista powiada mniej więcej coś takiego:

Gdybyśmy zbudowali kierowanego przez komputer robota, który zachowywałby się tak jakby był świadomy, to dlaczego nie mieli byśmy uznać, że jest świadomy (Searle 1997, str. 204; tłumaczenie autora).

Ponieważ interesuje nas tutaj nie świadomość, ale zachowanie, skupię się tylko na tym fragmencie repliki Searle'a, który dotyczy zachowania. Otóż, jeżeli dobrze rozumiem Searle'a, to jego stanowisko jest mniej więcej następujące. Oczywiście, można skonstruować obiekty, które będą zachowywać się tak jakby były świadome, ale cóż z tego. Zachowanie jest zupełnie nieistotne. Pogląd ten wsparty jest przykładami. Klepsydra, na przykład, mierzy czas równie dobrze jak zegarek noszony na ręce, ale jej zewnętrzne zachowanie jest nieistotne dla zrozumienia wewnętrznej budowy zegarka na rękę. Można zbudować komputer, który będzie krzyczał *ovej*, gdy mocniej uderzymy w klawisz, ale czy jest jakiś powód, aby uważać, że ma on wrażenia bólowe. Mówiąc krótko, zbudowanie artefaktu, który zachowywałby się tak jak człowiek, to rzecz może i trudna, ale możliwa, jednakże nie ma to nic do rzeczy. Zewnętrzne zachowanie jest nieistotne.

W dalszym ciągu będziemy starali się pokazać, że zbudowanie artefaktu, który będzie zachowywał się tak jak pewien człowiek jest niemożliwe. Mówiąc dokładniej, będziemy starali się pokazać, że zbudowanie artefaktu, który będzie zachowywał się tak samo jak mózg pewnego człowieka nie jest rzeczą możliwą. Odwołamy się w tym celu do działania neuronami.

Założmy, że pewien człowiek o imieniu Jan, w wyniku nieszczęśliwego wypadku jest całkowicie sparaliżowany. Założmy dalej, że mózg Jana nie uległ uszkodzeniu i działa tak jak przedtem. Mówiąc, że mózg Jana działa tak jak przedtem chcemy powiedzieć tylko tyle, że gdyby Jan odzyskał władzę w rękach, nogach, strunach głosowych itd., np. w wyniku odpowiednich interwencji chirurgicznych, to natychmiast mógłby rozpocząć normalne życie i robić to co zazwyczaj robił przed wypadkiem. Niestety, neurochirurgia jest nauką zbyt słabo rozwiniętą i najlepsi nawet specjaliści w tej dziedzinie nie potrafią pomóc naszemu bohaterowi. Ciało Jana — tak brzmi ich wyrok — na zawsze pozostanie martwe. Sytuacja nie jest jednak całkowicie beznadziejna. Wszak mózg Jana żyje i działa w dalszym ciągu. Jan jest w dalszym ciągu

zdolny do robienia wielu różnych rzeczy. Gdyby tylko udało się — mówiąc językiem Arystotelesa — zaktualizować potencje drzemiące w zdrowym i prawidłowo funkcjonującym mózgu Jana! Załóżmy, że zespół specjalistów z różnych dziedzin (neurobiologów, informatyków, psychologów itd.) podejmuje się, przynajmniej częściowo, rozwiązać ten problem. Tworzą oni dom będący z wyglądu repliką domu Jana. Nowy dom, w przeciwieństwie do starego, jest jednak silnie nasycony elektroniką; jest to tzw. inteligentny dom. Żyjący i zdolny do działania mózg Jana — *via* najrozmaitsze artefakty — zostaje podłączony do wielu różnych urządzeń. Rozpoczyna się długa i mozolna nauka. Jan stara się sterować neuronami (można tu użyć wielu innych określeń, np. mózgiem, myślą, siłą woli itd.) wieloma różnymi urządzeniami — kierować wózkiem na którym znajduje się jego ciało, zapalać i gasić światło w różnych pomieszczeniach, otwierać i zamykać drzwi i okna, korzystać z telewizora itd., itd. Załóżmy, że nauka trwa przez wiele miesięcy i że towarzyszą jej efekty charakterystyczne dla procesu uczenia się obsługi nowych urządzeń (tzn. Jan stopniowo opanowuje różne umiejętności, a w miarę trwania nauki myli się coraz rzadziej). Choć Jan jest pilnym i zdolnym uczniem, to sterowanie pewnymi urządzeniami okazuje się być bardzo trudne, a niekiedy wręcz niemożliwe. Wina nie zawsze leży po stronie Jana. Artefakty, na które Jan stara się wpłynąć są zbyt mało wrażliwe, albo reagują nie na to na co powinny reagować. Używając przyrodniczo-cybernetycznego żargonu można byłoby powiedzieć, że moce przyczynowe neuronów Jana są niekompatybilne z niektórymi artefaktami. Specjalny zespół pilnie śledzi wszystkie te trudności i w miarę trwania nauki elektroniczne środowisko, w którym wypadło działać neuronom Jana jest modyfikowane i ulepszone, zmieniane na coraz bardziej przyjazne neuronalnym możliwościom Jana. Wreszcie, po wielu miesiącach, nauka dobiega końca. Ostateczny efekt jest zadziwiający. Jan jest w stanie samodzielnie zrobić wiele różnych rzeczy — zapala i gasi światła w dowolnym pomieszczeniu, otwiera i zamyka dowolne drzwi oraz okna, samodzielnie korzysta z telewizora itd., itd. Zauważmy, że wszystkie te rzeczy Jan robił i przed wypadkiem, posługiwał się jednak rękami i nogami. Teraz — dzięki pośrednictwu wielu różnych artefaktów — robi je neuronami. Czy jednak organ jest tutaj taki istotny? Ręce czy neurony — co za różnica. Liczy się końcowy efekt. Ważne, że okno jest zamknięte. Załóżmy na dodatek, że dzięki specjalnemu syntezaotorowi mowy (zastępującemu aparat artykulacyjny) można się z Janem porozumieć i upewnić, że rzeczywiście robi to co chce. Dodajmy, że Jan potrafi być na swój sposób złośliwy (np. unieruchomić specjalnie windę), niegrzeczny (zamknąć drzwi przed nosem nieproszonemu gościowi), uparty itd., itd. Mówiąc krótko, dające się obserwować efekty działalności Jana posiadają wszelkie znamiona podmiotowego zachowania. W każdym razie, wydaje się, że w opisanej sytuacji można mówić zarówno o zachowaniu Jana (mózgu Jana) jak i o środowisku. Fakt, że środowisko to jest środowiskiem elektronicznym nie powinien nas odstraszać. Środowisko, w którym wypadło działać ludziom pełnosprawnym jest także w coraz większym stopniu środowiskiem elektronicznym.

Przejdźmy obecnie do drugiej części naszego wywodu. Załóżmy, że (kierując się tym razem względami czysto teoretycznymi) postanawiamy skonstruować obiekt będący *funkcjonalnym równoważnikiem* Jana. Okoliczność, iż Jan to po prostu mózg upraszcza nasze zadanie. Nie musimy się kłopotać o ręce i nogi. Wystarczy, że zbudujemy funkcjonalny równoważnik mózgu. Obiekt taki zwykle się określać mianem *symulatora mózgu*. Przyswójmy sobie ten żargon i określmy nasze zadanie. Chcemy zbudować symulator mózgu Jana!

Czym jednak jest symulator mózgu? Otóż, nie wiemy. Ci, którzy upowszechniają tego rodzaju idee, często wyobrażają sobie to jakoś tak. Zeskanujemy wszystkie informacje zawarte w mózgu Jana (jakimś superskanerem n-tej generacji) i wgrajmy je do komputerowego medium. Otrzymamy w ten sposób symulator mózgu Jana. (Wszak mózg służy do myślenia, a myśl to informacja). Brzmi to efektownie, ale praktyczna wartość tej wskazówki jest równa zero. Kłopot polega na tym, że nie wiemy, co właściwie mamy skanować, gdy każą nam skanować informacje zawarte w mózgu Jana. Nie przejmujemy się tym. Naszym zadaniem nie jest zdefiniowanie symulatora mózgu. My chcemy pokazać, że symulator mózgu nie będzie się nigdy zachowywał tak jak mózg. Aby to pokazać, nie trzeba wiedzieć, czym jest symulator mózgu. Wystarczy wiedzieć, że nie jest on repliką mózgu.

Weźmy pod uwagę symulator mózgu Jana i oznaczmy go przez **SJ**. Niech **SJ** spełnia wszelkie warunki, jakie tylko chcieliby nałożyć na niego funkcjoniści. Niech więc **SJ** posiada taką samą (jak mózg Jana) architekturę funkcjonalną, zawiera takie same informacje, niech przetwarza je w taki sam sposób (cokolwiek miałyby to znaczyć) itd. Jeżeli pewien funkcjonalista chciałby, aby **SJ** spełniał jakieś dodatkowe warunki, to chętnie je zaakceptujemy. Na **SJ** nakładamy tylko jedno ograniczenie. Otóż **SJ** nie jest repliką mózgu Jana, ale obiektem wykonanym z jakichś innych tworzyw. Ograniczenie takie jest oczywiste i powinno ono ucieszyć rzeczników funkcjonalistycznej ideologii.

Zwolennicy idei symulatora mózgu zgadzają się w jednym. Symulator mózgu Jana nie musi być jego repliką, tzn. kopią, tzn. kolejnym ludzkim mózgiem niczym się nieróżniącym od oryginału. Każdy (lub prawie każdy) zgodzi się z tym, że bliźniaczy (pod każdym względem) mózg będzie się zachowywał w bliźniaczy sposób. Ci, którzy mówią o skanowaniu informacji i symulatorze mózgu, nie mają, rzecz jasna, na myśli kolejnego ludzkiego mózgu. Byłoby to jakieś szyderstwo z idei symulatora mózgu i funkcjonalizmu w ogóle. Sednem funkcjonalizmu, tym co się podkreśla i akcentuje, jest ta oto okoliczność, że symulator jednego i tego samego mózgu może zostać urzeczywistniony na wiele różnych sposobów. Jest to owa słynna wielość realizacji. Nie musi on być wykonany z biologicznych neuronów. Tworzywo nie jest ważne — liczy się architektura funkcjonalna, oprogramowanie. Symulator mózgu można zrobić z czegokolwiek — z puszek po piwie lub z miliarda Chińczyków (Johnson-Laird 1988, Block 1990, Pylyshyn 1984). Przyjmijmy więc, że **SJ** nie jest repliką mózgu Jana, ale obiektem wykonanym z jakichś innych tworzyw. Teraz możemy ostatecznie sprecyzować nasze pytanie. Brzmi ono jak następuje: *Czy **SJ** oraz mózg*

Jana będą się zachowywać identycznie? Otóż, jeżeli przez zachowanie mamy rozumieć dające się obserwować efekty działalności *SJ* oraz *mózgu Jana* (a coś innego moglibyśmy rozumieć?!), to odpowiedź musi być negatywna. *SJ* i *mózg Jana* nie będą się zachowywać identycznie. Co więcej, można łatwo pokazać, że *SJ* i *mózg Jana* będą się zachowywać całkowicie odmiennie.

Weźmy pod uwagę dwa inteligentne domy D1 i D2. Załóżmy, że D1 i D2 są takie same pod każdym względem. D1 to dobrze nam znany inteligentny dom, w którym znajduje się *mózg Jana*, a D2 to jego elektroniczny klon wybudowany specjalnie na użytek *SJ*. Podłączmy *SJ* do D2 w odpowiedni sposób (najlepiej niech to zrobi jakiś funkcjonalista). Czy *mózg Jana* i *SJ* będą się zachowywać w taki sam sposób; tzn. czy *mózg Jana* i *SJ* będą w tym samym momencie otwierać i zamykać takie same drzwi, gasić takie same światła w takich samych pokojach itd., odpowiednio w D1 i D2? Można w to wątpić. Wszak środowisko, w którym działają neurony Jana, zostało specjalnie skonstruowane na użytek neuronów Jana i jest ono wrażliwe na całkiem określone wydarzenia fizyczne zachodzące w tych neuronach. W *SJ* zachodzą inne fizyczne wydarzenia, posiadające inne moce przyczynowe i wywierające inny wpływ na otaczające je elektroniczne środowisko. Okoliczność, że *SJ* jest symulatorem *mózgu Jana*, że posiada taką samą architekturę funkcjonalną, że zawiera takie same programy, informacje (cokolwiek miałyby to znaczyć), że zachodzi jakaś odpowiedniość pomiędzy wydarzeniami dochodzącymi w jednym i drugim obiekcie nie ma tu nic do rzeczy. Wydarzenia fizyczne zachodzące w *mózgu Jana* są całkowicie innymi przyrodniczymi wydarzeniami niż odpowiadające im wydarzenia zachodzące w *SJ*. Tego co powodują jedno nie są w stanie spowodować drugie i na odwrót.

Założmy jednak, że w wyniku cudu lub nadzwyczajnego zbiegu okoliczności *SJ* i *mózg Jana* zachowują się tak samo, tzn. w tym samym momencie w dwóch różnych domach zapalają i gaszą bliźniacze światła; otwierają i zamykają odpowiadające sobie bliźniacze drzwi itd. Tak się jakoś dziwnie złożyło, że wydarzenia zachodzące w *mózgu Jana* i *SJ*, aczkolwiek różne pod względem fizycznym, są w stanie na to akurat środowisko (przypomnijmy takie samo dla *mózgu Jana* i *SJ*) wywrzeć wpływ skutkujący bliźniaczymi końcowymi efektami. Otóż, gdyby nawet tak się złożyło (a jest to o wiele za mało, aby obronić funkcjonalizm) to i tak nie można byłoby powiedzieć, że *mózg Jana* i *SJ* zachowują się tak samo. Można byłoby i należało powiedzieć tylko tyle, że *mózg Jana* i *SJ* zachowują się tak samo w tym jednym konkretnym środowisku. W innym środowisku mogłyby się zachowywać różnie.

Jeśli idea tożsamości lub identyczności zachowań ma mieć sens, to musi obejmować wszelkie możliwe środowiska. Z całą pewnością nie można powiedzieć o dwóch różnych ludziach, zwierzętach, robotach itd., że zachowują się tak samo, o ile zachowują się tak samo tylko w jednym konkretnym środowisku. Jeżeli dwa obiekty zachowują się tak samo na pustyni, ale inaczej w wodzie; tak samo pod prysznicem, ale inaczej na ulicy, to rzecz jasna nie można o nich powiedzieć (z pominięciem wszelkiej relatywizacji), że po prostu zachowują się tak samo. Przyjmujemy odąd, że dwa obiekty zachowują się tak samo (niezależnie od tego, czy są to ludzie,

roboty, zwierzęta, mózgi czy symulatory), jeżeli zachowują się tak samo we wszelkich możliwych (i takich samych) środowiskach. Jeżeli zgodzimy się, że tak właśnie należy rozumieć identyczność zachowań, to nie ulega wątpliwości, że *mózg Jana* i *SJ* nie zachowują się tak samo. Aby wyciągnąć taki wniosek, wystarczy wiedzieć, że wydarzenia zachodzące w *mózgu Jana* i w *SJ* są wydarzeniami różnymi pod względem fizycznym. Aby osiągnąć pożądany efekt (tzn. odmiennosc zachowań), wystarczy więc uwrażliwić środowisko na te różnice. Z całą pewnością można to zrobić na wiele sposobów. Najprostszym pomysłem, który przychodzi laikowi do głowy, byłby następujący. *Mózg Jan* i *SJ* są wykonane z różnych materiałów (wedle zasady: *hardware* nie ma znaczenia, liczy się tylko *software*), wystarczyłoby więc uwrażliwić środowisko na „chemię” odpowiadających sobie wydarzeń. Jeżeli odpowiednie warunki chemiczne nie byłyby spełnione, to pośredniczące artefakty nie byłyby (mimo wszystko) uruchamiane. W ten sposób *mózg Jana* w dalszym ciągu gasiłby i zapalał światła, zamykał i otwierał drzwi, podczas gdy *SJ* w takim samym pod każdym względem środowisku nie robiłby niczego. Manipulując odpowiednio tymi warunkami i środowiskiem (długo by o tym pisać), łatwo można byłoby uzyskać bardziej spektakularne efekty. *Mózg Jana* zapalałby światło, a *SJ* (w takim samym środowisku) uruchamiał telewizor; *mózg Jana* otwierałby okno, a *SJ* (w tym samym czasie i w takim samym środowisku) otwierałby drzwi itd. Z całą pewnością nie są to takie same zachowania.

Stanowisko, które prezentujemy może budzić różne obiekcje. Rozważymy zaledwie kilka.

Obiekcja pierwsza. Autor tego tekstu zgodzi się zapewne, że gdyby *mózg Jana* i *SJ* umieszczono w różnych środowiskach elektronicznych, innymi słowy, gdyby podłączono je pod różne artefakty pośredniczące, to wówczas identyczność zachowań na poziomie zapalania świateł i otwierania okien byłaby już możliwa do osiągnięcia. To, gdzie kończy się *SJ*, a zaczyna środowisko, to jednak rzecz umowna. Dlaczego nie mielibyśmy przyjąć, że artefakty pośredniczące stanowią integralną część symulatora, a środowiskiem są po prostu drzwi, okna, telewizor...? Odpowiadamy na to jak następuje. Istotnie, to gdzie kończy się symulator, a zaczyna środowisko, to w znacznej mierze kwestia decyzji. Rzecz jednak w tym, że to gdzie kończy się *mózg Jana*, a zaczyna środowisko, w którym on działa, to nie jest kwestia umowy. *Mózg Jana* kończy się tam, gdzie się kończy, a *SJ* ma być symulatorem tego właśnie mózgu, a nie czegoś innego. *Mózgu Jana* zaś nie można „podłączyć” bezpośrednio do drzwi i okien. Niezbędne są artefakty pośredniczące i właśnie one stanowią bezpośrednie środowisko neuronów Jana. Identyczność zachowań na poziomie drzwi i okien osiągnięta tą drogą, byłaby więc osiągnięta kosztem umieszczenia *mózgu Jana* i *SJ* w całkowicie różnych środowiskach. Dla mózgu byłyby to artefakty pośredniczące, a dla symulatora — drzwi i okna. Identyczność taka jest niewiele warta i nie świadczy bynajmniej o podobieństwie wchodzących w grę obiektów. Gdyby ktoś twierdził, że Piotr i Stefan są równie silni dlatego, że pierwszy podnosi 300 kg na Ziemi, a drugi 300 kg na księżycu, to uznalibyśmy to za nonsens. Dlaczego tym ra-

zem mielibyśmy stosować jakieś inne miary? Gdybyśmy raz uznali, że środowisko nie jest istotne, a ważne są tylko końcowe efekty, to wówczas manipulując odpowiednio różnymi artefaktami pośredniczącymi osiągalibyśmy zadziwiające rezultaty. Mogłoby się wówczas okazać, że wcale dobrym symulatorem *mózgu Jana* jest *mózg pewnego szczura*, a dobrym symulatorem *mózgu szczura* jest *żołądek Jana*.

Obiekcja druga. Jeżeli prezentowana tutaj koncepcja jest trafna, to należałoby uznać, że jakiegokolwiek dwa obiekty, różne pod względem fizycznym, nigdy nie będą zachowywać się tak samo. Z całą pewnością jest to dziwna i mocno podejrzana idea. Nieustannie przecież jesteśmy świadkami czegoś innego. Uczni specjaliści nieomal każdego dnia wytwarzają różne pod względem fizycznym obiekty, które zachowują się podobnie lub wręcz identycznie. Odpowiadamy jak następuje. Różne pod względem fizycznym obiekty zachowują się podobnie lub identycznie pod pewnym względem i w pewnych tylko środowiskach. Zdarza się również, że podobieństwo jest osiągnięte kosztem odmienności środowisk, czego wcale często nie zauważamy, obserwując tylko końcowe efekty. Jeżeli raz ustalimy, co jest środowiskiem, a co obiektem, to dwa różne obiekty nigdy nie będą zachowywać się tak samo. Jeżeli dwa obiekty różnią się dla nas czymś więcej niż tylko liczbą, to zawsze znajdziemy (lub skonstruujemy) takie środowisko, które „wykryje” te różnice.

Naszym zdaniem idea całkowicie różnych pod względem fizycznym obiektów, ale zachowujących się tak samo, jest wyrazem bardzo starego przekonania, że o zachowaniu pewnych obiektów (np. ludzi lub zwierząt) decyduje nie fizyczne ciało i jego własności, ale niematerialna dusza. Dwa różne pod względem cielesnym obiekty mogą wszak posiadać taką samą duszę. Funkcjonalizm, zwłaszcza w swej komputerowej wersji, jest w naszej ocenie kolejną wersją tego poglądu. Rolę duszy pełni tym razem program. *Mózg Jana* i *SJ*, choć tak różne pod względem fizycznym, miałyby zachowywać się tak samo, gdyż posiadają takie same programy, czyli takie same dusze. Dusza pojęta jako program może, podobnie jak wiele jej szacownych antenatek, wędrować, tj. przenosić się z jednego medium do drugiego. Możliwości, które się otwierają przed wojażującym plikiem umysłu są przy tym nieporównanie większe i bardziej ekscytujące niż te, które obiecują biura podróży i tradycyjne religie Dalekiego Wschodu. Dusza (ze względu na wielość realizacji) może „zamieszkać” nieomal wszędzie — we wnętrzu komputera lub w miliardzie Chińczyków. Wędrując w ten sposób, można zaspokoić nie tylko ciekawość nowych miejsc, lecz także urzeczywistnić pragnienie nieśmiertelności.

Sądzymy, że kryzys tradycyjnych religii jest jedną z ważnych przyczyn popularności funkcjonalizmu.

Obiekcja trzecia. Cała krytyka funkcjonalizmu przedstawiona w tej pracy oparta jest na spostrzeżeniu, że różne pod względem fizycznym (a w szczególności wykonane z różnych materiałów) obiekty posiadają różne moce przyczynowe. Z tego właśnie spostrzeżenia wyciąga się następnie wniosek, że obiekty takie w pewnych środowiskach i w pewnych sytuacjach będą zachowywać się odmiennie. Autorem wspomnianego spostrzeżenia jest jednak z całą pewnością Searle. Filozof ten wielokrotnie za-

uważa (Searle 1994, 1997), że tzw. program nie posiada innych mocy przyczynowych niż medium, w którym jest realizowany. Czemu zatem, można zapytać, autor ten przywoływany jest w charakterze negatywnego bohatera? Otóż, zgadzamy się ze wszystkim, co John Searle pisze na temat mocy przyczynowych i programów. A ponadto, podzielamy jego stanowisko w kwestii świadomości, zwane *biologicznym materializmem*. Zawdzięczamy temu filozofowi bardzo dużo i chcemy to wyraźnie podkreślić. Rzecz jednak w tym, że — jak się wydaje — John Searle nie zauważa niektórych konsekwencji swych poglądów. Zachowanie — w opinii Searle'a — nie jest niczym ważnym. Różnice lub podobieństwa w zachowaniu nie wydają wiarygodnego świadectwa o rzeczywistych własnościach czy to robota, czy człowieka. Zachowanie to czynnik, który w rozważaniach o umyśle w zasadzie można pominąć. Wydaje się, że antypatia do behawioryzmu indukuje u tego autora skłonność do bagatelizowania samego fenomenu zachowania. Jeśli Searle wspomina o zachowaniu, to po to, aby zauważyć, że nie jest ono ważne. Liczą się przyczyny zachowania, samo zachowanie jest nieistotne. My staramy się dowodzić czegoś innego. Sądzimy, że jeżeli przyczyny zachowania są różne, to i zachowania muszą być różne. Różne przyczyny muszą wszak powodować w pewnych środowiskach odmienne zachowania.

Wielu filozofom umysłu argumenty Searlea wydają się mało przekonujące. Sądzimy, że jedną z przyczyn tego stanu rzeczy jest właśnie to, że autor *The Rediscovery of the Mind* tak mało pisze o zachowaniu. Behawioryzm nie jest już modny. Wyparły go różne odmiany funkcjonalizmu. Funkcjoniści odziedziczyli jednak po behawioryzmie pełen szacunek, a niekiedy wręcz religijny stosunek do zachowania. Sądzimy, że do takich filozofów znacznie łatwiej jest dotrzeć z pewnymi argumentami, odwołując się do fenomenu zachowania. Jeśli uda się ich przekonać, że różne pod względem fizycznym obiekty muszą się różnić na poziomie zachowania, to zarazem uda się ich przekonać, że w funkcjonalistycznym obrazie umysłu tkwi jakaś luka.

Na zakończenie chciałbym się zająć jeszcze jedną sprawą. W pracy tej sporo pisałem o zachowaniu podmiotowym. W kontekście działania neuronami określenie to jednak może wywoływać sprzeciw.

3.2.3. DZIAŁANIE NEURONAMI I OSOBA SPRAWCY

Mówienie o *robieniu neuronami* może się wydać podejrzane z innego jeszcze względu. Ilekroć bowiem mówimy o robieniu lub działaniu, musi — jak mniemamy — istnieć *sprawca, podmiot* działania. Poszukiwanie sprawcy lub podmiotu to zajęcie, któremu każdego dnia oddaje się wielu różnych ludzi. Prokurator chce wiedzieć *Kto podłożył bombę*, a nauczyciel w szkole *Kto wybił szybę*. Poszukiwania sprawcy ustają — taka jest codzienna praktyka — gdy uda się nam wskazać pewnego konkretnego człowieka. W codziennej praktyce nie mamy również kłopotów z wyjaśnieniem, w jaki sposób coś zostało zrobione i jakimi środkami lub narzędziami posłużył się sprawca. Wyjaśnienia typu *Jan wziął kamień do ręki i rzucił nim w szybę* całkowicie nas zadowolają. To co wystarcza nauczycielom i prokuratorom, budzi nieufność filozofów. Kamień i ręka to przecież tylko narzędzia. Jan mógł zrobić to samo w zu-

pełnie inny sposób. Mógł kopnąć szybę (założmy, że znajduje się ona na parterze, a Jan ma długie nogi) lub naprzeć na nią barkiem. Przebieg zdarzeń byłby inny, ale sprawca — jesteśmy o tym przekonani — dokładnie ten sam. Nasuwa się w związku z tym pytanie. *Kim lub czym jest ten, kto wybija szyby, posługując się wedle upodobania ręką, nogą lub barkiem? Kto, w charakterze narzędzi, wykorzystuje kije, kamienie, ręce lub nogi? Kim jest właściwy sprawca i jaka jest jego prawdziwa natura?* Filozofowie rozpoczynają swoje poszukiwania dokładnie w tym miejscu, w którym kończą je nauczyciele i prokuratorzy.

Kim jest właściwy sprawca i jaka jest jego natura? Tradycyjna odpowiedź na to pytanie jest dobrze znana. Kojarzy się ją na ogół z osobą Kartezjusza (choć można byłoby wskazać bardzo wielu prekursorów i pomysłodawców tego konceptu). Odpowiedź ta brzmi: rzeczywisty podmiot-sprawca to *duch*. Prawdziwa natura *sprawcy-ducha* jest zupełnie inna niż natura kijów, kamieni, rąk i nóg. Te ostatnie są *materialne*, a duch jest *niematerialny*. To podmiot-duch zaplanował wybicie szyby, a następnie, wykorzystując tajemną władzę, którą sprawuje nad ciałem, zrealizował swój zamiar. Odpowiedź wskazująca na ducha zadowalała zwykłych ludzi i większość filozofów przez wiele set lat. Czasy się jednak zmieniły i obecnie kartezjaniści stanowią (przynajmniej wśród filozofów) zdecydowaną mniejszość. Niektórzy filozofowie utrzymują nawet, że podmiot-sprawca to fikcja, to tylko punkt ciężkości narracji językowej (Dennett 1991), inni zaś, że podmiot-duch to relikwiny epoki magicznej (Churchland P. M. 1981, 1994) itd. Pytanie, *kim lub czym jest podmiot-sprawca*, należy z pewnością do najważniejszych i najbardziej intrygujących pytań współczesnej filozofii umysłu. Jest to ponadto jedno z najbardziej dramatycznych pytań. Od odpowiedzi nań zależy bardzo wiele, znacznie więcej niż kształt przyszłych teorii działania.

Kim lub czym jest więc podmiot-sprawca działania lub robienia różnych rzeczy? Nie wiem, nie potrafię odpowiedzieć na to pytanie. Nie jest ono zresztą ani przedmiotem tej pracy, ani tych uwag. Ich sens jest inny. Otóż, mówienie o *robieniu neuronami* nie prowadzi, co się tyczy podmiotu-sprawcy, do żadnych dodatkowych problemów. Nie jest ani trudniejsze, ani bardziej tajemnicze od takiego samego pytania zadanego na okoliczność robienia rękami lub nogami. Pytanie *Kto zrobił to neuronami?* brzmi — zgoda — dziwnie, ale wcale nie jest bardziej niepokojące niż pytanie *A kto zrobił to rękami?* Na jedno i drugie — jak sądzę — odpowiedzielibyśmy tak samo. W zależności od przeświadczeń filozoficznych i okoliczności mówilibyśmy: *Zrobił to Jan Kowalski* albo *Zrobił to duch*, albo *Podmiot to fikcja* — *oba pytania są źle postawione*, albo *Przy aktualnym stanie wiedzy nie sposób na to pytanie odpowiedzieć* itd.

Założmy, że pewien Jaś doznał całkowitego paraliżu w wieku lat pięciu. Nim Jaś zachorował umiał już robić wiele różnych rzeczy, w tym — oczywista — włączać i wyłączać telewizor. Przyjmijmy, że mózg Jasia funkcjonuje normalnie i że Jaś widzi, słyszy i rozumie co się do niego mówi. Opiekunowie Jasia domyślają się, że wielką ulgą w cierpieniach dziecka byłaby możliwość samodzielnego korzystania z telewizora. Jaś nie może posłużyć się rękami, ale może użyć mózgu. Ten funkcjonuje bez

zarzutu. Czy jest możliwe, aby Jaś włączał i wyłączał telewizor neuronami? Z pewnością. Skoro szczer potrafi uruchomić przy pomocy neuronów odpowiednie artefakty, to zapewne potrafi to i człowiek. Załóżmy, że odpowiednie urządzenia zostały już zainstalowane i że Jaś po kilkutygodniowym treningu *nauczył się* włączać i wyłączać telewizor. Pisząc, że „Jaś nauczył się”, chcę powiedzieć tyle, że nauce towarzyszyły fenomeny charakterystyczne dla nabywania nowej umiejętności. Z początku Jaś się mylił, telewizor włączał się i wyłączał przez przypadek itd. Wreszcie jednak Jaś uzyskał kontrolę nad tym urządzeniem i włącza oraz wyłącza telewizor kiedy chce. Jaś jest np. w stanie wyłączyć i włączyć telewizor ilekroć jego opiekunowie formułują takie polecenie (Jaś rozumie co się do niego mówi i jest posłusznym dzieckiem). W każdym razie rozróżnienie pomiędzy *robieniem* i *wydarzaniem się* zachowuje swój sens zarówno z punktu widzenia Jasia, jak i jego opiekunów. Czy prowadziłyby to do jakichś nowych trudności w kwestii podmiotu-sprawcy? Nie wydaje się.

Wyobraźmy sobie, że w pokoju, gdzie przebywa Jasio, znajduje się jeszcze dwoje innych dzieci — Kasia i Małgosia. Załóżmy, że dzieci mają zakaz oglądania telewizji. Wchodzimy do pokoju i zauważamy, że telewizor jest pomimo to włączony. Przeprowadzamy „śledztwo”: *Kto włączył telewizor?* Po ustaleniu sprawcy, powiedzieliśmy, że zrobiła to Kasia lub że Małgosia, lub że... Jasio.

Załóżmy, Jasio odzyskał władzę w rękach i nogach. Teraz, tak jak inne dzieci, może zapalać telewizor, używając rąk. Załóżmy ponadto, że urządzenia, z których korzystał Jasio, pozostały w domu, nadają się do użytku i że Jasio lubi niekiedy włączać telewizor neuronami. Wyobraźmy sobie, że Jasio zaczyna coraz częściej łamać zakaz oglądania telewizji w określonych godzinach. Przy czym, niekiedy włącza telewizor, posługując się rękami, innym razem — neuronami. Czy stanowiłoby to dla nas jakiś problem? Czy na pytanie *Kto włączył telewizor?* mielibyśmy dwie różne odpowiedzi, uzależnione od tego, czy Jaś użył neuronów czy rąk? W obu przypadkach powiedzielibyśmy po prostu, że *Jaś włączył telewizor*. Na dalsze pytania, pytania typu *Kto to jednak jest ów Jaś i jaka jest jego prawdziwa natura?* nie ma prostej i ogólnie przyjętej odpowiedzi. Ani wtedy, gdy Jaś robi coś neuronami, ani wtedy, gdy robi coś rękami.

BIBLIOGRAFIA

- Block, N. 1990. *The Computer Model of the Mind*, [w:] D. Osherson and E.E. Smith ed., *An Invitation to Cognitive Science* 3. Cambridge, MA: MIT Press.
- Byrne, A. 1994. *Behaviourism*, [w:] *A Companion to the Philosophy of Mind*, ed. Samuel Guttenplan, Basil Blackwell.
- Chalmers, D. J. 1996. *The Conscious Mind*, Oxford University Press.
- Churchland, P. M. 1981. *Eliminative materialism and the propositional attitudes*, „Journal of Philosophy”, 78, 67-90.
- Churchland, P. M. 1989. *A Neurocomputational perspective*, Cambridge, MA: MIT Press.
- Churchland, P. M. 1994. *Folk psychology*, [w:] *A Companion to the Philosophy of Mind*. ed. Samuel Guttenplan, Basil Blackwell.

- Davidson, D. 1970. *Mental events*, [w:] *Expirience and Theory*, ed. L. Foster and J. W. Swanson: Amherst, MA: University of Massachusetts Press.
- Dennett, D. C. 1991. *Conscioussness Explained*. London.
- Johnson-Laird, P. N. 1988 *The Computer and the Mind*. Cambridge, MA: Harvard University Press.
- Kurzweil, R. 1999. *Zespolenie ducha z maszyną*, „Świat Nauki” Nr 12(100) s. 72-76.
- Putnam, H. 1965. *Brains and behavior*, [w:] *Analytical Philosophy*, Vol 2, ed. R. J. Butler.
- Putnam, H. 1994. *Putnam, Hilary*, [w:] *A Companion to the Philosophy of Mind*, ed. Samuel Guttenplan, Basil Blackwell.
- Pylyshyn, Z.W. 1984. *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, MA: MIT Press.
- Searle, J. 1994. *The Rediscovery of Mind*, The MIT Press.
- Searle, J. 1995a. *Umysty, mózgi i programy*, [w:] *Filozofia umystu*, Warszawa.
- Searle, J. 1995b. *Umyst, mózg i nauka*, Warszawa.
- Searle, J. 1997. *The Mystery of Consciousness*, The New York Review of Books.
- Smart, J. J. C. 1962. *Senstations and brain processes*, [w:] *The Philosophy of Mind*, ed. V.C. Chappell. Englewood Cliffs, N. J.: Prentice-Hall, Inc.
- Wittgenstain, L. 1972. *Dociekania Filozoficzne*. Warszawa.
- Zucker, M. 2000. *Umyst pokonuje opór materii*. „Świat Nauki” Nr 1(101) s. 27-28.