Filozofia Nauki
Rok III, 1995, Nr 4(12)

Jan Żytkow

Redukcjonizm naukowy i komputerowy a treść świadomości

Redukcja naukowa i funkcjonalizm komputerowy

Nauka przynosi poznanie rzeczywistości, rozkładając obiekty i zjawiska na składniki. Bada empirycznie te składniki i oddziaływania między nimi, opisując je za pomocą symboli, liczb, struktur i formuł matematycznych. Poznane w ten sposób składniki i ich oddziaływania używane są następnie do budowy modeli złożonych obiektów i zjawisk. W trakcie budowy modelu jego formalizm składany jest z elementów symboliczno-matematycznych odpowiadających poszczególnym składnikom, ich strukturze i ich oddziaływaniom. Adekwatność modelu można wykazać empirycznie — pokazując, że obiekt modelowany rzeczywiście składa się z postulowanych składników i struktur, i że jego stwierdzone empirycznie zachowanie odpowiada zachowaniu przewidywanemu przez model. Gdy model dobrze przewiduje obserwowane zachowania, nabieramy przekonania, że zjawisko dokonuje się tak, jak opisuje to model, a więc że elementy modelu mają swe odpowiedniki w rzeczywistości fizycznej, i że wszystkie obiekty fizyczne, które mają wpływ na dane zjawisko, zostały uchwycone. Na początku XVIII wieku, na przykład, na podstawie praw Newtona Halley przewidział trajektorię komety nazwanej później jego imieniem, a w szczególności jej pojawienie się blisko Ziemi w grudniu 1758. Gdy przewidywanie to się spełniło, teoria Newtona uzyskała efektowne potwierdzenie. Prawa Newtona zastosowane do kilku punktów materialnych okazały się wystarczające dla wyjaśnienia ciekawego zjawiska astronomicznego.

Te same podstawowe składniki, na przykład prawa Newtona i obiekty reprezentowane jako punkty materialne, mogą być komponowane w ogromną liczbę modeli, wyjaśniając wiele różnorodnych zjawisk. Bogactwo istniejących zastosowań — i postęp w tworzeniu nowych — rozwija w nas przekonanie o nieograniczonych zastoso-

waniach nauki i rodzi wiarę, że każde zjawisko zostanie wcześniej czy później wyjaśnione.

W ten sposób nauka przekształca się w ideologię. Od konkretnych modeli — coraz bardziej różnorodnych i precyzyjnych — przechodzi do uniwersalnej tezy, że wszystko w świecie materialnym da się wyjaśnić za pomocą modeli naukowych, a więc, że wszystko składa się ze struktur materialnych o określonych składnikach i oddziaływaniach. W historii nauki przekonanie o jej uniwersalności było dodatkowo uzupełniane przekonaniem, że podstawowe elementy są już znane. Kiedyś podstawą taką był mechanicyzm; dziś jest to kombinacja teorii czterech oddziaływań. Podstawowe teorie w dziedzinie chemii czy biologii nie muszą sięgać do cząstek elementarnych. Współczesna biologia odwołuje się do molekuł organicznych i ich oddziaływań jako wyjściowych składników, z których buduje się modele komórek i zjawisk w organizmie.

Nauka nie widzi zasadniczych granic dla swej metody. Zawsze można uważać, że zjawisko, które dziś nie posiada wyjaśnienia, znajdzie je wcześniej czy później. Być może istniejące elementy wiedzy okażą się wystarczające, a być może odkryte zostaną nowe składniki, nowe struktury bądź oddziaływania. Nowe odkrycia nie zmieniają zasadniczo metody naukowej; wzbogacają ją bowiem o elementy podobne do już posiadanych. Ciężar dowodu spychany jest na przeciwnika tezy o uniwersalizmie nauki. Od oponenta żąda się: pokaż mi sytuacje, których nie może wyjaśnić model naukowy; pokaż mi obiekty, które posiadają własności nie mierzalne przez naukę; przedstaw zasadnicze powody, dla których nie da się naukowo wyjaśnić tych obiektów i sytuacji.

Jeżeli się bierze na siebie ciężar dowodu, to nie wystarczy wskazać, że jakieś zjawisko nie ma modelu. Trzeba pokazać, że go mieć nie może. Na poziomie zjawisk świadomości — od dowodu takiego wymaga się wskazania składników, struktur bądź oddziaływań, które wykraczają poza zakres nauki. Tymczasem metody naukowe nie widzą w mózgu niczego, co by nie składało się z atomów kilkunastu pierwiastków i ich kombinacji. Brak konkretnych wyjaśnień zjawisk świadomości racjonalizowany jest przez brak konkretnej wiedzy szczegółowej o materialnej strukturze mózgu bądź przez trudności z wnioskowaniem na temat struktur złożonych.

Modele komputerowe funkcjonują podobnie do modeli w naukach przyrodniczych, reprezentując obiekty, struktury i oddziaływania. Reprezentacja ta, równie formalna i schematyczna jak w modelach matematycznych, dokonuje się za pomocą struktur budowanych w pamięci komputera. Ale pojawia się tu nowy ważny element w postaci kodyfikacji wnioskowań i obliczeń przez algorytmy. Algorytm, nazywany też programem, jest strukturą składającą się z dobrze określonych instrukcji operujących na danych w pamięci komputera. Chociaż można śledzić funkcjonowanie programu bez użycia komputera, kolejną jego zaletą jest automatyczne wykonywanie ogromnej liczby instrukcji i operowanie wielką liczbą różnorodnych danych. Prowadzi to często do wyników, których uzyskanie bez udziału komputera byłoby niemożliwe.

Specyficzny dla konkretnego modelu jest program komputerowy, nie zaś procesor i pamięć, reprezentujące uniwersalny komputer, na którym można wykonywać wszelkie algorytmy. Materialna budowa komputera jest oczywiście ważna, gdyż niewielkie choćby odstępstwa od zaplanowanej konstrukcji powodują, że komputer przestaje funkcjonować jako komputer. Lecz chociaż elementy konstrukcyjne komputera składają się oczywiście ze składników atomowych, używając komputerów we właściwych warunkach nie musimy się liczyć z wpływem zakłóceń na niższym poziomie na wynik obliczenia.

Ostatnio pojawiają się możliwości, dzięki którym modele komputerowe stosowane są do modelowania zjawisk zachodzących w naszych umysłach. Program komputerowy i dane, którymi program ten operuje modelują funkcjonowanie myślącego podmiotu działającego. Reprezentują cele i schematy wnioskowania, wyjaśniają procesy myślowe za pomocą struktur algorytmicznych, składających się z prostych, dobrze określonych instrukcji, operujących na danych w pamięci komputera. Określony program i jego dane mogą wykazywać podobieństwo do zjawisk zachodzących w umysłach ludzi, np. do procesu zapamiętywania dużej liczby cyfr i odtwarzania ich z pamięci bądź też rozwiązywania określonego problemu. Wiele modeli wykazuje dużą zgodność z empirycznymi rezultatami uzyskiwanymi w badaniach z ludźmi.

W obecnym stanie rzeczy modelowanie w naukach przyrodniczych odwołuje się do reguł ściślejszych niż komputerowe modelowanie umysłu. Modelowanie komputerowe znajduje się w fazie przyjmowania istnienia ukrytych składników i przypisywania im określonych struktur: algorytmów i struktur danych. Nie jest jasne, czym są ich fizyczne odpowiedniki w umysłach ludzi. Jest to zapewne związane z wczesnym stadium rozwoju kognitywistyki. Na razie sukcesem są jakieś w miarę adekwatne modele dla poszczególnych zjawisk. W miarę tworzenia takich modeli można się spodziewać, że kanon redukcji będzie uściślany, co doprowadzi w końcu do niewielkiego zestawu elementarnych składników i funkcji, których fizyczne odpowiedniki będą coraz lepiej zlokalizowane. Sieci neuronowe — to modele komputerowe, które upraszczają problem fizycznych odpowiedników, wychodząc od znanych składników mózgu: neuronów i ich połączen synaptycznych. Ale problemem staje się wtedy reprezentacja celów myślącego podmiotu działającego, wiedzy wyższego rzędu i świadomych rozumowań.

Adekwatny model naukowy zakłada, że odpowiedniki elementów modelu istnieją w rzeczywistości fizycznej, i że relacje między nimi są wiernie opisane przez model. Ponieważ model komputerowy jest funkcjonującym układem fizycznym, więc w porównaniu z tradycyjnymi modelami naukowymi ma tę przewagę, że może także funkcjonować podobnie do modelowanego zjawiska. Komputerowe próby modelowania umysłu i świadomych zachowań, choć jeszcze bardzo ograniczone, każą zastanowić się nad tym, czy funkcjonalnie adekwatny model świadomości nie byłby sam systemem fizycznym obdarzonym świadomością. Materialne podłoże komputera jest inne niż materialne podłoże naszych umysłów, ale jeśli funkcjonowanie komputerowego podmiotu działającego odpowiada na poziomie zewnętrznej obserwacji funkcjonowaniu

naszych umysłów, to może wystarcza to by uznać, że system komputerowy jest umysłem obdarzonym świadomością.

Modele a świadomość

Czy nasza świadomość jest w konflikcie z materialistycznym redukcjonizmem, czy też da się wyjaśnić naukowo? Czy świadomość wykracza poza modele komputerowe, czy też jesteśmy komputerami o nieznanej jeszcze dokładnie strukturze i sposobie działania? Mimo że można bezpiecznie przewidywać dalsze sukcesy nauki i programów komputerowych w modelowaniu zjawisk związanych ze świadomością, trudno pozbyć się wrażenia, że nasza świadomość jest czymś szczególnym, nie dającym się sprowadzić do funkcjonowania procesora lub układu składników fizycznych i chemicznych. Mamy też wrażenie, że redukcja gubi przyczynowy wpływ na nasze zachowanie świadomego obrazu świata i celów, które stawia sobie nasza świadomość, jak też świadomie podejmowanych działań.

Ci, którzy odrzucają nieodparte dane o przyczynowym wpływie naszych świadomych postrzeżeń, myśli i działań na świat materialny — robią to, by uniknąć konfliktu z naukowym redukcjonizmem. Postaram się wykazać, że konfliktu tego uniknąć się nie da, że trzeba go zaakceptować i wyciągnąć z niego wnioski. Podstawowy składnik wszystkiego, co jest w naszej świadomości, nie da się wyjaśnić za pomocą modeli naukowych ani komputerowych. Właściwy wywód będę musiał poprzedzić ważnym rozróżnieniem pojęciowym.

Treści bezpośrednio uświadamiane

To wszystko, co postrzegamy, czujemy i myślimy, nazwiemy ogólnie treściami naszej świadomości. Zasadniczą rolę w moim wywodzie odgrywa rozróżnienie w treści naszej świadomości kilku warstw. Zacznijmy od najprostszych postrzeżeń, powiedzmy wrażenia czerwieni. Różne czerwone obiekty postrzegamy podobnie. Postrzegamy je inaczej niż obiekty zielone. Wrażenia czerwieni są nam dane bezpośrednio, choć wiemy też, że kolor czerwony da się określić fizycznie przez stwierdzenie odpowiednich długości fal w promieniowaniu emitowanym bądź odbijanym przez dane ciało.

Aby wyrazić te fakty, odróżnimy:

(1) podział obiektów na grupy;

(2) fizyczną treść podziału i opartą na niej fizyczną zasadę przynależności do każdej grupy i

(3) treść wewnętrzną narzucającą się bezpośrednio naszej świadomosci i opartą na niej zasadę podziału.

Do rozpoznawania długości fal radiowych używamy treści fizycznej. Rozpoznając kolory posługujemy się treścią wewnętrzną, ale możemy też używać treści fizycznej, badając długości i intensywności linii spektralnych w widmie danego obiektu.

Treści wewnętrzne naszej świadomości możemy odróżnić wyraźnie przez porównanie z sytuacjami, w których treść wewnętrzna nie występuje lub musi różnić się od tej,

którą mamy. Rozważmy daltonistę. Jak postrzega on czerwień i zieleń? Na pewno inaczej niż ja, bo mnie różnica między przedmiotem czerwonym i zielonym rzuca się w oczy, a on tych kolorów nie odróżnia. Treściami wewnętrznymi są moje postrzeżenia czerwieni i zieleni, i postrzeżenia jakiejś jednej jakości przez daltonistę. Rozważmy inne przykłady. Jakie kolory postrzegają zwierzęta, tam gdzie widzenie ich wykracza poza zakres widzenia ludzi? Jakie są wrażenia nietoperzy, gdy widzą za pomocą ultradźwięków? Analizując te przykłady możemy dotrzeć do dwu pierwszych składników, ale brakuje nam treści wewnętrznej. Za pomocą doświadczeń w rozpoznawaniu obiektów fizycznych możemy stwierdzić, że dwa poznające podmioty dokonują takiego samego podziału na grupy, lub też, że ich podziały się różnią. Ale w ten sposób nie docieramy do treści wewnętrznych, określających ich podziały. Podobnie możemy pokazać, że dwie fizyczne zasady podziału są równoważne, bez odwoływania się do treści wewnętrznych.

W naukowej metodzie klasyfikowania używa się pomiarów. Możemy zorientować się za pomocą instrumentów naukowych nie tylko co do koloru powierzchni, ale też co do jej zdolności do odbijania fal radarowych. Tylko pierwsza z tych jakości jest nam dana we wrażeniach zmysłowych. Treść wewnętrzna jest tym, co wiemy o czerwieni, a czego nie mamy w odniesieniu do fal radarowych. Trudno jednak wykluczyć, że są istoty, które postrzegają fale radarowe tak, jak my postrzegamy promieniowanie widzialne. Pytanie, jakie są ich wrażenia, ma sens, który wykracza poza sens naukowy określony przez pomiary.

Moje treści wewnętrzne są dane tylko mnie i nie jestem ich w stanie nikomu bezpośrednio przekazać. Ale jestem w stanie tworzyć sytuacje, które mogą wywołać w innych tę samą treść wewnętrzną. Metodę tę stosujemy w definicji ostensywnej, kiedy to kojarzymy słowo z odpowiednio skonstruowanymi sytuacjami. Sytuacje te mają w uczącym się wywołać takie samo wrażenie, jak we mnie, i skojarzyć to wrażenie ze słowem, które służy do jego identyfikacji. Rezultaty uczenia są oceniane przez podobieństwo w ocenie przynależności obiektów do grupy między mną i osobą, którą uczę.

Wewnętrzne treści postrzeżeń nie ograniczają się do prostych barw czy kształtów, a wiele cech fizycznie trudnych do zdefiniowania jest łatwo uchwytnych za pomocą treści wewnętrznej. Przyjemny wygląd czy stanowczy głos narzucają nam się bezpośrednio. Składnik treści wewnętrznej, obecny we wszystkich naszych postrzeżeniach świata zewnętrznego, rozciąga się też na myśli i odczucia wewnętrzne. Można nawet twierdzić, że bezpośrednio nie obcujemy z niczym więcej niż z treściami wewnętrznymi. Podziały na klasy są wtórne i nie towarzyszą wielu naszym postrzeżeniom, a treść fizyczna pojawia się pod postacią metod pomiarowych w rezultacie badań naukowych i jest często nieistotna, jak w wypadku pojęć moralnych. Oceniamy moralnie za pomocą obserwacji, ale tylko w stopniu, w którym potrafimy sytuacje fizyczne powiązać z wartościami moralnymi w nas samych i w innych podmiotach moralnych.

Przeżycia moralne czy estetyczne odnoszą się do sytuacji w świecie zewnętrznym, ale dominuje w nich treść wewnętrzna.

Czy możemy zdefiniować pojęcie treści wewnętrznej? Słowna definicja jest niemożliwa ani dla konkretnych treści, takich jak 'czerwony', ani dla ogólnego pojęcia treści wewnętrznej. Ale możliwy jest system definicji ostensywnych. „Treść wewnętrzna" — to nazwa ogólna dla konkretnych elementów naszej świadomości. Jej znaczenie budowane jest z wielu przykładów definiowanych ostensywnie, a następnie używanych w definicjach ostensywnych wyższego rzędu. Proces definiowania i odróżniania treści wewnętrznej od treści fizycznej zaczyna się na niskim poziomie. Przykład z kolorem i radarem odróżnia sytuacje bezpośredniego przeżycia i jego braku, a przykład daltonisty wskazuje na sytuacje, w których wewnętrzne przeżycia muszą się różnić.

Każdy z nas uczy się początkowo poszczególnych treści wewnętrznych przez definicje ostensywne. Treści te pozwalają nam nazywać i rozpoznawać określone sytuacje oraz komunikować o nich innym, ale nie zdajemy sobie sprawy z tego, że są to treści wewnętrzne. Znacznie później uczymy się pojęć zakresu i treści fizycznej. Rozszerzają one nasze możliwości rozumienia i wyrażania, ale jednocześnie, nakładając się na treść wewnętrzną — komplikują jej rozumienie. W rezultacie musimy wyodrębnić treści wewnętrzne, kontrastując je z innymi danymi naszej świadomości.

Proste sytuacje pozwalają nauczyć dziecko terminów takich, jak „czerwony" lub „trójkąt". Znajomość wielu takich terminów pozwala przejść na wyższy poziom i uczyć terminów takich, jak „kolor", czy „kształt". Tu uczymy również za pomocą przykładów, ale odwołują się one nie do poszczególnych barwnych plam czy przedmiotów o określonym kształcie, ale do terminów określających kolory bądź kształty, a dopiero pośrednio, przez te terminy, do ich treści wewnętrznych. Definiując ostensywnie termin „kolor", używamy terminów — nazywających poszczególne znane kolory — jako przykładów, a w miarę potrzeby innych znanych własności, takich jak kształty czy desenie, jako kontrprzykładów. Gdy zabieg ten jest skuteczny, nowy kolor nazwiemy kolorem a nie kształtem, gdy zobaczymy go po raz pierwszy. Na jeszcze wyższym poziomie używamy terminów „kolor" i „kształt" by zdefiniować termin „własność". W analogiczny sposób możemy stopniowo rozróżnić treści wewnętrzne i treści fizyczne poszczególnych własności, a następnie uchwycić pojęcie samej treści wewnętrznej własności. Podobny zabieg można zastosować do złożonych struktur i sytuacji, odróżniając ich treść wewnętrzną od treści fizycznej.

Naukowe wyjaśnienia świadomości

Wykażę teraz, że bezpośrednio uświadamiana treść wewnętrzna nie da się wyjaśnić przez żaden model naukowy. Załóżmy, że nauka bardzo dokładnie wyjaśniła funkcjonowanie mózgu. Załóżmy dalej, że świadomość ze wszystkimi jej detalami, takimi jak spostrzeżenia, myśli i działania, została utożsamiona z określonym zbiorem struktur materialnych i procesów fizycznych, a poszczególne stany świadomości — z konkret-

nymi strukturami i procesami, zapewne o wielkiej złożoności. Zgodnie z tym założeniem, w modelu tym zostało uchwycone w języku fizyki i chemii to wszystko, co da się naukowo powiedzieć o stanach naszej świadomości, w szczególności przyczynowe mechanizmy ich pojawiania się i mechanizmy przechodzenia z jednych stanów w inne. Ale z samego modelu nie wynika, jakie treści wewnętrzne odpowiadają określonym stanom modelu. Ich identyfikacja musi być indywidualnie dodawana do tego modelu dla każdej poszczególnej treści wewnętrznej. Nawet gdy da się rozłożyć poszczególny stan opisany naukowo na składniki i nawet gdy dla każdego składnika znamy odpowiadającą mu treść wewnętrzną, nie posiadamy jeszcze treści wewnętrznej całego stanu. Co innego bowiem jest dysponować dokładnym opisem obrazu, a co innego widzieć ten obraz.

Można korelować określone stany fizyko-chemiczne modelu z konkretnymi treściami wewnętrznymi, zauważając na przykład, że określony stan fizyko-chemiczny pojawia się w mózgu każdego człowieka razem z wewnętrznie postrzeganym wrażeniem czerwieni. Identyfikacji tej możemy dokonać z naszym własnym wrażeniem czerwieni. Następnie, obserwując w mózgu różnych ludzi analogiczny stan raz już zidentyfikowany, możemy przewidywać, że odpowiadająca mu treść wewnętrzna to wrażenie czerwieni. Tożsamość stanów obserwowalnych u różnych ludzi przy oglądaniu plam barwnych tego samego koloru utwierdziłaby nas w przekonaniu, że stany wewnętrzne są u różnych ludzi identyczne. Każdej wewnętrznej treści naszej świadomości odpowiadałyby od strony naukowej określone struktury i procesy. Być może, dałoby się wykazać np., że daltonista widzi kolory zielony i czerwony tak, jak my czerwony, ale nadal nie można by mu wskazać, jaka jest treść wewnętrzna stanów, które się nigdy w jego świadomości nie pojawiają. W dalszym ciągu nie wiedzielibyśmy, co nietoperze postrzegają za pomocą ultradźwięków, bądź jak owady widzą nadfiolet. Model mózgu nietoperza może świadczyć o podobieństwie stanów postrzeżeniowych nietoperza i człowieka. Ale w ten sposób nie dotrzemy do stanów świadomości nietoperza.

Dane nam w bezpośrednich przeżyciach treści — mimo że korelowalne ze stanami modelu — są naukowo zbędne. W wyjaśnieniach wystarczy używać stanów modelu, nie wspominając treści wewnętrznych. Jedne stany modelu są powiązane z innymi stanami za pomocą zawartych w modelu mechanizmów przyczynowo-skutkowych, podczas gdy treści wewnętrzne pozostają poza zasięgiem tych mechanizmów, gdyż są indywidualnie łączone ze stanami modelu. Jeśli usunąć te połączenia i powiązane z nimi treści wewnętrzne, to model przedstawiałby istotę fizycznie identyczną z nami, która zachowywałaby się dla zewnętrznego obserwatora identycznie jak my, jeśli brać pod uwagę to wszystko, co ów obserwator mógłby mierzyć. W przeciwieństwie do nas jednak, ta bez-świadoma istota nie miałaby żadnych wewnętrznych treści świadomości.

Mamy tu do czynienia z paradoksem. Istniejemy jako istoty świadome, a każda z treści naszej świadomości jest tego dowodem. Tymczasem z punktu widzenia nauki wszystkie te treści są nieistotne. Są one doczepione do mechanizmu fizyko-chemicznego, który się nimi nie posługuje i może się bez nich obejść. Gdyby te treści nie istniały,

nie trzeba by niczego zmieniać w modelu. Paradoks ten nie pojawia się natomiast z punktu widzenia owej bez-świadomej istoty. Nie ma sprzeczności między założeniem o modelu, który wyjaśnia wszystko, co dostępne obserwacji, a obserwacjami tej istoty. Bez-świadoma istota nie mogłaby zrozumieć, co znaczą treści wewnętrzne i nie widziałaby niczego, czego by brakowało modelowi.

Rozumowanie to można zastosować indywidualnie do każdej treści T naszej świadomości. Może to być np. wrażenie czerwieni w naszym polu widzenia. Może to być ból głowy. Załóżmy, że treść ta została wyjaśniona naukowo przez model M, przedstawiający pewną strukturę i proces P, które pojawiają się równocześnie z treścią T. Model M, jak każdy inny model fizyczny, opisuje obiekty, procesy i ich wzajemne relacje. Aby zweryfikować wyjaśnienie P treści T powinniśmy wykazać, że sytuacja P współwystępuje z T. Jeśli T wpływa na nasze zachowanie, jak to się dzieje w przypadku bólu głowy, model M może to zachowanie przewidzieć za pomocą efektów stanu P. To co da się przewidzieć w języku fizycznym zostaje wyjaśnione przez następstwa stanu P, natomiast treść wewnętrzna jest w tym wyjaśnieniu zbędnym dodatkiem, stowarzyszonym z P, ale niepotrzebnym dla mechanizmu opisywanego przez model.

Komputerowe modele świadomości

Wszelka redukcja umysłu do systemu komputerowego napotyka na ten sam problem, co modele naukowe. Możemy starać się, by model komputerowy odzwierciedlał świadome treści posiadane przez umysł, ale funkcjonowanie modelu ogranicza się do elementów pamięci komputera i programu, który nimi operuje. Modele komputerowe można rozpatrywać na wielu poziomach, od poziomu «maszynowego» do poziomu języków dowolnie wysokiego rzędu. Na poziomie maszynowym program komputerowy składa się z elementów pamięci w postaci ciągów zer i jedynek, i działań procesora, który przetwarza w swych układach elektrycznych stare ciągi zer i jedynek w nowe ciągi, i zapisuje nowe ciągi w pamięci. Żadna treść, która byłaby dodatkiem do elementów pamięci i kroków obliczeniowych procesora nie jest potrzebna. Program w języku wyższego rzędu używa bardziej złożonych struktur danych i instrukcji, ktore nimi operują. I tu nie jest potrzebna żadna dodatkowa treść, by śledzić funkcjonowanie programu. Teoretycznie nie jest tu nawet niezbędny komputer, choć praktycznie jest on potrzebny, gdy liczba wykonywanych instrukcji jest duża.

Roboty — to komputery wyposażone w sensory i manipulatory, oraz w programy, które interpretują dane pochodzące z sensorów, podejmują decyzje i dokonują manipulacji w świecie fizycznym. Komputer wysyła do sensorów i manipulatorów polecenia w postaci ciągów liter w prostych językach, zrozumiałych dla procesorów w tych urządzeniach. Otrzymuje od sensorów dane w postaci ciągów symboli, w zawczasu ustalonym formacie. Żadne treści wewnętrzne nie są potrzebne by wyjaśnić funkcjonowanie robota. Trudno wykazać, że roboty nie posiadają treści wewnętrznych świadomości, ale z całą pewnoscią mogą ich nie mieć. Nie widać ani jak, ani po co takie stany wewnętrzne mogłyby być do systemu komputerowego dołączone.

Prześledźmy typowe wątpliwości na przykładzie. Przypisuje się świadomości rolę centralnego systemu decyzyjnego, uzasadniając to tym, że system ten jest niezbędny dla integracji działań człowieka bądź zwierzęcia. Zauważa się słusznie, że rolą świadomości jest wybrać cele, rozpoznać bieżącą sytuację i biorąc je pod uwagę zadecydować o podjęciu stosownych działań. Programy komputerowe w dziedzinie sztucznej inteligencji odtwarzają ten schemat w różnych wersjach. Programy takie sterują zachowaniem inteligentnych autonomicznych robotów. Ale nasze rozróżnienie na funkcjonowanie programu komputerowego i wewnętrzne stany świadomości stosuje się oczywiście do komputerowych modeli świadomości tak samo, jak i do wszelkich innych modeli. Algorytm, który podejmuje decyzje, nie odtwarza tej warstwy świadomości, która zawiera stany wewnętrzne. Stany te są w modelu zbędnym dodatkiem, bez którego można opisać zachowanie się robota w każdym szczególe.

Wydaje się, że każda dobrze określona behawioralnie rola emocji, celów i innych stanów świadomości może być odtworzona przez system komputerowy. Ale system ten, przejawiając owe zachowania, nie musi i najprawdopodobniej nie posiada treści wewnętrznych, związanych z celami, emocjami, i wszelkimi innymi stanami.

Można mówić, że robot coś zauważył, bądź że się zastanawia, ale terminologia ta jest potrzebna nam — dla lepszego zrozumienia systemu komputerowego, a nie robotom, dla których zauważyć, to wpisać w określony obszar pamięci symbole odpowiednie do zewnętrznej sytuacji, bądź wywołać odpowiednią procedurę. Przez kontrast, komputer jest dobrym narzędziem dla zrozumienia przez nas naszych treści wewnętrznych. Wszystko to, o czym wiemy, że jest zauważane przez komputer, różni się od treści wewnętrznych.

Interakcjonizm

Przeżywamy niezwykłe bogactwo treści wewnętrznych. Nasze przeżycia są dla nas bardziej niewątpliwe niż cokolwiek innego. „Myślę, więc jestem" jest jednym z takich nieodpartych, niepodważalnych przeżyć. Wydaje się niemożliwe, aby całe to bogactwo doświadczeń nie odgrywało przyczynowej roli i było zbędnym dodatkiem do mechanizmu fizyko-chemicznego. Epifenomenalizm, stanowisko filozoficzne, które głosi, że tak jest, nie był nigdy popularny. Jesteśmy przekonani, że świadomość odgrywa zasadniczą rolę w naszym działaniu. Zastanowimy się teraz nad stanowiskiem interakcjonizmu, by wykazać, że jest ono zgodne z całokształtem naszych obserwacji na temat nas samych.

Czy świadomość jest przejawem funkcjonowania jakiejś substancji, jakiegoś trwałego «nośnika» wszystkich wewnętrznych treści? By znaleźć klucz do odpowiedzi na to pytanie, zastanówmy się, w jaki sposób nabieramy przekonania o istnieniu jakiegoś obiektu materialnego. Widzimy kształt. Widzimy kolor bądź deseń. Doświadczamy chłodu, gładkości i innych własności powierzchni. Nie doświadczamy żadnym zmysłem tego, co nazywamy substancją materialną. Docieramy do własności a nie do substancji, którą własności te reprezentują. Ale widzimy te własności we wzajemnym

związku. Widzimy kolor wypełniajacy kształt, nie zaś kolor i kształt oderwane od siebie. Twardość, chłód i szorstkość dotyczą tego samego fragmentu powierzchni. Są odczuwane jednocześnie i są trwałe, oczywiście w pewnych granicach. Tam gdzie doświadczamy takiego zestawienia własności, mówimy o obiektach i o ich substancji materialnej.

Gdy zastanawiamy się bliżej nad treściami, które pojawiają się w naszej świadomości, zauważamy duże podobieństwo do własności obiektów. Wewnętrzne treści mogą być zapamiętane przez długi czas. Potrafimy rozpoznać po latach smak dawno nie próbowanej potrawy. Potrafimy niemal natychmiast wrócić pamięcią do dawno minionych czasów. Czasem lepiej pamiętamy miejsca, w których byliśmy, sytuacje przeżyte czy przedmioty otrzymane przed laty. Wszyscy pamiętamy nasze obecne mieszkania i potrafimy w wyobraźni «iść przez nie» z pokoju do pokoju. Doświadczenia z naszą pamięcią są powtarzalne i każdy może je przeprowadzić. Choć każdy ma dostęp tylko do swych własnych wspomnień, możemy za ich pomocą weryfikować wspólną wiedzę o naszych treściach wewnętrznych. Mamy więc do czynienia z sytuacją podobną do nauk empirycznych, gdzie wymagana jest powtarzalność doświadczeń.

W naszej świadomości może się pojawić kilka treści na raz. Widokowi przed naszymi oczami towarzyszą myśli. W świadomości naszej wiążą się nasze wrażenia, myśli i wspomnienia. Świadomość stanu teraźniejszego i wrażenia z przeszłości pojawiają się jednocześnie, dając nam poczucie tożsamości nas teraz z nami z przeszłości. Doświadczenia te są również powtarzalne.

To współwystępowanie elementów świadomości, ich trwałość i powtarzalność, przemawiają za istnieniem ich nośników, nazywanych substancjami duchowymi. Uzasadnienie to jest podobne do uzasadnienia istnienia substancji materialnych. Substancja duchowa jest substancją w sensie analogicznym do materialnej, a treści w naszej świadomosci są analogiczne do własności obiektów materialnych: występują wspólnie i są trwałe. Interakcjonizm jest stanowiskiem, które głosi istnienie substancji obu rodzajów. Ponieważ dla naukowego redukcjonisty istnienie substancji materialnych nie budzi wątpliwości, można zaproponować mu nastepującą grę: przedstaw mi argument na rzecz istnienia substancji materialnej, a ja przerobię go na argument na rzecz istnienia substancji duchowej. Podobną grę można zaproponować w odniesieniu do argumentów za nieistnieniem: każdy argument za nieistnieniem substancji duchowej można zmodyfikować tak, by dowodził nieistnienia substancji materialnej. Rezultat taki powinien skłonić redukcjonistę do odrzucenia danego argumentu za nieistnieniem substancji duchowej.

Trwałość substancji uzasadniana bywa przez odwołanie się do zasad filozoficznych. Na przykład: zasada filozoficzna wypowiedziana przez Lukrecjusza „Nic nie powstaje z niczego, bo gdyby coś powstawało z niczego, to wszystko powstawałoby ze wszystkiego bez żadnych zarodków" dotyczy dowolnej substancji, nie tylko materialnej. Da się ją powtórzyć na temat ginięcia: nic, co istnieje, nie może zamienić się w nicość.

Skoro więc wierzymy w trwałość materii, powinniśmy również wierzyć w trwałość substancji duchowej.

Pora na kilka truizmów, niezbędnych do tego, aby zamknąć naszą argumentację na rzecz interakcjonizmu. Jest oczywiste, że obiekty materialne wpływają na stan naszej świadomości — przez poczucie zimna czy gorąca, wrażenia kolorów, kształtów czy smaków. Nasza świadomość wpływa też na świat materialny: świadome decyzje prowadzą do czynów, które zmieniają sytuację materialną wokół nas. Biorąc pod uwagę wszystkie te fakty, nie sposób nie być interakcjonistą.

„Nie przekonam się, dopóki nie poznam mechanizmu"

Stawiano zarzuty pod adresem Kartezjusza, że nie pokazał w szczegółach, jak oddziałują ciało i dusza. Ten sam zarzut można postawić każdej koncepcji interakcjonizmu, włączając panpsychizm. Od Kartezjusza oczekiwano modelu mechanicznego, których budował on wiele dla wyjaśnień naukowych. Dziś naukowo zadowalającą odpowiedzią byłoby jakiekolwiek najszerzej rozumiane wyjaśnienie fizyczne czy komputerowe.

Postulat redukcjonisty pod adresem interakcjonisty — „Uwierzę Ci, gdy pokażesz mi mechanizm" — jest nie do przyjęcia dla obu stron. Jak wykazaliśmy, świadomość wykracza poza poznanie naukowe. Żaden mechanizm ją wyjaśniający nie istnieje, a więc nie ma sensu go szukać. Dla redukcjonisty zaś istnienie takiego mechanizmu byłoby argumentem na rzecz tego, że interakcjonizm jest błędny. A więc taki mechanizm, zamiast być dowodem «za» byłby dowodem «przeciw» interakcjonizmowi. Utwierdziłby więc a nie zmienił pogląd redukcjonisty. Nie jest więc rzeczą właściwą, aby wymagać go jako warunku dla zmiany przekonań.

Przewaga ewolucyjna świadomości

Z punktu widzenia teorii ewolucji uzasadnia się istnienie świadomości przez próbę wskazania na jej przewagę ewolucyjną. Ale z moich wcześniejszych rozważań wynika, że jakakolwiek przewaga ewolucyjna, wyrażona za pomocą mechanizmu naukowego, jest także przewagą ewolucyjną owej wcześniej opisanej bez-świadomej istoty. Nie możemy więc twierdzić, że jest to ewolucyjna przewaga jakiegoś elementu naszej świadomości. Przyjemności, przykrości, czy uczucia moralne można analizować w kontekście przeżycia osobnika bądź grupy oraz wyprodukowania liczniejszego potomstwa. Na pewno zwierzęta czy ludzie, którzy podążaliby w kierunku przykrości a unikali przyjemności, zostaliby szybko wyeliminowani ze świata. Ale by móc serio twierdzić, że treści wewnętrzne przyjemności i przykrości zostały wytworzone przez ewolucję, trzeba pokazać, że składają się z dostępnych dla mechanizmu ewolucji części i funkcjonują jako struktury budowane pod kontrolą kodu genetycznego. Ewolucja może jedynie budować z dostępnych elementów. Kości z substancji mocniejszej niż stal, lżejszej niż aluminium, niełamliwej i samoregenerującej, dawałyby nam wielką

przewagę ewolucyjną, ale zapewne nie ma substancji i metody, za pomocą której kości takie mogłyby być wytworzone w żywych organizmach.

Ktoś może powiedzieć, że świadomość istnieje, co dowodzi, że można ją skonstruować. Co więcej, analiza naukowa naszego organizmu natrafia wyłącznie na atomy i struktury z nich złożone, a więc tylko one mogą służyć nauce jako substancja, z której konstruuje się świadomość. Ale jest to aprioryczne podejście do świadomości, które pozostaje aktem wiary, dopóki nie zbuduje się odpowiedniego modelu. A tego zrobić się nie da. Trzeba by pokazać, że określone struktury molekularne, budowane pod kontrolą DNA, i kolejne struktury nad nimi nadbudowywane, prowadzą do obiektów posiadających świadomość i wykazują przewagę ewolucyjną w porównaniu z innymi obiektami żywymi nie posiadającymi tych struktur. Możemy więc uzasadnić przewagę ewolucyjną danych struktur, ale przewaga ta dotyczy w równym stopniu istoty bez-świadomej. Nasz dowód przez wskazanie zbędności treści wewnętrznych stosuje się oczywiście do modeli ewolucyjnych.

Postęp nauki i modeli komputerowych

Systemy komputerowe w zakresie sztucznej inteligencji odtwarzają różne mechanizmy myślenia, funkcjonowania pamięci i innych zachowań związanych ze świadomością. Wiele twierdzeń, że komputery nie są zdolne do wykonania jakiejś konkretnej funkcji, zostało sfalsyfikowanych przez budowę odpowiednich systemów komputerowych. Daje to podstawę, by wierzyć, że każda dobrze zdefiniowana klasa zachowań, które wiążą się z jakąś rolą świadomościową, może być reprezentowana przez program komputerowy. Ale jest zasadnicza różnica między jakąś pojedynczą, dobrze zdefiniowaną ludzką możliwością — a nieograniczoną różnorodnością i wielością ludzkich możliwości i kontekstów, w których nasz umysł może się nimi wykazać. Ograniczone rozwiązania wielu problemów nie są równoważne pełnej symulacji ludzkiego myślenia, funkcjonalnie mu równoważnej. Wcześniej już dowodziłem, że żaden system komputerowy nie odtworzy w pełni procesów psychicznych w nas zachodzących. Znaczy to, że nie będziemy nigdy postawieni przed sytuacją, w której bez-świadomy system komputerowy może wykazać się tymi samymi możliwościami, co świadomy umysł ludzki. Rola świadomości jest zasadnicza, nie reprezentowalna przez żaden mechanizm. Paradoksalnie, sukcesy inteligentnych systemów komputerowych będą lepiej ilustrowały granice między człowiekiem a robotem i unaoczniały wewnętrzne treści naszej świadomości.

Nauka nie daje nam żadnych konstruktywnych wskazówek na temat treści wewnętrznych, związanych ze strukturami materialnymi, i wydaje się, że ich nigdy nie dostarczy. Postęp nauki w empirycznym badaniu świadomości, «tkwiącej» w obiektach materialnych, będzie bardzo powolny, gdyż metoda naukowa odnosi sukcesy przez analizę prostych sytuacji. Ponieważ zjawiska świadomości występują tylko w strukturach złożonych, a nie dostrzegamy ich w układach prostych, jest mała szansa byśmy mogli empirycznie uchwycić sytuację, w której pojawia się świadomość. Eksperymenty

wokół tej granicy dawałyby szanse badania przyporządkowania struktur fizycznych i odpowiadających im zjawisk świadomości. Podobieństwo zachowań, które jest podstawą przekonania o istnieniu stanów psychicznych u zwierząt, załamuje się, gdy rozważamy obiekty coraz mniejsze. Podejście analityczne, właściwe dla metody naukowej, dzieli obiekty i procesy na prostsze składniki i bada je osobno. Nie wiadomo, jak można by w tych składnikach stwierdzić stany świadomości. Na poziomie bakterii, cząstek białka, bądź DNA wydaje się to niemożliwe. Nie wiadomo też, jak łączyć stany świadomości, znalezione w prostych sytuacjach, w stany właściwe sytuacjom złożonym.

Etyczne problemy eksperymentów nad istotami świadomymi powodują dalsze ograniczenie metody analizy naukowej. Nawiasem mówiąc, nie mamy żadnych problemów moralnych z eksperymentami na robotach, bo nie wierzymy, aby posiadały one świadomość. Komputerowe rekonstrukcje świadomości pozwalają więc na podejście analityczne. Choć system komputerowy, za pomocą którego staramy się rekonstruować świadomość, może być bardzo złożony, to jego funkcjonowanie można przeanalizować w każdym szczególe. Można też przeprowadzać różnorodne eksperymenty, wydzielając w nim różne części i modyfikując je na różne sposoby.

Niektórzy uważali, że istnieją substancje bardziej złożone niż poszczególne osoby: «duch narodu» czy «świadomość społeczna». Mamy wrażenie, że substancje takie nie istnieją — i wskazujemy, że zjawiska społeczne są «wypadkową» decyzji i oddziaływań poszczególnych osób. Można co prawda wskazać na zadziwiającą trwałość, z jaką istnieją wspólnoty narodowe i religijne, ale czy możemy stąd przejść do wniosku, że istnieje «duch narodu»? Jest to problem tego samego rodzaju, co próba wykazania istnienia świadomości na podstawie materialnych zachowań.

Umysły nasze są anomalią w wizji świata stworzonej przez uniwersalizm naukowo-komputerowy. Akceptacja interakcjonizmu nie przeszkadza nam być naukowcami i konstruktorami inteligentnych systemów komputerowych. Trudno o bardziej fascynujące zajęcie — niż próby zrozumienia naszego umysłu. W trakcie tych poszukiwań zbudujemy zapewne wiele użytecznych systemów komputerowych, ale nie stworzymy umysłu, który działałby podobnie do umysłu ludzkiego.

Podsumowanie

Starałem się wykazać, że świadomość, która przejawia się w naszych wewnętrznych przeżyciach, jest zbędnym dodatkiem do modeli fizycznych i komputerowych. Z punktu widzenia naukowego redukcjonizmu i komputerowego funkcjonalizmu nasze przeżycia sa ornamentem bez wpływu na modele naukowe i komputerowe. Jeśli towarzyszą obiektom fizycznym, to dzieje się to w sposób niemożliwy do uchwycenia przez żaden opisywalny mechanizm. Całe bogactwo naszych stanów wewnętrznych jest poza zakresem wyjaśnień naukowych. Podejście naukowe modeluje istoty bez-świadome, ograniczając się do zgodności z naukowo dostępnymi faktami o nas. Gdyby modele komputerowe były w pełni skuteczne, można by je traktować jako istoty zachowujące

się tak samo jak my, ale nie posiadające wewnętrznych treści świadomości. Oczywiście jest to wada z naszego punktu widzenia, bo istoty te nie dostrzegałyby jej, nie wiedząc, czym są owe treści wewnętrzne. Modele samych siebie, które by istoty te budowały, mogłyby być zgodne z całokształtem dostępnych im danych empirycznych. Ponieważ istniejemy jako istoty świadome, nie są możliwe adekwatne modele naszej świadomej działalności, komputery i roboty dostarczają zaś nowych argumentów na rzecz różnicy między nami a tym, co mogą osiągnąć bez-świadome istoty.

Istnienie wewnętrznych treści naszej świadomości nie pozwala zaakceptować redukcjonizmu. Empiryczna analiza tych treści i ich udział we wszystkim, co robimy, nakazują przyjąć stanowisko interakcjonizmu. Doświadczenia z modelowaniem procesów poznawczych i decyzyjnych przy użyciu komputerów i robotów dostarczać będą nowych argumentów na rzecz różnicy między nami a istotami bez-świadomymi, a więc nowych argumentów przeciw redukcjonizmowi.

* * *

Pragnę podziękować Małgorzacie Żytkow, która pomogła mi lepiej zrozumieć i przedstawić wiele idei tego artykułu. Za cenne uwagi dziękuję też Piotrowi Wreśniewskiemu.