Włodzisław Duch

Katedra Metod Komputerowych, Uniwersytet Mikołaja Kopernika,
ul. Grudziądzka 5, 87-100 Toruń

duch@phys.uni.torun.pl

Jaka teoria umysłu w pełni nas zadowoli?

ABSTRAKT

Naturalistyczne rozwiązanie problemu umysłu jest w zgodzie z coraz większą liczbą danych eksperymentalnych. Jednocześnie współczesne dyskusje na temat natury umysłu, a zwłaszcza świadomości, opierają się na fantazjach usiłując rozwiązać nieistniejące problemy. Odróżnienie rzeczywistych problemów nauk o poznaniu od problemów pozornych otwiera drogę do akceptacji rozwiązań naturalistycznych i prowadzi do zadawalającej teorii umysłu odpowiadającej na trudne pytania dotyczące świadomości. Należy odróżnić przynajmniej dwa znaczenia pojęcia "zrozumieć": operacyjnie definiowalne zrozumienie intelektualne i subiektywnie definiowalne zrozumienie egzystencjalne. To rozróżnienie, wraz z ograniczeniami funkcjonalizmu, pozwala podważyć argumenty odmawiające możliwości powstania rozumienia i świadomości w systemach sztucznych. Utożsamienie świadomości z procesami poznawczymi wynikającym ze zdolności do dyskryminacji ciągłych reprezentacji wewnętrznych na poziomie globalnej dynamiki mózgu wydaje się być dobrą podstawą dla naturalistycznej teorii umysłu.

1. Wstęp 16

Jak mogłaby wyglądać teoria umysłu, która w pełni zadawalałaby większość badaczy? W materii tej panuje nadal "zamęt i brak zgodności poglądów" (Searl 1999, s. 323). Co oznacza "zrozumienie umysłu"? O jakiego rodzaju zrozumienie tu chodzi? Z jednej strony Noam Chomsky (Horgan 1999, s. 192) twierdzi, że nauka nie zrobiła absolutnie żadnego postępu w badaniach świadomości czy wolnej woli. Z drugiej strony w naukach kognitywnych obowiązujące kredo jest bardzo proste: umysł jest tym, co robi mózg (por. Crick 1998, Dennett 1991), a więc wyjaśnienia neurobiologiczne wystarczą. Naturalizm biologiczny nieco zgrabniej formułuje John Searl (1999, s. 15): "zjawiska mentalne są wywoływane przez procesy neurofizjologiczne zachodzące w mózgu, one same zaś są własnościami mózgu".

W ostatnim ćwierćwieczu nauki o mózgu, opierając się na takim rozumieniu umysłu, poczyniły ogromny postęp w wyjaśnianiu różnych aspektów zachowania zwierząt i ludzi. Zakładając, że kredo kognitywistyki jest słuszne można wątpić, czy w przyszłości wiedza ta, przełożona na język zrozumiały dla laików, ulegnie znacznej zmianie. Na poziomie podręczników szkoły podstawowej prawdopodobnie nie pojawią się żadne nowe elementy, na poziomie szkoły średniej zapewne nieliczne, poważniejszych zmian spodziewać się można jedynie w podręcznikach uniwersyteckich i publikacjach specjalistycznych. Oczekiwania wielkiej rewolucji w nauce, wynikającej z rzekomej konieczności nowej teorii umysłu (por. Penrose 1994, Gardner 1996) są prawdopodobnie błędne. Jeśli wiedza ogólna dotycząca działania mózgu się nie zmieni, to czy umysł może przestać nas zadziwiać?

Filozofia umysłu skoncentrowała się w ostatnich latach nad dyskusjami dotyczącymi świadomości, problemu jakości wrażeń, intencjonalności i problemów pokrewnych. Jakiego rodzaju rozumienie będzie tu zadawalające? Przede wszystkim należy przedyskutować znaczenie pojęcia "rozumieć". W tym względzie w filozofii dominują teorie formalne, oparte na logice, podczas gdy bardziej istotne są związki z procesami zachodzącymi w mózgu. W następnym rozdziale uzasadniam

konieczność odróżnienia rozumienia intelektualnego, weryfikowalnego w obiektywnym sensie, od rozumienia egzystencjalnego, które jest subiek tywnym stanem umysłu. Następnie przedstawię kilka problemów, którymi zajmują się filozofowie umysłu próbując wykazać błędne podstawy ich rozumowania. Spróbuję też uza sadnić, że właściwym podejściem do teorii umysłu, włączając w to świadomość, jest rozważenie zbieżności zachowań ciągu coraz to bardziej dokładnych modeli mózgu do zachowań właściwych dla umysłu. Z tego punktu widzenia widać, że świadomość jest procesem po znawczym. Prowadzi to do – bez wątpienia iluzorycznego – poczucia "zrozumienia", czym jest umysł. Na zakończenie uzasadniam dlaczego takie podejście do rozumienie umysłu wydaje mi się zadowalające.

2. Mózg i rozumienie umysłu

Dyskusji wymaga przede wszystkim samo pojęcie "zrozumienia". Umysł nie jest jednolitą, monolityczną konstrukcją, prostą "substancją duchową". Przypomina raczej zbiór współpracujących ekspertów, "społeczeństwo umysłów" Marvina Minsky'ego (Minsky 1985) lub "społeczeństwo mózgów" Waltera Freemana (Freeman 1996). Zrozumienie intelektualne, abstrakcyjne, jest czymś innym niż zrozumienie egzystencjalne, związane z bezpośrednim przeżyciem jakiegoś stanu umysłowego. W języku potocznym, jak i w filozofii umysłu, te dwa rodzaje rozumienia nie są wyraźnie odróżniane, co prowadzi do licznych nieporozumień. Mechanizmy działania mózgu są w obu przypadkach całkiem odmienne.

Zrozumienie intelektualne to w istocie obiektywne "rozumienie operacyjne", pozwalające sensownie odpowiadać na pytania związane z jakąś teorią, wymagające analizy konstrukcji gramatycznych lub matematycznych, angażujące płaty czołowe i skroniowe mózgu. Tak jest również w odniesieniu do teorii dotyczących własnej jaźni. Zrozumienie intelektualne jest empirycznie weryfikowalne. **Zrozumienie egzystencjalne** związane jest z subiektywną zdolnością do przeżywania wrażeń i stanów emocjonalnych. Pozwala nam to zrozumieć drugiego człowieka, angażując obszary sensoryczne kory mózgowej i struktury układu lim bicznego mózgu, związane z pamięcią epizodyczną i ekspresją emocji. Chociaż jako stan subiektywny organizmu zrozumienie egzystencjalne nie jest chwilowo empirycznie weryfikowalne, można się spodziewać, że postępy w dziedzinie obrazowania stanów mózgu po zwolą już w niedalekiej przyszłości na w miarę precyzyjne określenie stanu emocjonalnego, a nawet kategorii świadomie przeżywanych wrażeń (por. Tononi i Edelman 1998). Obiektywna weryfikacja zrozumienia egz ystencjalnego mogłaby też polegać na badaniu auten tyczności reakcji emocjonalnych i określaniu podobieństwa stanów mózgu badanej osoby do stanów wzorcowych. Nie usunie to oczywiście wewnętrznego, nieredukowalnego punktu widzenia pierwszej osoby, związanego z indywidualnym sposobem przezywania świata. Może jedynie pokazać, że procesy wewnętrzne zachodzą obiektywnie w oparciu o stany neurofizjologiczne mózgu.

Podział ten przypomina rozróżnienie pamięci semantycznej i epizodycznej, zarówno pod względem funkcjonalnym jak i z punktu widzenia obszarów mózgu zaangażowanych w wykonywanie tych funkcji. Przypomina to również rozróżnienie Raya Jackendoffa "umysłu obliczeniowego" i "umysłu fenomenologicznego". Obydwa rodzaje zrozumienia są ze sobą powiązane, gdyż wszystkie procesy zachodzące w mózgu są ze sobą, przynajmniej pośrednio, powiązane (w tym przypadku istotne są powiązania układu limbicznego i płatów czołowych, wpływające na rozumowanie, por. Damasio 1999). Można się więc spierać, że nic takiego "naprawdę" w mózgu nie ma miejsca, jest to tylko kwestia naszych interpretacji stanów neurofizjologicznych (por. Searl 1990). Wszystkie koncepcje odnoszące się do mózgu, czy też do wszelkich obiektów fizycznych, są

metaforyczne a podziały sensowne tylko w przybliżeniu, jednakże teoria nie może się obyć bez koncepcji. Postaram się pokazać, że odróżnienie rozumienia intelektualnego i egzystencjalnego na obecnym etapie rozważań nad umysłem jest przydatne i uzasadnione.

Poczucie zrozumienia towarzyszące rozumieniu intelektualnemu wydaje się być wynikiem poznawczej interpretacji stanu mózgu będącego sygnałem zakończenia procesu analizy jakiejś porcji informacji (np. zdania). Może to, chociaż nie musi, wywołać wtórnie poczucie zrozumienia egzystencjalnego i związane z tym emocje. Słuchając złożonego zdania w cza sie wykładu lub czytając trudne zdanie w książce mózg rezerwuje większość swoich mocy przetwarzania informacji do analizy tego zdania, po zakończeniu przesyłając sygnał gotowości do dalszego działania: “zrozumiałem, co dalej?” Jest to konieczne ze względu na sto sunkowo długi czas, potrzebny do analizy postrzeganej sytuacji lub sensu zdania. Poczucie zrozumienia może być jednak zwodnicze i dopiero próba odpowiedzi na pytania egzaminacyjne upewnia nas, na ile udało się nam naprawdę zagadnienie zrozumieć. Bardzo silne po czucie “zrozumienia wszystkiego” wywołać mogą substancje halucynogenne. Fałszywe przekonanie, że się rozumie może się też pojawić w niektórych chorobach umysłowych. Z drugiej strony może być również odwrotnie, możemy coś rozumieć w sensie intelektualnym zanim nie pojawi się poczucie zrozumienia. Byłoby więc rzeczą ryzykowną wiązać wrażenia towarzyszące procesom rozumienia z samym rozumieniem. W szczególności można sobie wyobrazić systemy przetwarzające inform ację, zdolne do rozumienia w sensie operacyjnym, chociaż nie posiadające wrażeń, a w szczególności poczucia rozumienia.

Analiza zawartości semantycznej z logicznego punktu widzenia była od dawna dyskutowana w filozofii (por. Putnam 1975, 1987), jednakże przeprowadzone powyżej rozróżnienie po między zrozumieniem egzystencjalnym a operacyjnym wykracza poza obszar filozofii języka. Zmiana nastawienia, pojawienia się poczucia “rozumiem” i związanych z nim przeko nań, wydaje się być głębszym procesem, związanym z działaniem mózgu na poziomie bardziej podstawowym niż procesy myślenia. Wiele koncepcji uznanych zostało za “zrozumia łe” dopiero po bardzo długim okresie od ich powstania, zapewne dopiero po wprowadzeniu ich do powszechnego nauczania. Zrozumienie ruchu było przez dwa tysiące lat bardzo trudnym zagadnieniem i nawet Kepler wierzył w harmonię sfer i anioły popychające planety. Newton i jego następcy sądzili, że działanie na odległość jest możliwe tylko przy założeniu, że przestrzeń rozumieć należy jako boskie sensorium (por. Gregory 1981). Podobne trudności miał Maxwell, wyobrażając sobie mechaniczny eter, w którym rozchodzić się miały fale elektromagnetyczne. Jeszcze w XVII wieku wielcy matematycy twierdzili, że zrozumienie, dlaczego dla niewiadomych i ich znaków (*-a)(-b)=ab,* przekracza możliwości ludz kiego umysłu. Podobnie chemicy po odkryciu, że woda składa się z dwóch gazów nie mogli w to uwierzyć.

Dlaczego obecnie zagadnienia te wydają się nam zrozumiałe? Być może dlatego, że udało się nam je głęboko wbudować w intelektualny model świata dostatecznie wcześnie, w okresie zwiększonej plastyczności mózgu. Umysł podąża za takimi głęboko zakorzenionymi skojarzeniami w naturalny sposób przestając się dziwić. Dużo większe trudności sprawia nam zrozumienie pojęcia zakrzywionej czasoprzestrzeni czy dualizmu falowo-cząstkowego mechaniki kwantowej, gdyż uczymy się o nich zbyt późno i nie mamy z nimi do czynienia na co dzień, a więc pojęcia te są znacznie słabiej związane z dobrze ugruntowanymi koncepcjami, będącymi podstawą naszego sposobu myślenia. W efekcie tylko eksperci mają poczucie, że dobrze rozumieją teorię grawitacji czy brak granic skończonego Wszechświata, a większość ludzi nadal zadaje sobie pytanie “co jest poza tymi granicami”, doznając przy tym wrażenia tajemnicy. Rozumienie podstaw mechaniki

kwantowej jest nadal bardzo kontrowersyjne, podobnie jak powstanie materii z próżni (z powodu niestabilności pustej próżni), zwijanie dodatkowych wymiarów w teoriach superstrun i inne koncepcje fizyki teoretycznej . Najlepiej ugruntowane wydają się być koncepcje związane z rozumieniem egzystencjalnym, bliskie percepcji. Należy jednak pamiętać, że fizyka Arystotelesa, niezgodna z prosty mi obserwacjami, a więc nieugruntowana w doświadczeniu (np. paraboliczny tor rzuconego kamienia powinien, zgodnie z Arystotelesem, być prostoliniowy i zakończony gwałtownym spadkiem po wyczerpaniu się "impetu") przetrwała niekwestionowana przez 2000 lat. Widocznie rozumienie wymaga dopasowania nowych faktów do istniejących modeli umysło wych, a te, raz zakorzenione, bardzo trudno jest zastąpić innymi. Naturalną skłonnością umysłu jest odczuwać większe zrozumienie po dodaniu epicykli do znajomego modelu niż do zastąpienia go modelem całkiem nowym.

Równie naturalne jest poszukiwanie model i prostych: mózg jest zbyt skomplikowany, by stanowić dobrą podstawę do dyskusji umysłu, który jawi się nam w doświadczeniu we wnętrznym jako coś jednolitego. Wyjaśnia to z jednej strony dążenie do zrozumienia umysłu na poziomie fundamentalnych praw fizyki (cf. Penrose 1994), a z drugiej popularność najbardziej uproszczonych schematów działania mózgu: podziału funkcji pomiędzy dwie półkule oraz podziału na korę, układ limbiczny i pień mózgu. Podział ten, dokonany przez Paula McLeana (1973, 1990), kierownika Laboratorium Ewolucji i Zachowania się Mózgu w NIMH (National Institute of Mental Health), przypomina również zaproponowany przez Freuda podział umysłu na id, ego i superego. Id, czyli **to** , uznać można za prymitywną natu rę zwierzęcą, realizowaną przez mózgi gadów głównie za pomocą pnia mózgu i podwzgórza. Ego, a więc emocje, struktura osobowości, odpowiadają układowi limbicznemu. Supe rego jest natomiast nośnikiem świadomości społecznej, moralności i odpowiedzialności, cech za które odpowiedzialna jest przede wszystkim kora mózgu.

Filozofia umysłu posługuje się pojęciami i modelami rozwiniętymi w bardzo długim okresie czasu. Chociaż wiele problemów dotyczących natury umysłu ma proste rozwiązania, nie wywołują one odpowiedniego rezonansu w umysłach niektórych filozofów i naukowców. Fizyka i biologia odrzuciła całkowicie średniowieczny obraz świata (Lewis 1995), jednakże część filozofii wydaje się nadal pod jego wpływem, przypisując umysłowi lub przynajmniej neuronom (cf. Searl 1990, Kloch 1996) jakieś tajemne moce przyczynowe. Umysł rozumiany jako funkcja mózgu może się w związku z tym wydawać pomniejszeniem godności człowieka (por. Popper, Eccles 1977). Przetwarzanie informacji, percepcja i funkcje afek tywne są najwyraźniej związane z działaniem mózgu i nowe pokolenia nie będą prawdopodobnie widzieć w tym niczego dziwnego. Na powszechnie obecnie panujące wrażenie ta jemniczej natury umysłu mogą mieć też wpływ niejasne powiązania z takimi (nie dającymi się zrozumieć) pojęciami religijnymi jak duch czy dusza, które były w przeszłości podstawą wszelkich dyskusji w filozofii umysłu.

3. Problemy pozorne i problemy prawdziwe w filozofii umysłu.

   W żadnej gałęzi nauki ani filozofii nie ma tylu dziwacznych poglądów, co w filozofii umy słu. Przywodzi tu grupa filozofów określana jako "neomysterianie". Collin McGinn i jego koledzy atakują możliwość zrozumienia umysłu, wyciągając takie argumenty jak niemożliwość zrozumienia nieprzestrzennej natury umysłu (McGinn 1995). W jaki sposób to, co nieprzestrzenne (umysł), może powstać z tego, co zlokalizowane jest w przestrzeni (mózg)? Tego typu problemy pokazują, do jakiego stopnia myślenie niektórych filozofów jest nadal zakorzenione w średniowiecznych koncepcjach, traktujących umysł jako rodzaj substancji, coś podobnego do przedmiotów fizycznych. Umysł rozumiany jako funkcja mózgu ma strukturę

relacyjną, a świadome treści odpowiadają kolejnym stanom globalnej dynamiki mózgu. Stany te mają pewną strukturę temporalno-logiczną, zależną od zapisanych w mózgu śladów pamięci, które porównać można do kolein wyżłobionych w materii neuronalnej przez doznania zmysłowe i wewnętrzne stany mózgu. Struktura mózgu i docierające do niego przez zmysły dane decydują o przyjmowanych przez niego stanach. Niestety nie mamy dobrych analogii tego procesu. Co prawda na dyskach kompaktowych czy dyskach wi deo zapisane są ślady pamięci stanów dynamicznych przyjmowanych przez membranę mikrofonu czy fotoczułe elementy kamery, ślady aktualizowane w zrozumiałej dla człowieka formie przez odpowiednie elektroniczne urządzenia odtwarzające, jednak struktura tempo ralna tych stanów jest ustalona, nie oddziałują one ze środowiskiem ani nie potrafią się same modyfikować, więc jest to jedynie częściowa analogia. Jednakże nawet w tym przypadku nie ma sensu przypisywać stanom dynamicznym odtwarzacza płyt wideo własności przestrzennych, chociaż urządzenie mieści się w małej skrzynce. Stany te można określić przez podanie aktywności różnych elementów elektronicznych, podobnie jak i stany mózgu okre ślić można przez aktywność neuronów. Istotny jest nie tyle sam opis tych stanów (można je zdefiniować w wysokowymiarowej abstrakcyjnej przestrzeni aktywności neuronów lub logicznych bramek), co relacje pomiędzy nimi.

Paradygmat komputacyjny dostarczył nam dobrych przykładów relacji nieprzestrzennych pomiędzy abstrakcyjnymi obiektami używanymi w czasie obliczeń. Putnam (1975) posługuje się tu analogią stanów mentalnych i fizycznych, wykazujących podobieństwo do stanów logicznych i strukturalnych maszyny Turinga. Maszyny Turinga nie są dobrym mode lem zachodzących w mózgu procesów, lepiej wyobrazić sobie zbiór oddziałujących na siebie rezonatorów elektrycznych (odpowiadających kolumnom korowym mózgu) lub mechanicznych, na wzór sprężyn drgających w dużym materacu. Globalne stany dynamiczne takiego zbioru charakteryzować się mogą złożoną logiką temporalną. Jeśli część z tych rezonatorów specjalizuje się w analizie sygnałów wizualnych, a część w lingwistycznych komentarzach stanów tych pierwszych, otrzymamy system, którego stany wewnętrzne można analizować częściowo stosując relacje przestrzenne (do obiektów występujących w części analizującej sygnały wizualne) jak i nieprzestrzenne (do symboli lingwistycznych). Mentalna rotacja jest przykładem operacji umysłowej, która dotyczy reprezentacji wizualnych i ma strukturę przestrzenną, ale wrażenia związane ze słuchaniem muzyki już takiej struktury nie posiadają, podobnie jak procesy myślenia abstrakcyjnego w czasie gry w szachy czy rozwiązywania zadań matematycznych. Mózg zlokalizowany jest w czasie i przestrzeni, ale nie ma powo dów by treści umysłu, wynikające z relacji pomiędzy dynamicznymi stanami mózgu, miały strukturę przestrzenną.

Nie tylko filozofowie martwią się takimi pseudoproblemami, również niektórzy fizycy i matematycy uznają nielokalność umysłu za poważny problem. Dochodzi nawet do propozycji traktowania mózgu za obiekt nielokalny i uznania umysłu za kluczowy aspekt wszech świata (Clarke 1995). Nie wpłynie to oczywiście na nasze rozumienie procesów poznawczych czy syndromów neuropsych ologicznych, ma jedynie wyjaśnić nielokalną naturę umy słu, która nie wymaga żadnego wyjaśnienia, gdyż jest problemem pozornym. Dziwne, że nie padają pytania: jak dźwięk lub obraz przesłać można przez drut? Musi tu być jakaś pomyłka kategorialna. Mając dobre przykłady technicznych urządzeń wydaje się nam, że rozumiemy jak to jest możliwe. Brak prostych modeli, w oparciu o które możemy sobie wyobrazić, jak powstają funkcje umysłu, prowadzi do porzucenia lub też zaprzeczenia sensowności modeli kognitywnych (p or. Searl 1990). Dobry przykład takich trudności znajdujemy u Hilary Putnama w serii wykładów “Dewey Lectures” i w dodatku do Royce Lectures II (1998).

Putnam twierdzi, że przekonanie o istnieniu jakichś wewnętrznych reprezentacji umysłowych, pojawiających się na przykład gdy widzimy jakiś obiekt, jest błędne i odpowiedzialne za wiele problemów w filozofii umysłu. Dla klasycznej kognitywistyki w ujęciu Allena Ne wella (1990) reprezentacja wewnętrzna jest pojęciem podstawowym, koniecznym do zrozumienia umysłu. Według Putnama stany umysłu nie odpowiadają żadnym wewnętrznym stanom fenomenalnym, które pełniłyby rolę "wspólnego mianownika" tak, że ich pojawienie w mózgu byłoby równoważne pojawieniu się odpowiadającym im treści doświadczeń świa domych. Jeśli zjawisko (stan umysłu) wydaje się identyczne, nie oznacza to jeszcze identyczności stanów fenomenalnych, gdyż bycie w tym samym stanie jest relacją przechodnią a rozróżnialność stanów umysłu nie. Na przykład jeśli będziemy kolejno malować 100 kart białą farbą, do której przy każdej kolejnej karcie dolewać będziemy jedną kroplę farby czerwonej, dwie kolejne karty w tym ciągu będą miały nierozróżnialną barwę, a jednak karta pierwsza będzie całkiem biała a ostatnia całkiem czerwona. Jeśli jednak karty sąsiednie wydają się identyczne i odpowiadają temu identyczne stany umysłu, to biorąc kolejno sąsiednie pary kart, tj. pary (1,2), (2,3), ... (n,n+1), możemy twierdzić, że za każdym razem mamy do czynienia z tym samym stanem fenomenalnym, gdyż wrażenia są nierozróżnialne. Stąd, twierdzi Putnam, stany fenomenalne dla kart 1 i 100 są takie same. Można podać jesz cze prostszy argument: jeśli patrzymy na małą wskazówkę zegara i co 15 sekund jesteśmy pytani, czy jej pozycja się zmieniła, będziemy mówić, że nie, ale po paru minutach jest ona najwyraźniej różna.

Argumenty logiczne zastosowane do układów o dynamice ciągłej prowadzą do paradoksów znanych już od czasów Zenona z Elei i Parmanidesa. W tym przypadku również nie jest inaczej, jednakże są tu dwa aspekty. Po pierwsze, nie jesteśmy zdolni do rozpoznania dowolnie małych różnic stanów swojego mózgu, a więc identyczność w sensie praktycznym nie oznacza identyczności w sensie matematycznym, a jedynie podobieństwo, przynależność do określonego przedziału. Stosowanie logiki rozmytej czy przybliżonej do kategoryzacji (por. psychologiczne modele kategoryzacji, np. Cohen i Massaro 1992) jest znacznie bar dziej naturalne i nie prowadzi do takich paradoksów, gdyż relacja przechodniości jest wówczas słuszna jedynie w przybliżeniu. Jest tu jednak drugi aspekt, bardziej subtelny aspekt zagadnienia odpowiedniości stanów mózgu i stanów umysłu. Nawet w mózgu węszącego królika mapy elektroencelograficzne (Freeman 1996) wykonane w przeciągu kilku dni dla tych samych bodźców zapachowych, wywołujących te same reakcje królika, są całkiem odmienne. Nie ma mowy, by w takim systemie jak mózg były identyczne stany fenomenalne odpowiadające tym samym wrażeniom. Czyżby więc Putnam miał rację nie z powodów logicznych, lecz empirycznych? Bynajmniej!

Zachowanie mózgu opisać można badając atraktory dynamiki wielkich grup (rzędu milionów) neuronów, zdefiniowane w przestrzeni pobudzeń tych neuronów. Pomimo braku identyczności stanów powierzchniowych mózgu (np. badanych za pomocą EEG) można się w dynamice jego działania doszukać stałych relacji pomiędzy tymi atraktorami, czyli względnie stabilnymi (w skali czasowej rzędu sekund) stanami neurofizjologicznymi. Od powiednio potraktowane matematycznie stany dynamiczne mózgu charakteryzują się więc pewnymi głębszymi relacjami, które czasami mogą mieć prostą reprezentację logiczną, np. w postaci implikacji: jeśli A to B. Odpowiedniość stanów umysłu i mózgu nie dotyczy stanów powierzchniowych, mających naturę efemeryczną, lecz wykazujących się pewną stabil nością stanów atraktorowych. Reprezentacje wewnętrzne muszą być stabilne nie ma jednakże powodu, by były to reprezentacje powierzchniowe. Mówiąc metaforycznie, dopiero przyglądając się z pewnego oddalenia widać ogólny charakter konstrukcji umysłu, która może być trudna do ogarnięcia przy badaniach

mikroskopowych.

Tłumaczy to również pewne paradoksy językowe dotyczące przekonań, często dyskutowane w filozofii umysłu. Podobnie jak wielu innych ludzi jestem przekonany, że “Paryż jest stolicą Francji” (por. Putnam 1987). Jakiego rodzaju stany fenomenalne mózgu odpowiadają tego typu stwierdzeniom? Neurofizjologiczne modele działania mózgu (Duch 1997) można upraszczać tak, by pobudzenia grup neuronów kory mózgu kodujących koncepcje w pamięci przedstawić za pomocą sieci neuronowych z rekurencją, a następnie w postaci jesz cze prostszego modelu w postaci węzłów sieci semantycznej (jest to często spotykany w sztucznej inteligencji sposób reprezentacji wiedzy). Węzły te reprezentują pojęcia “stolica”, “Paryż”, “Francja” i ich powiązania między sobą. Połączenia między nimi symbolizują połączenia grup neuronów biorących udział w kodowaniu danej koncepcji. Można twierdzić, że pomimo powierzchniowych różnic stanów neurofizjologicznych osób, które takie przeko nanie żywią, uproszczenie modelu dynamiki stanów neurofizjologicznych ich mózgu będą zawierać podobne fragmenty sieci semantycznych (dysponujemy w tym zakresie jedynie pośrednimi dowodami z badań na małpach i modelami komputerowymi). Patrząc na szczegóły trudno jest dostrzec ogólniejsze prawidłowości, dlatego ujęcie istoty jakiegoś zjawiska wymaga zwykle abstrahowania od wielu nieistotnych jego aspektów a przejście do sieci semantycznych jest wynikiem takiego abstrahowania. Posiadanie danego przekonania jest więc równoznaczne z zachodzeniem w naszych mózgach pewnych relacji, chociaż w każdym konkretnym mózgu mogą one być realizowane w nieco odmienny sposób.

Na szczęście zagadnienia tego typu można będzie zweryfikować empirycznie w niedalekiej przyszłości – jeśli jakaś forma reprezentacji wewnętrznych istnieje, muszą się one ujawnić jako pewne niezmienniki w badaniach stanów dynamicznych mózgu. Metody obrazowania mózgu rozwinęły się na dobre dopiero w ostatnim dziesięcioleciu a praca Tononi i Edelmana (1998) jest pierwszą udaną próbą powiązania świadomych wrażeń z mierzalnymi obiek tywnie parametrami. Są to jednakże trudności techniczne, związane ze złożonością badanego obiektu i koniecznością prowadzenia badań nieinwazyjnych (w przypadku ludzi), a nie trudności fundamentalne. Kognitywistyka poczyniła w ostatnich latach ogromne postępy i nie ma powodu, by nie udało się zrozumieć dokładniej sposobu działania mózgu. Nie ma też wątpliwości, że możliwości naszego umysłu są ograniczone. W istocie nie jesteśmy zdolni poznać w pełni żadnego obiektu fizycznego, nawet atomu wodoru, który jest nadal intensywnie badany a ostatnich lata przyniosły cały szereg interesujących odkryć związanych z jego zachowaniem w silnych polach czy bardzo ciekawymi stanami wzbudzonymi. Nie oznacza to jednak, byśmy nie mogli poznać ogólnego planu działania mózgu i sposobu, w jaki to działanie wiąże się ze stanami mentalnymi.

W ostatnich latach nasiliły się próby pokazania trudności filozoficznych, związanych z naturą umysłu, a w szczególności z naturą wrażeń świadomych. David Chalmers (1995-1997) wywołał burzliwą dyskusję wokół starego problemu dotyczącego jakości wrażeń fenomenalnych (zwanych z łacińska *qualis*), przedstawiając go jako “trudny problem świadomo ści”. Przetwarzaniu informacji towarzyszą wrażenia niemożliwe do zwerbalizowania, takie jak poczucie smaku lub wrażenia koloru. Komputery, przetwarzające te same informacje nie odczuwają podobnych wrażeń. Czy jest to istotnie trudny problem związany z umysłem? Sądząc po propozycjach jego rozwiązania (Chalmers 1997) może się tak w istocie wyda wać, albowiem żadna z tych propozycji nie oferuje przewidywań lub wyjaśnień dotyczących natury wrażeń świadomych – większość z nich nie traktuje zresztą świadomości jako procesu poznawczego. Chalmers opiera się na dwóch, pozornie całkiem

rozsądnych, zasadach. Pierwsza z nich, zasada strukturalnej koherencji, głosi, że struktura świadomych doświadczeń wynika z treści informacji dostępnej świadomości (jest to równoważne stwierdzeniu, iż dostępność informacji jest warunkiem koniecznym do powstania wrażeń świadomych). Druga zasada, zwana zasadą niezmienniczości organizacyjnej, głosi, że liczy się jedynie spe cyficzna struktura powiązań przyczynowych pomiędzy elementami systemu, a nie jego fizyczna budowa. Jest to oczywiście kredo funkcjonalizmu.

Nie będę tu szczegółowo omawiał wyników dyskusji nad trudnym problemem świadomości, czyli próby odpowiedzi na pytanie skąd biorą się i czym są wrażenia świadome. Zadawala jąca teoria umysłu powinna dawać odpowiedź na konkretne pytania, np. pytania dotyczące warunków powstawania wrażeń świadomych w mózgu, czasu koniecznego do ich powstania, przyczyn złudzeń wzrokowych czy słuchowych, halucynacji i symptomów schizofrenii, struktury wrażeń różnej modalności itp. Teorie wymyślone specjalnie dla rozwiązania jed nego, pozornie istotnego pytania: “czym są jakości wrażeń”, nie mogą być poznawczo płodne i żadna z przedstawionych w książce Chalmersa (1996) teorii taką się nie wydaje. Z drugiej strony zadawalająca teoria powinna również wyjaśnić przyczyny istnienia świadomych wrażeń, a nie tylko stwierdzić, że są one tylko kwestią dyspozycji lub wynikiem sta nów neurofizjologicznych mózgu Dlaczego tak trudno podać sensowne rozwiązanie? Być może winę ponosi jedna z oczywistych zasad, przyjmowanych za podstawę dyskusji.

4. Problemy funkcjonalizmu

Zasada organizacyjnej niezmienniczości jest podstawą licznych eksperymentów myślowych, stanowiąc podstawę rozumowania wielu filozofów umysłu. Warto jednak pamiętać, że jest to zasada jedynie przybliżona, a jej zastosowanie do wymieniania neuronów mózgu na sztuczne elementy (por. Chalmers 1995) jest wielce wątpliwe. Cóż to bowiem oznacza, że z funkcjonalnego punktu widzenia elementy są identyczne? Elementy krzemowe i elementy białkowe oparte na węglu są w istotny sposób różne i nie ma sposobu, by jedne zamienić drugimi nie wpływając na pracę całości. Oddziaływania atomów i związków węgla są odmienne niż oddziaływania atomów i związków krzemu, różne są wiązania chemiczne, nie można utworzyć podobnych kanałów jonowych ani zrobić krzemowych struktur reaguj ących w ten sam sposób co białka na neurotransmitery obecne w szczelinach synaptycznych. Kwantowe własności różnych atomów nie pozwalają na ich funkcjonalnie identyczną zamianę nawet “w zasadzie”. Nie można więc eksperymentów myślowych opartych na wy mianie kolejnych neuronów na krzemowe płytki traktować poważnie.

Możliwość zamiany wrażeń związanych z widzeniem koloru czerwonego i zielonego przez zamianę połączeń w mózgu (*inverted qualia*) jest również czystą fantazją (por. np. Dennett 1991 czy Searl 1999, gdzie argumenty tego typu są stosowane), albowiem konieczna byłaby całkowita reorganizacja połączeń w większej części kory mózgu, związanych z pamięcią epizodyczną zawierającą odniesienia do koloru. Takie zamiany muszą prowadzić do daleko idących zaburzeń w działaniu mózgu a eksperymenty myślowe, które zakładają, że tak nie jest, są po prostu błędne. Mózg nie jest układem elektronicznym z przewodami, które można bezkarnie przekładać. Identyczność funkcji w granicy dokładności tworzenia modeli mózgu implikuje identyczność struktury fizycznej.

Moje wątpliwości budzi samo twierdzenie, że wrażenie czerwieni i wrażenie zieleni, niezależnie od wszelkich dyspozycji pamięci, jest odmienne, tak, że jest między nimi jakościowa różnica na poziomie mentalnym. Osobiście nie jestem skłonny przyznać, że są to wrażenia odmienne. Wpatrując się w kolorową płaszczyznę mam

różne skojarzenia, lecz jeśli je odrzucę, mogę jedynie powiedzieć, że mam wrażenie wzrokowe. Sądziłem, że jest to wynikiem mojej ma rnej pamięci kolorów, ale podobnego zdania były również pytane przez mnie osoby obdarzone dobrą pamięcią wzrokową. Zapewne niektórzy ludzie wyrażą przekonanie, że są to wrażenia odmienne, sądzę jednak, że jest to zagadnienie kontrowersyjne i nie należy pochopnie przyjmować argumentów na nim opartych. Jest natomiast rzeczą niewąt pliwą, że wrażenia związane z kolorami są wynikiem zdolności do dyskryminacji, pamięci kolorów i uczenia się informacji wzrokowej. Wystarczy spędzić pół godziny nad kolorową układanką (*puzzle*) by zauważyć, jak nasz układ wzrokowy reaguje na subtelne odcienie barw, których wcześniej wcale w układanym obrazie nie dostrzegaliśmy. Nowe wrażenia związane są z pewnością z działaniem pamięci i są wynikiem zdolności układu wzrokowego do dyskryminacji barw. Przekonanie o naszej zdolności do odmiennego postrzegania "jakości kolorów" wynika zapewne z reakcji emocjonalnych, jakie są rezultatem pobudzeń pa mięci, a przez to pośrednich pobudzeń układu limbicznego, wywołującego wrażenia emocjonalne. Jeśli odrzucę jednak takie wrażenia wtórne nie pozostaje w mojej świadomości nic, co pozwoliło by mi powiedzieć, że mam istotnie odmienne wrażenie. Pozostaje tylko odróżnienie wrażenia wzrokowego od tła. To samo dotyczy innych modalności: wszelkie wrażenia mają jakości dzięki możliwości porównania, wynikającej z istnienia pamięci poprzednich wrażeń, oraz mechanizmów dyskryminacji stanów o różnych modalnościach na poziomie globalnej dynamiki mózgu.

Na problemy w zastosowaniu zasady niezmienniczości organizacyjnej wygodnie jest patrzeć z punktu widzenia osiągalnej dokładności aproksymacji funkcji realizowanych przez funkcjonalne odpowiedniki. W przypadku elementów cyfrowych możliwa jest doskonała emulacja, w przypadku neuronów jedynie gruba aproksymacja. Inteligentne przetwarzanie infor macji może być możliwe w oparciu o krzem, nie ma jednak powodu by przypuszczać, że stany neurochemiczne, decydujące o naszych wrażeniach mentalnych (por. Black 1994), da się aproksymować dostatecznie dokładnie za pomocą urządzeń cyfrowych. Nawet niewielkie zaburzenia struktury elektrochemicznej mózgu prowadzą do istotnych zmian na poziomie zjawisk mentalnych (np. do chorób umysłowych). Zjawiska te są rezultatem globalnej dynamiki mózgu, zależnej od stanu wszystkich znajdujących się w nim struktur. W mózgu nie ma słabo sprzężonych obszarów, które można wyciąć nie zmieniając jego działania, chociaż czasami zmiany te mogą być trudno zauważalne.

Podana tu argumentacja prowadzi do wniosków zgodnych z naturalizmem biologicznym Searla (1999): neurofizjologia mózgu decyduje o charakterze umysłu. Nie oznacza to, by inteligencja czy jakaś forma umysłu nie mogły powstać z powodu oddziaływań elementów opartych na krzemie czy innych nie-węglowych związkach. Takie umysły będą jednak miały odmie nną naturę od naszego, ich sposób przeżywania świata będzie odmienny. Różne struktury mózgu muszą być przyczyną różnych umysłów. Można się jedynie zastanawiać, co jeszcze warto nazwać umysłem, a więc jaki minimalny zbiór własności przejawiać powinien dany system by można jeszcze mówić o umyśle. Załóżmy, że jest to intencjonalność, rozumienie i świadomość. Czy cechy te przejawiać mogą tylko systemy biologiczne? Ewo lucja zachowań robotów, sterowanych przez programy zachowujące mózgopodobną organizację przetwarzania informacji, którym zaprogramowano jedynie ogólne wartości (takie jak "lepiej jest mieć wrażenia, niż ich nie mieć"), pozwala uznać ich zachowanie za krok w kierunku intencjonalności (por. Edelman 1999). Nie widać zasadniczych powodów, dla których granicą zbieżności coraz doskonalszych modeli tego typu nie miałyby być systemy intencjonalne, stawiające sobie coraz bardziej złożone cele.

John Searl próbował pokazać, że systemy obliczeniowe nie są zdolne do przekroczenia bariery semantyki. Argument oparty na "chińskim pokoju" (szczegółowo omówiony w: Kloch 1996) miał podważyć wiarę specjalistów od sztucznej inteligencji w test Turinga. Czy samo przejście testu Turinga wystarcza do uznania programu komputerowego za równoważny umysłowi? Skoro człowiek, będący częścią systemu realizującego program do konwersacji w języku chińskim, nie rozumie ani pytań, ani odpowiedzi, nie należy sądzić, że program komputerowy je rozumie, chociaż daje sensowne odpowiedzi. Niektórzy filozofowie uznali, że argument ten całkowicie unieważnia test Turinga, zapominając przy tym, że przejście testu Turinga, niezależnie od jego interpretacji, oznaczać będzie powstanie sztucznego my ślenia znacznie przekraczającego możliwości naszego mózgu (choćby ze względu na szybkość działania i zdolność do dowolnie długiej koncentracji). Tymczasem argument Searla nie jest testem, gdyż stosowany jest w taki sposób, że w każdym przypadku daje negatywny wynik. Stosując go nie widać rozumienia nie tylko w systemach formalnych, ale i w naszych mózgach: umieszczenie obserwatora w mózgu Chińczyka nie da mu poczucia rozumienia języka chińskiego. Już Leibniz w swojej *Monadologii* zrozumiał, że jeślibyśmy weszli do myślącej, czującej i postrzegającej maszyny nie znaleźlibyśmy w jej środku fruwających myśli ani nie dostrzegli sensu w obrotach jej mechanizmu. Zamiast rozważyć warunki ko nieczne dla rozumienia przez człowieka Searl twierdzi, że skoro ludzie rozumieją dzięki swoim mózgom, to musi w nich tkwić jakaś siła sprawcza, tajemnicza "siła przyczynowa" neuronów. Dlatego ludziom można przypisać rozumienie a programom komputerowym nie.

Rozumowanie to łatwo odwrócić. Roboty, wyposażone w programy zdolne do przejścia testu Turinga, mogłyby na podstawie argumentu chińskiego pokoju dojść do dokładnie odwrotnych wniosków: ponieważ wiemy, że rozumiemy, to elementy krzemowe muszą mieć jakąś siłę sprawczą, której zapewne elementy białkowe nie mają. Skąd się więc bierze siła argumentu Searla? Moim zdaniem z braku dyskusji sensu pojęcia "rozumieć", które w jego pracy utożsamiane jest z rozumieniem opowiadań w językach obcych. Rozumie on opowiadania w języku angielskim i możemy to zweryfikować zadając mu pytania, natomiast nie rozumie wcale języka chińskiego, i to również można zweryfikować, zadając mu pytania. Do rozumienia intelektualnego i przypisania komuś (człowiekowi lub maszynie) inteligencji konieczne jest przejście testu Turinga, a nie poczucie rozumienia. Nawet jeśli nie mam ta kiego poczucia mogę "intuicyjnie" odpowiadać na pytania, a mój egzaminator może uznać, że rozumiem zagadnienie. "Intuicyjnie" oznacza tu jedynie tyle, że pomimo braku poczucia zrozumienia i przekonania, że zagadnienie rozumiem, jestem w stanie dać prawidłowe odpowiedzi. Być może nie musimy wymagać od inteligentnego programu komputerowego, by nie tylko rozumiał pytania w sensie operacyjnym (co możemy zweryfikować sprawdzając sens odpowiedzi), ale jeszcze miał poczucie, że je rozumie. Rozumienie intelektualne nie zawsze jest związane z rozumieniem egzystencjalnym.

W jaki sposób moglibyśmy się przekonać, że jakiś system rozumie w sensie egzystencjalnym? Searl (1999) opiera się tu na mocy przyczynowej neuronów, której nie potrafi jednak bliżej określić. Wydaje się, że wystarczającym warunkiem jest "wejście w rezonans" z takim systemem, przez konwersację lub obserwację jego zachowania. Nasze możliwości zrozu mienia innych ludzi lub zwierząt są uwarunkowane nie tylko podobieństwem naszych mózgów, lecz również kultury, w której wzrastaliśmy, pozwalającej na interpretację pewnych zachowań. Jesteśmy zdolni do egzystencjalnego rozumienia jedynie tego, co jest w miarę podobne do nas. Gdyby struktura "mózgu" sztucznego systemu była dostatecznie podobna do naszej można by sobie wyobrazić (przynajmniej jako eksperyment myślowy) bezpośrednie pobudzanie jednego mózgu przez drugi, a więc pewnego rodzaju

telepatyczną łączność. W ten sposób obserwator w chińskim pokoju mógłby nie tylko zacząć rozumieć, ale i miałby poczucie rozumienia. Jakie są warunki pojawienia się rozumienia w jego umyśle? Rozumiemy tylko te rzeczy, których się nauczyliśmy, które możemy skojarzyć z już posiadaną wiedzą. Nauczenie się języka chińskiego tak, by możliwe stało się rozumienie w egzystencjalnym sensie, wymaga rozpoczęcia od asocjacji symboli z wrażeniami zmysłowymi. We wnętrzna reprezentacja wrażeń w mózgu ani w systemach sztucznych nie da się przedstawić w postaci symbolicznej. Obserwator wewnątrz systemu mógłby jednak skorzystać z przekładu sygnałów elektrycznych z kamery czy mikrofonu na zrozumiałe dla niego wraże nia zmysłowe (np. odtwarzania obrazów czy dźwięków w formie zrozumiałej dla człowieka) i nauczyć się właściwych skojarzeń, dzięki czemu jego rozumienie formalne stało by się również rozumieniem egzystencjalnym.

Nie sądzę, by przejście pełnego testu Turinga bez zrozumienia z sensie egzystencjalnym było możliwe, separacja pomiędzy intelektem a uczuciami jest bowiem niezbyt głęboka (por. Damasio 1999). Bez implementacji zachowań związanych z uczuciami możliwa jest jedynie zgrubna aproksymacja działania umysłu, choć nie jest rzeczą jasną, jakiego rodzaju pytania przekraczać będą możliwości systemu dysponującego jedynie rozumieniem intelektualnym. Czy dokładniejsza aproksymacja jest możliwa i na czym miałaby polegać?

5. Zbieżność funkcjonalna modeli mózgu

Płodnym punktem widzenia na relację umysłu i mózgu jest próba określenia ciągu modeli, prowadzących od szczegółowego opisu procesów zachodzących w mózgu na poziomie molekularnym, przez kolejne uproszczenia modelu aż do wielce uproszczonego opisu, po zwalającego na dyskutowanie stanów psychologicznych i zachowania się systemu. Zamiast więc zadawać pytania: czy umysł jest komputerem lub maszyną Turinga (oczywiście, że nie jest), zadajmy pytanie: jakiego rodzaju systemy mają umysły? Odpowiedź jest prosta: mózgi są przyczyną umysłów i w idać oczywistą zależność złożoności form umysłu (wyrażających się w postaci złożonych form zachowań) od złożoności mózgu. Mózg jest nie tylko konieczny do wytworzenia umysłu, lecz musi jeszcze działać w odpowiedni sposób. Rodzą się tu dwa istotne pytania. Czy możemy zrozumieć jego działanie za pomocą modeli na tyle prostych, by otrzymać w pełni zadawalającą teorię umysłu? Jeśli wyobrazimy sobie ciąg coraz bardziej wyrafinowanych modeli mózgu, czy w granicy otrzymamy system posiadający istotne cechy naszego umysłu?

Wydaje mi się, że na obydwa pytania należy udzielić odpowiedzi twierdzącej. Proste modele, oparte na zrozumieniu funkcji różnych “podukładów” mózgu, które dają się zdefiniować funkcjonalnie jedynie w przybliżony sposób, pozwalają na zadawalające zrozumienie wiek szości syndromów neuropsychologicznych, a nawet psychiatrycznych (por. Ruppin 1995). Nawet najprostsze modele koneksjonistyczne pamięci wykazują szereg cech znanych z psychologii pamięci, np. adresowalność kontekstową, rozpoznawanie uszkodzonych wzorców czy niezależność czasu odpowiedzi od liczby nauczonych faktów. Takie modele pozwalają zrozumieć naturę reprezentacji wewnętrznych a nawet pochodzenie halucynacji, powstających w wyniku uszkodzeń sieci neuronowej, powodujących rozbicie reprezentacji we wnętrznych na fragmenty, które mogą być pobudzane tworząc bezsensowne konfiguracje. Dobra teoria umysłu powinna odpowiadać na takie pytania jak: dlaczego możliwa jest dysleksja bez dysgrafii? Jest to wynikiem specyficznych uszkodzeń zakrętu kątowego (por. Górska i inni 1997). Na czym polega syndrom Capgrasa, w którym pacjent nabiera przeko nania, że ktoś z jego rodziny (a nawet on sam), wyglądający i zachowujący się z pozoru normalnie, został opanowany przez demona lub pozaziemską istotę? Mamy tu do czynienia z dysocjacją rozpoznawania kognitywnego (w oparciu o procesy w korze mózgu) i afektywnego

(w oparciu o procesy w układzie limbicznym). Chociaż szczegóły nie są do końca znane zrozumienie takich zjawisk nie wiąże się z trudnościami fundamentalnymi, a jedynie technicznymi.

Nie ma wątpliwości, że coraz dokładniejsze modele komputerowe, tworzone w oparciu o dane neurofizjologiczne, pozwolą na coraz bardziej szczegółowe wyjaśnienia wielu aspek tów działania umysłu. Subiektywność i intencjonalność nie wydaje się być szczególnym problemem. Każdy system samoorganizujący się, np. taki jak steruje robotem Darwin (por. Edelman 1999), wykazuje zachowania intencjonalne, wynikające z realizacji ogólnych wartości i potrzeb systemu, będących rezultatem jego budowy biologicznej lub konstrukcji technicznej. Do takich potrzeb należy nie tylko pragnienie czy jedzenie, lecz również po trzeba doznań, powodująca na drodze oddziaływania ze środowiskiem powstanie wyrafinowanych form zachowań, które nie są wynikiem programowania przez twórcę systemu. Jego działanie można rozpatrywać zarówno z wewnętrznego punktu widzenia pierwszej osoby jak i z zewnętrznego punktu widzenia osoby trzeciej. W granicy rozwoju tego typu modeli, dokonujących wzorowanego na działaniu mózgu przetwarzania informacji, spo dziewać się można coraz bardziej złożonych form zachowań. Nie widać fundamentalnych powodów, dla których nie mielibyśmy przypisać robotom tego rodzaju jakiejś formy umysłu. Jest to w znacznej mierze kwestia zdefiniowania minimalnych warunków, jakie powi nien spełniać dany system, by można było mówić o jego umyśle. Fakt, że zamiast procesów analogowych mamy tu do czynienia z procesami cyfrowymi nie wydaje mi się tu znaczący (por. dyskusję argumentu chińskiego pokoju dokonaną powyżej).

Jedynym poważnym problemem pozostaje kwestia świadomości. Czy w ciągu coraz bardziej doskonałych modeli mózgu pojawią się również zachowania świadome, czy też po trzebujemy do tego materii biologicznej, dysponującej odpowiednią “mocą przyczynową” (Searl 1999)? Czy systemy takie mogą mieć wrażenia (qualia)? Wiele napisano o nieredukowalności świadomości, lecz zamiast wdawać się w dyskusje tego rodzaju spróbujmy po patrzeć na to zagadnienie od innej strony.

W naszych mózgach zachodzą realne procesy neurofizjologiczne, będące przyczyną powstawania wrażeń świadomych. Pozwolę sobie na kilka spekulacji dotyczących sposobu zachodzenia tego procesu. Nie jest przy tym istotne, na ile szczegóły tych spekulacji będą zgodne z przyszłymi teoriami mózgu, chodzi jedynie o charakter całego rozumowania. Wiele procesów rozwiązywanych jest przez dobrze zlokalizowane obszary mózgu, obserwowalne za pomocą tomografii komputerowej lub innych metod tego typu. Należą do nich zagadnienia analizy danych zmysłowych, takie jak segmentacja obrazu, dzięki czemu wi dzimy przedmioty, a nie barwne plamy. Jednakże procesy wymagające współdziałania kilku modalności zmysłowych, lub odwołania się do pamięci epizodycznej, która nie jest zlokalizowana w konkretnym obszarze kory mózgu i zawiera aspekty określonej modalności, nie mogą być realizowane lokalnie. Konieczny jest do tego mechanizm rozprzestrzeniania in formacji i łączenia wyników otrzymanych z poszczególnych obszarów mózgu w jedną całość (por. Newman i Baars 1993). Mamy więc do czynienia z fragmentami reprezentacji mentalnych dostarczonymi przez zmysły lub wewnętrzne pobudzenia obszarów przetwa rzających informację zmysłową, oraz z całościową reprezentacją, zawierającą aktualnie pobudzone fragmenty. Można o niej myśleć jako o meta-reprezentacji, realizowanej prawdo podobnie przez globalną dynamikę bioelektrycznych wyładowań mózgu, obserwowalną w postaci fal EEG.

Postawmy teraz następującą hipotezę: treści umysłu, wrażenia świadome, zawarte są w dynamice globalnej bioelektrycznych stanów mózgu. Nie potrafimy jeszcze analizować aktyw ności bioelektrycznej mózgu dostatecznie dokładnie by to

zobaczyć (w szczególności elektrody EEG obserwują tyko aktywność powierzchniową, podczas gdy konieczne są również informacje o aktywności ośrodków podkorowych). Jest jednak dość prawdopodobne, że tak jest w istocie, w każdym razie jest to hipoteza weryfikowalna i ma daleko idące konse kwencje. Treść umysłu można więc uważać za pewien komentarz do stanów mózgu, przy czym wkład do niej mają procesy na różnych poziomach przetwarzania informacji. W systemie nie ma żadnego homunkulusa, który "widzi" obrazy korzystając z wyjść najwyższych pięter układu wzrokowego. Pojawienie się informacji w mózgu wystarczy do podjęcia działania, takiego jak skomentowanie werbalne tej informacji lub jej udostępnienie innym wyspecjalizowanym obszarom mózgu. Pojemność informacyjna tego dynamicznego systemu jest niewielka, a czas trwania pojawiającej się w nim informacji ograniczony (por. symulacje komputerowe pamięci krótkotrwałej pokazujące stabilność około 7 reprezentacji, Ingber 1995). Z drugiej strony liczba zachodzących w mózgu procesów jest ogromna, ale tylko te najbardziej aktywne mają znaczący wkład do globalnej dynamiki. Używam tu pojęcia przetwarzania informacji w ramach modelu, nie przypisując przetwarzania informacji mó zgowi, który, podobnie jak każdy układ analogowy, nie przetwarza informacji w większym stopniu niż prąd w odbiorniku radiowym rozwiązuje równania Ohma i Kirchoffa, por. Searle 1990).

Rozważmy teraz typowe zadanie, przed jakim stoi ssak, np. szczur, odżywiający się różnorodnym pokarmem. Smak pożywienia nie jest wielkością dyskretną, nie można go opisać symbolicznie, jest natomiast zbiorem pobudzeń czopków smakowych. Smak ten należy po równać z zapamiętanymi smakami by określić, czy jest to pożywienie bezpieczne. Reprezentacja wrażenia smaku musi więc zostać rozesłana do wszystkich obszarów mózgu, które zawierać mogą istotne informacje, po czym musi zostać utrzymana w pamięci krótkotrwałej dostatecznie długo, by rozejść się po całym mózgu i by nastąpiły w nim procesy asocjacji. Mamy więc do czynienia z ciągłą, niewerbalną aktualizacją wrażeń smakowych w pamięci krótkotrwałej. Ciągła aktualizacja konieczna jest ze względu na szybki zanik informacji w pamięci krótkotrwałej. Jeśli nie ma żadnych negatywnych skojarzeń szczur zacznie jeść, jeśli jednak smak (lub zapach) skojarzy mu się negatywnie kolejnym stanem jego mózgu będzie pobudzenie ośrodków strachu i porzucenie jedzenia. Najważniejszym procesem jest więc dyskryminacja wrażeń wpływających na zachowanie. Jeśli skojarzenia są pozytywne szczur rozpoznaje ulubione pożywienie jako nagrodę – pobudzeniu ulegają ośrodki przyjemności i informacja o tym pojawia się w treści jego umysłu.

Szczur wie, a jego wiedza ma charakter wr ażeń wynikających z działania mechanizmów poznawczych. Gdyby jego mózg posiadał zdolność werbalnego komentowania stanów umy słu, tak jak posiada ją mózg ludzki, z pewnością stwierdziłby, że odczuwa smak, widzi przedmioty i ma z tym skojarzone wrażenia, gdyż są one realną reprezentacją fizycznych stanów mózgu. Moje wrażenie widzenia mają konkretną jakość, nie są jedynie "dyspozycjami" by twierdzić, że mam wrażenia (por. Dennet 1991), lecz realnymi stanami mózgu. Twierdzę, że każdy model działający na podobnej zasadzie, jeśli będzie dostatecznie złożony, by móc komentować pojawienie się takich wrażeń, będzie twierdził, że jest świadomy swoich wrażenia, a struktura tych wrażeń może być dowolnie podobna do struktury naszych wrażeń. W granicy modele tego typu musiałyby wytworzyć pewną reprezentację sie bie, odgraniczającą "ja" od "nie-ja", powalającą odróżnić "siebie" od reszty świata. Każdy system podlegający ewolucji musiał wytworzyć taką reprezentację, chociaż czasami identy fikacja gatunkowa jest ważniejsza od osobniczej. Nasze poczucie tożsamości wywodzi się prawdopodobnie z proprocepcji a rozwój mózgu związany był przede wszystkim z koniecznością powstania wewnętrznego modelu ciała w celu przewidywania wyników ruchu

kończyn (por. Cotterill 1996).

Świadomość w tym ujęciu nie jest więc niczym innym jak zdolnością do dyskryminacji ciągłych reprezentacji stanów mózgu, tworzących pewną "przestrzeń wewnętrzną" dla su biektywności. Wszelkie zaburzenia świadomości muszą być związane z upośledzeniami funkcji kognitywnych. Ślepota korowa (wynikająca z uszkodzenia kory wzrokowej) prowadzi do upośledzenia funkcji wzrokowych, pomimo dostępności informacji z nerwu wzro kowego w wielu strukturach mózgu (por. Milner, Goodale 1995). Informacja ta ma jednak całkiem inny wkład do globalnej dynamiki mózgu i stąd dyskryminacja tych stanów nie prowadzi do wrażeń natury wzrokowej, chociaż związana jest z licznymi, trudnymi do określenia wrażeniami – interpretacji tych wrażeń trzeba się powoli nauczyć, ale poziom kompetencji wzrokowej nig dy nie będzie zbyt wysoki, gdyż do precyzyjnej analizy sygna łów z nerwu wzrokowego brakuje wyspecjalizowanych struktur.

W literaturze filozoficznej często przywołuje się logiczną możliwość istnienia *zombi*, pozbawionego świadomości, lecz z pozoru normalnie zachowującego się człowieka. Do pew nego stopnia osiągnięcie takiego stanu jest możliwe po zastosowaniu wyciągu z pewnych roślin, używanego w Haitańskim kulcie voodu. Wade Davis, antropolog z Harvardu, opisał (Davis 1988) stosowane w obrzędach voodu substancje powodujące pozorną śmierć. W tym stanie ludzie zostają pochowani żywcem, a po paru dniach odkopuje się ich i podaje środki halucynogenne, powodujące amnezję i paraliżujące ich wolę. W wyniku niedotlenienia może nastąpić trwałe uszkodzenie niektórych funkcji mózgu. Po takich zabiegach trudno się jednak spodziewać, by normalne funkcje mózgu *zombi* nie uległy upośledzeniu. W szczególności złożone funkcje poznawcze, związane z planowaniem czy twórczym myśleniem, wymagają zdolności do kojarzenia na meta-poziomie integrującym informacje z wszystkich obszarów mózgu, do których *zombi* nie będzie zdolny. Uzasadnionym wydaje się twierdzenie, że takie funkcje wymagają istnienia świadomości, i że nie istnieją *zombi* zdolne do wykonywania takich czynności przy jednoczesnym braku świadomości.

Sugestie pohipnotyczne wpływać mogą na globalną dynamikę mózgu powodując po podaniu skojarzonego bodźca (np. słowa lub dźwięku) dziwne zachowania, które dana osoba usiłuje uzasadnić uciekając się do konfabulacji. Z pozoru jest to bardzo niezrozumiałe zja wisko, jednakże mechanizm jest prawdopodobnie podobny do tego, jaki powoduje, że o określonej godzinie czy określonym miejscu pojawia się w naszym umyśle potrzeba załatwienia jakiejś sprawy. W jaki sposób, nie myśląc o tym przez cały dzień, przypomniałem sobie o ważnym telefonie o 6-tej godzinie? Pewne grupy neuronów w moim mózgu musiały realizować "proces w tle", uaktywniając się o określonej godzinie na tyle, by zapamiętana przez nie informacja pojawiła się jako treść umysłu. Sugestia hipnotyczna prawdopodobnie uruchamia również podobne pro cesy, skojarzone z różnymi bodźcami słownymi lub innymi.

6. Konkluzje

John Searl w znakomitej książce "Umysł na nowo odkryty" (1999) ostro skrytykował klasyczny kognitywizm, traktujący umysł jako pewnego rodzaju program realizowany przez mózg. Jest to podejście naiwne, mylące model z rzeczywistością. "Mózg wytwarza stany świadome [...] i na tym koniec. Jeśli chodzi o umysł wszystko już zostało powiedziane" (Searl 1999, s. 300). Nie ma według niego żadnych działań zgodnych z regułami, przetwa rzania informacji, nieświadomych wnioskowań i innych, występujących w kognitywistycznych modelach koncepcji. Nie wydaje mi się to dobrą podstawą do zbudowania teorii dzia łania umysłu i w istocie Searl nie daje żadnej odpowiedzi na pytanie "jak jest możliwe, by istniała jakakolwiek teoria umy słu?".

Koncepcje kognitywistów są na tyle przydatne, na ile są dobrym przybliżeniem do modeli opisujących działanie mózgu na poziomie neurofizjologicznym. Jeśli dostrzegamy w tym działaniu pewne regularności próbujemy je ująć w formie reguł logicznego postępowania. Jak słusznie wskazuje Searl nie można jednak tych reguł traktować jako intencjonalnych reguł mentalnego postępowania. Nie zmienia to jednak faktu, że wygodnie jest w pewnych uproszczonych modelach posługiwać się pojęciami klasycznej kognitywistyki, w tym lo gicznymi regułami do tłumaczenia stanów umysłu. Jak już wspomniałem najbardziej zadawalającym uzasadnieniem i zakreśleniem obszaru stosowalności takich koncepcji byłoby stopniowe upraszczanie modeli biofizycznych (Duch 1997, gdzie modele na różnych po ziomach i zagadnienia redukcji opisane są dokładniej), począwszy od poziomu genetyki i budowy molekularnej komórek, zjawisk neurochemicznych, działania neuronów, grup neuronów, jąder neuronowych i innych anatomicznie odróżnialnych obszarów mózgu, aż do globalnej dynamiki zjawisk bioelektrycznych, która wydaje się korelować ze stanami umysłu (Tononi, Edelman 1998).

Zrozumienie relacji mózg-umysł wymaga stosowania uproszczonych modeli, nie możemy się tu zadowolić ogólnym stwierdzeniem “mózg wytwarza świadomość”. Zadanie to nie przekracza jednak naszych możliwości i coraz lepiej rozumiemy wiele szczegółów dotyczących działania mózgu. Potrafimy symulować działanie pojedynczych neuronów z dużą dokładnością i nic nie wskazuje na to, by miały one jakieś szczególne “moce przyczynowe”. Świadomość nie jest wynikiem pobudzeń neuronów, lecz procesów poznawczych związa nych z dyskryminacją ciągłych reprezentacji wewnętrznych. Szczególny status przypisywany wrażeniom (qualia) jest iluzją. Nie widzę fundamentalnych powodów, dla których nie dałoby się skonstruować takich wzorowanych na mózgu systemów sztucznych, które będą twierdzić, że są świadome, ani też powodów, dla których mielibyśmy takie twierdzenia odrzucić. Czy ma to istotne znaczenie? Wydaje mi się, że ważniejsze będą pytania natury etycznej. System o stopniu komplikacji dorównującym ludzkiemu mózgowi, który rozwinie swoje wyobrażenia o świecie oddziaływując z jakąś grupą ludzi, będzie dla nich cennym partnerem, stanowiąc istotną wartość niezależnie od naszych przekonań dotyczących zdol ności systemów sztucznych do posiadania świadomych wrażeń. Inflacja wartości wynikać może jedynie z łatwości powielania lub odtwarzania stanu sztucznego umysłu.

Ostanie dyskusje dotyczące natury umysłu pokazują (por. Dennett 1991, Chalmers 1997, Searle 1990), jak dziwne idee wydają się różnym ludziom zadowalające (lub przynajmniej akceptowalne). Ponieważ większość z tych pomysłów nie prowadzi do sensownych odpowiedzi na konkretne pytania, stawiane przez nauki poznawcze, nie mogą one być podstawą do dobrych teorii umysłu. Przedstawiłem tu argumenty podważające powszechne przekona nie o niezdolności modeli opartych na symulacjach komputerowych do osiągnięcia prawdziwego “zrozumienia” czy świadomości. Symulowany umysł długo jeszcze nie będzie identyczny z biologicznym ze względu na trudności techniczne tego typu symulacji, ale w granicy takie modele mogą być przekonane, że mają świadome wrażenia, wynikające z in terpretacji globalnego stanu swoich mózgów. Granice dokładności symulacji takich modeli za pomocą zwykłych komputerów nie są jeszcze jasne. Przedstawione tu naturalistyczne rozwiązanie prowadzi do licznych przewidywań empirycznych (Duch 1999, w przygotowaniu) i może stać się podstawą do stworzenia zadawalającej teorii umysłu.

Literatura

1. Black Ira B, *Information in the Brain. A Molecular Perspective* (A Bradford Book, 1994)
2. Chalmers D.J, *Facing up to the problem of consciousness*. J. of Consciousness Studies

2 (1995) 200-219
3. Chalmers D.J, *The Conscious Mind: In Search of a Fundamental Theory* (Oxford University Press 1996)
4. Chalmers D.J, *Moving forward on the problem of consciousness*. J. of Consciousness Studies 4 (1997) 3-46
5. Cohen M.M, Massaro D.W, *On the similarity of categorization models*, In: F.G. Ashby, ed. *Multidimensional models of perception and cognition* (LEA, Hillsdale, NJ 1992), rozdz. 15.
6. Cotterill R, *Prediction of internal feedback in conscious perception*. J. of Consciousness Studies 3 (1996) 245-266
7. Crick F, *Zdumiewająca hipoteza czyli nauka w poszukiwaniu duszy* (Prószyński i S-ka, Warszawa 1997)
8. Duch W, *Platonic model of mind as an approximation to neurodynamics*. In: *Brain-like computing and intelligent information systems*, ed. S-i. Amari, N. Kasabov (Springer, Singapore 1997), chap. 20, pp. 491-512
9. Damasio A.R, *Błąd Kartezjusza* (Rebis 1999, seria "Nowe horyzonty")
10. Davis W, *Zombification*, Science 240 (1988) 1715-1716
11. Dennett, D. C, *Consciousness explained* (Little-Brown 1991)
12. Edelman G. M, *Przenikliwe powietrze, jasny ogień. O materii umysłu* SIZE="3"> (PIW, Warszawa 1999)
13. Freeman W. J, *Societies of Brains: A study in the neuroscience of love and hate* (Lawrence Erlbaum Associates 1996)
14. Górska T, Grabowska A, Zagrodzka J. (red.) *Mózg a zachowanie* (Wydawnictwo Naukowe PWN, Warszawa 1997)
15. Gardner M, *Computers near the threshold*. J. of Consciousness Studies 3 (1996) 89-94
16. Gregory R, *Mind in Science: A History of Explanations in Psychology and Physics*. (Penguin Books 1981)
17. Horgan J, *Koniec nauki, czyli o granicach wiedzy u schyłku ery naukowej* (Prószyński i S-ka, Warszawa 1999)
18. Ingber L, *Statistical mechanics of multiple scales of neocortical interactions*. In: *Neocortical dynamics and Human EEG Rhythms*, ed. Nunez P.L (Oxford University Press 1995), p. 628-681
19. Jackendoff R, *Consciousness and the computational mind* (MIT Press, Cambridge, MA, 1987)
20. Kloch J, *Świadomość komputerów?* (Ośrodek Badań Interdyscyplinarnych, Kraków, i BIBLOS, Tarnów 1996)
21. Lewis C.S, *Odrzucony obraz* (Kraków 1995)
22. McLean P. D, *A triune concept of the brain and behavior* (University of Toronto Press, Toronto 1973); *The Triune Brain in Evolution* (University of Toronto Press, Toronto 1990)
23. McGinn, C, *Consciousness and Space*. J. of Consciousness Studies 2 (1995) 220-230
24. Milner A.D., M. Goodale, *The visual brain in action* (Oxford University Press, 1995)
25. Minsky M, *The Society of Mind* (Simon and Schuster, Nowy Jork 1985)
26. Nevell A, *Unified theories of cognition* (Harvard University Press, Cambridge, Massachusetts, 1990)
27. Newman J, Baars B.J, *Neural Global Workspace Model*, Concepts in Neuroscience 4 (1993) 255-290
28. Penrose R, *Shadows of the mind* (Oxford University Press 1994)
29. Popper K, Eccles J.C, *The Self and its Brain* (Springer Verlag, Berlin 1977)
30. Putnam, H, *The meaning of 'meaning'*. Minnesota Studies in the Philosophy of Science 7 (1975) 131-193
31. Putnam H, *Mind, language and reality* (Cambridge University Press, 1975)
32. Putnam, H, *Representation and Reality* (MIT Press 1987)
33. Putnam H, dodatek do *The Royce Lectures in the Philosophy of Mind*, wykładów wygłoszonych na Brown University, 1998 (dziękuję autorowi za przesłanie mi manuskryptu).
34. Ruppin E, *Neural modeling of psychiatric disorders*, Network 6 (1995) 635-656
35. Searle J, *Świadomość, inwersja wyjaśnień i nauki kognitywne.* Brain and Behavioral Sciences 13 (1990) 585-642. Przedruk w "*Modele umysłu*" red. Z. Chlewiński (Wyd. Naukowe PWN, Warszawa 1999)
36. Searle J, *Umysł na nowo odkryty* (PIW, Warszawa 1999).
37. Tononi G, Edelman G. M, *Consciousness and Complexity*, Science 282 (1998) 1846-1851.