
Neurokognitywna teoria świadomości

John Searle w artykule o przyszłości filozofii, napisanym dla milenijnego numeru *Philosophical Transactions of Royal Society* przypomina, iż wiele problemów uważanych za filozoficzne udało się sprowadzić na grunt nauk szczegółowych (Searle 1999). Debaty na temat różnic pomiędzy materią „bezwładną” a „ożywioną” na początku XX wieku toczono z wielką intensywnością, lecz dopiero rozwój genetyki i biologii molekularnej zmienił zupełnie meritum tej dyskusji. Podobnie obecne debaty dotyczące filozofii umysłu mogą po latach odejść w niepamięć dzięki osiągnięciom nauk kognitywnych, a w szczególności neurobiologii i nauk pokrewnych. Za największą przeszkodę w znalezieniu poprawnego rozwiązania problemu psychofizycznego Searle uznał „zbiór tradycyjnych, lecz przestarzałych, pojęć takich jak ciało i umysł, materia i duch, mentalne i fizyczne”. Zamiast martwić się, w jaki sposób mózg może stworzyć umysł należy według niego przyjąć, że tak po prostu jest. Umysły, kwarki, polityka czy kultura są w różny sposób częścią jednego świata.

Spróbuję tu pokazać, że umysły w istocie należą do tego świata, chociaż tworzą w nim świat autonomiczny, nieredukowalny do samych zjawisk fizycznych. Użyję przy tym argumentu, który wydaje mi się dość oryginalny. Przedstawię mianowicie prosty schemat mózgowopodobnego układu – nazwijmy go artilektem – którego pobudzenia wykazywać będą podobne relacje wzajemne, jakie wykazują stany umysłu. Sama reprezentacja czy stany wewnętrzne nie są tu istotne, liczą się jedynie relacje wzajemne. Artilekt, interpretując w symboliczny sposób swoje stany dynamiczne, będzie musiał twierdzić, że jest świadomy i ma wrażenia. Nasze mózgi są przykładem realizacji układu o takiej strukturze. Pozwala to spojrzeć na zagadnienie świadomości z nowej strony i prowadzi do wielu wniosków dotyczących natury wrażeń świadomych. Najpierw jednak przedstawię bardzo krótko trudny problem świadomości i dotychczasowe próby jego rozwiązania.

1. Trudny problem świadomości

Świadomość stała się w ostatnim dziesięcioleciu modnym zagadnieniem (por. Chalmers 1995, 1996). Ponieważ jest to pojęcie często stosowane w języku potocznym, używane w bardzo wielu kontekstach, wywołujące wiele skojarzeń, nie da się go ująć w ramach teorii naukowej, która pokryłaby wszystkie jego znaczenia. W najprostszym sensie problem świadomości sprowadza się do znalezienia neurofizjologicznych korelatów odróżniających sytuacje, w których jesteśmy jakiegoś bodźca świadomi, od sytuacji, w których nie jest on uświadamiany. Są realne szanse na znalezienie takich korelatów dzięki analizie czynności bioelektrycznej i biomagnetycznej mózgu (por. Sporns, Tononi, Edelman 2000). Globalna dynamika mózgu jest aktualizacją, w danej chwili czasu, niektórych śladów pamięci, skojarzeń przez nie wywołanych, stanu obszarów analizujących dane zmysłowe i obszarów kontrolujących ruch. Następujące po sobie stany charakteryzują się pewną czasoprzestrzenną strukturą, która zapewne odpowiada stanom pamięci roboczej, strumieniowi świadomych wrażeń, czyli treści umysłu. Nie mamy jeszcze na to dowodów, jednakże taka możliwość jest wielce prawdopodobna i nie budzi większych kontrowersji.

Z filozoficznego punktu widzenia najtrudniejszym zagadnieniem jest problem subiektywnych wrażeń lub ściślej rzecz biorąc problem specyficznych, subiektywnych jakości, jakie wiążą się z wrażeniami. Jest to stary problem przedstawiony w nowym ujęciu przez Davida Chalmersa (1995,96) i szczegółowo omawiany na łamach „Journal of Consciousness Studies”. Łacińskie słówko *qualis* oznacza własność w oderwaniu od rzeczy, posiadających daną własność, stąd mówi się potocznie o „problemie qualiów”. Większość zagadnień dotyczących działania umysłu sprowadza się do wskazania na specyficzne funkcje mózgu, które są za nie odpowiedzialne. Możemy wyjaśnić przyczyny powstawania halucynacji pod wpływem środków halucynogennych wskazując na załamanie się mechanizmów pamięci skojarzeniowej w wyniku zaburzeń procesów neurochemicznych, koniecznych do jego prawidłowego działania. Nie pomoże to nam jednak zrozumieć, jakiego rodzaju wrażenia ma osoba doznająca takich halucynacji. Do zrozumienia odczuć związanych z wrażeniami nie wystarczy wyjaśnić funkcji odpowiedzialnych za ich powstanie. Zrozumienie sposobu integracji wrażeń o różnych modalnościach w jedną całość, dostępności informacji pozwalającej na odpowiednie zachowania, kwestia tworzenia raportów na temat stanu wewnętrznego, rozwijania się dialogu wewnętrznego i tym podobne problemy wyjaśnić można znając mechanizmy działania mózgu, a więc w sposób redukcjonistyczny. Wrażenia subiektywne nie wydają się sprowadzać do funkcji. Wrażenie barwy nie daje się rozłożyć na czynniki prostsze, stąd wyjaśnienia redukcjonistyczne nie są możliwe.

Jest to zagadnienie ważne, gdyż bez jego zrozumienia nie potrafimy odpowiedzieć na pytanie „czy komputer lub inny sztuczny system może mieć wrażenia?” Zaprogramować można jedynie „różnice, które prowadzą do różnych stanów”, a więc stany informacyjne. Istnienie wrażeń nie jest obserwowalną różnicą – tak się przynajmniej twierdzi. Przetwarzanie informacji, przynajmniej w niektórych warunkach, może zachodzić bez wrażeń wzrokowych (np. ślepotą korową, por. Dennett 1991). Z drugiej strony wrażenia wzrokowe mogą pojawiać się pomimo utraty wzroku, jeśli ta utrata związana jest z uszkodzeniami na poziomie pierwotnej kory wzrokowej lub szlaków do niej dochodzących (Sacks 1999). Często dyskutowana jest logiczna możliwość przetwarzania informacji „w ciemności” i istnienia zombi, a więc istot, które zachowując się zewnętrznie podobnie jak istoty świadome nie mają żadnych wrażeń związanych z przetwarzaniem informacji (por. Block 1995). Trudno jest jednak wyprowadzić takie wnioski z przypadków ślepoty korowej. Przypadki te pokazują raczej, że uszkodzenia układu wzrokowego bardzo mocno redukują możliwości posługiwania się docierającymi do mózgu sygnałami ze strony siatkówki. Wszelkie uszkodzenia mózgu prowadzą do obniżenia kompetencji behawioralnej. Możliwość działania pomimo braku wrażeń wzrokowych nie oznacza braku wszelkich wrażeń, a jedynie nieumiejętność interpretacji tych „niejasnych” wrażeń, które się pojawiają. Po pewnym czasie mózg uczy się je prawidłowo interpretować (tzn. wykorzystywać do eksploracji przestrzeni i do działania, por. O'Regan i Nöe 2001), co znajduje wyraz w nowych wrażeniach świadomych, odmiennych od zwykłych wrażeń wzrokowych. Ślepotą korową jest argumentem za związkiem pomiędzy właściwą interpretacją stanów mózgu i powstawaniem świadomych wrażeń – do zagadnienia tego jeszcze powrócę.

W dyskusji po artykule Chalmersa formułującym trudny problem świadomości, czyli problem zrozumienia jakości wrażeń świadomych, zaproponowano kilka rozwiązań, o których warto wspomnieć.

1. Czy jest tu jakiś realny problem?

Daniel Dennett (1996) i kilku innych filozofów uznało ten problem, podobnie jak problem wyjaśnienia istoty życia, czy pojęcia takie jak „fajność”, za problem sztuczny. Czy jest sens dyskutować o wyjaśnieniu istoty „fajności” poza opisem, co uznać można za fajne, a co nie? Według niego wrażenia subiektywne są kwestią mentalnych dyspozycji. Ta odpowiedź wskazuje na trudności w dostrzeżeniu sedna „trudnego” problemu. Próba wyobrażenia sobie, jak można by zaprogramować wrażenia w sztucznym systemie pokazuje dobitnie, że nie jest to tylko kwestia dyspozycji systemu do produkowania pisku po naciśnięciu klawisza. Bardzo łatwo jest takie dyspozycje zaprogramować, trudniej za to uzasadnić, że piszcząca maszyna ma podobne wrażenia, co piszczące zwierzę. W XVIII wieku mechanicyści podjęli próbę zrozumienia działania organizmów i redukcji życia psychicznego do mechanicznych związków elementów, ale koncepcje te niewiele wyjaśniły. Próba sprawdzenia działania umysłu do przetwarzania informacji daje znacznie większe możliwości modelowania stanów wewnętrznych, ale załamuje się na problemie zrozumienia jakości wrażeń. Życie czy fajność to koncepcje abstrakcyjne, wrażenia są przeżywane w bezpośredni sposób. Czy sztuczne systemy mogą mieć wrażenia? W miarę budowania coraz bardziej złożonych systemów pytanie to stanie się bardzo ważne.

2. Rozwiązanie nie-redukcyjne.

Chalmers powrócił do starego pomysłu, który można wywieść jeszcze od Spinozy (1677). Skoro przetwarzanie informacji przez mózgi wywołuje w tych mózgach wrażenia, to może istnieje podwójny aspekt informacji – fizyczny i fenomenalny? Otrzymamy w ten sposób naturalistyczny dualizm, który dużo łatwiej zaakceptować niż tradycyjny dualizm Kartezjański. Ten pomysł wydaje się mieć wielu zwolenników, ale w istocie nie jest to rozwiązanie problemu, a raczej rezygnacja z próby rozwiązania: doświadczenia subiektywne istnieją, bo tak jest, wynika to z fundamentalnego prawa natury. Nadal jednak nie wiemy nie tylko dlaczego tak jest, ale nie potrafimy rozstrzygnąć, jakie systemy będą miały wrażenia, a jakie nie, dlaczego pewne bodźce przetwarzane przez mózg wywołują w nas wrażenia, a na inne reagujemy automatycznie, bez wrażeń świadomych, dlaczego w pewnych stanach mózgu jesteśmy świadomi, a w innych (np. podczas anestezji) nie? Pomysł ten wydaje się bezpłodny, nie pozwala zrozumieć specyfiki wrażeń o różnych modalnościach, odpowiedzieć na pytania „dlaczego wrażenia są takie a nie inne”?

W istocie nie jest to lepsze rozwiązanie niż jawna rezygnacja z prób zrozumienia umysłu, które mają przekraczać nasze zdolności intelektualne (McGinn 1990). Być może ludzki umysł nie może zgłębić wszystkich tajemnic natury, ale nie oznacza to, że należy szybko zrezygnować z poszukiwań dobrego rozwiązania, zwłaszcza w obliczu postępów neurobiologii w ostatnich latach.

3. Panpsychizm.

Seager (1995) i Rosenberg (1996) opowiadają się za słabą wersją panpsychizmu, który można określić jako panprotopsychizm. Materia ma pewną własność X, która dopiero w połączeniu ze złożoną materią tworzy świadomość. Jeśli za złożoną materię uznać taką, która przetwarza informację, jest to postawa bliska podwójnemu aspektowi informacji Chalmersa. Świadomość nie pojawia się we wszystkich złożonych systemach, a tylko w takich, które przetwarzają informację w specyficzny sposób. Co odróżnia superszybkie, niezwykle złożone komputery, które nie mają świadomości, od mózgów, które ją mają? Na czym ma polegać ta złożoność, konieczna do powstania świadomości? Jakiego rodzaju

własność jest potrzebna do powstania świadomości i czy można ją jakoś dodawać bądź usuwać ze złożonego systemu? Przypomina to jakąś duszę, której obecność miała być konieczna do ożywienia materii. Jest tu więcej pytań, niż odpowiedzi a sam pomysł nigdzie nie prowadzi.

4. Protofenomeny.

Podstawowe elementy świadomości, które można by nazwać „protofenomenami”, w złożonych systemach wywołują doświadczenia świadome. Jak jednak te protofenomeny wiążą procesy fizyczne z umysłowymi? Czym są i jak wiążą się z neuronami? Przypomina to próbę „kwantyzacji pola świadomości”. J. Eccles (1985) sądził, że niematerialny umysł steruje mózgiem zmieniając prawdopodobieństwa transmisji synaptycznych za pomocą hipotetycznych *psychonów*, pełniących rolę kwantów pola świadomości. Nie ma jednak żadnych poszlak świadczących o tym, że coś takiego istnieje i jest do czegoś potrzebne. Wręcz przeciwnie, im dokładniej potrafimy modelować korę i inne struktury mózgu za pomocą sztucznych sieci neuronowych, tym więcej funkcji typowych dla umysłu udaje się w ich zachowaniu otrzymać. Czy należy się spodziewać na pewnym etapie doskonalenia modeli mózgu pojawienia się trudności z powodu braku psychonów lub innych tajemnych elementów w komputerach? Widać tu znowu nadzieję na konieczność wprowadzenia (w nieco zmechanicyzowanej formie) jakiegoś ducha, który ożywi materię. Istnienie protofenomenów lub psychonów wydaje się wielce nieprawdopodobne.

5. Świadomość kwantowa.

Od ponad 40 lat (Wigner 1962) próbuje się połączyć problemy związane z interpretacją podstaw mechaniki kwantowej, w szczególności tzw. problem obserwatora i kolapsu funkcji falowej, z zagadnieniem świadomości. Świat kwantowy opisany jest przez funkcję falową, opisującą wszystkie potencjalnie możliwe stany, które dopiero w wyniku pomiaru przyjmują konkretne wartości. Czy potrzebna jest do tego świadomość obserwatora czy wystarczy nieodwracalny zapis wyniku pomiaru? Jest wiele alternatywnych teorii pomiaru w mechanice kwantowej, które nie odwołują się do świadomości (Giulini i inni 1996). Sam Wigner, od którego pochodzi pomysł związku świadomości z problemem pomiaru (Wigner, 1962, str. 284-302), wycofał się z niego (Mehra i Wightman, 1995, str. 271) widząc trudności związane z istnieniem stanów kwantowych w mózgu. Jak wykazał ostatnio Tegmark (2000) efekty kwantowe mogą wystąpić w komórkach tylko w niesłyszalnie krótkich czasach, mniejszych niż jedna bilionowa część sekundy, podczas gdy na stan neuronu mają wpływ procesy trwające przynajmniej miliard razy dłużej. Ten kierunek myślenia nie przyniósł więc żadnych rezultatów.

Czy nielokalny model umysłu ma coś wspólnego z nielokalnością w mechanice kwantowej, jak sądzi Stapp (1993) czy Clarke (1995)? Czy do rozwiązania problemu świadomości potrzebna jest kwantowa teoria grawitacji, jak sądzi Penrose (1994)? Nie ma żadnych poszlak doświadczalnych popierających takie tezy. Efekty kwantowe w mózgu muszą ulec szybkiemu zniszczeniu z powodu jego wysokiej temperatury, a silne pola magnetyczne tomografów komputerowych zupełnie by je zmieniły, ludzie nie odczuwają jednak żadnych zaburzeń świadomości w takich warunkach. Rozwiązania kwantowe, chociaż upłynął już dość długi czas od ich zaproponowania, nie tłumaczą żadnego dającego się badać zjawiska związanego z świadomością.

6. Neurobiologiczna teoria znaczenia.

Francis Crick i Chris Koch (1995) podkreślają, że w przetwarzaniu informacji przez mózg tylko te bodźce, które mają dla organizmu istotne znaczenie, stają się świadome. Jest to zapewne stwierdzenie prawdziwe, ale z filozoficznego punktu widzenia mało przydatne. Trudno jest bowiem zdefiniować pojęcie znaczenia i zrozumieć, dlaczego znaczeniu miałyby towarzyszyć doświadczenia świadome. Na podobnej zasadzie, co przetwarzaniu informacji? Jeśli uznać umysł za system kontrolny na najwyższym szczeblu decyzyjnym, to oczywiście powinien on dostawać informacje przefiltrowane, tylko te, które potrzebne są do podejmowania decyzji, a więc posiadające znaczenie.

Nie wszystkie zachodzące w mózgu procesy wydają się być konieczne dla powstawania wrażeń świadomych. Sprzężenia z niektórymi grupami neuronów są na tyle powolne, że nie mogą one bezpośrednio wpływać na pojawianie się wrażeń świadomych, stąd pomysł Cricka i Kocha poszukiwania „neuronów świadomości”. Ponieważ wszystkie sprzężenia w mózgu są dość silne jest całkiem prawdopodobne, że nie da się go podzielić na rozłączne obszary, które można analizować niezależnie i przypisać im funkcje „realizacji świadomości”. Stany umysłu odpowiadają globalnym stanom dynamiki a obszary reagujące wolniej mają początkowo słaby, a potem narastający, wkład do powstających wrażeń.

7. Teoria Baarsa.

Bernard Baars, twórca psychologicznej kognitywnej teorii świadomości, proponuje uznać istnienie świadomości jako „brutalny fakt”, z jakimi często mają do czynienia psychologowie (1988, 1997). Globalnie dostępna informacja staje się w jakiś sposób świadoma. Zaletą tego podejścia jest możliwość odróżnienia, kiedy informacja może być uświadamiana (gdy jest globalnie dostępna), a kiedy nie. Decydują o tym procesy skojarzeniowe zachodzące w mózgu, w szczególności szybka ocena informacji przez ośrodki podkorowe, zaangażowane w powstawanie emocji. Jest to więc krok w dobrym kierunku. Jednakże wiele pytań pozostaje bez odpowiedzi. Czy sztuczne systemy mogą mieć wrażenia i jakie warunki muszą spełnić? W jaki sposób powstają wrażenia w oparciu o informację, która jest globalnie dostępna w systemie? W tym momencie Baars odwołuje się do „brutalnego faktu” rezygnując z próby dalszych wyjaśnień.

Żaden z tych punktów widzenia nie prowadzi więc do realnego postępu w zrozumieniu trudnego problemu świadomości.

2. Sztuczne umysły

Komputery i maszyny Turinga nie są właściwą metaforą dla działania umysłu. Architektura maszyny Turinga nie jest dobrą analogią opisującą działanie mózgu. Analogia oparta jest jedynie na przetwarzaniu informacji (komputery i mózgi przetwarzają informację), ale nie można jej rozciągać na inne aspekty działania umysłu, takie jak wrażenia. Chociaż rezultaty działania umysłu i programów komputerowych mogą być podobne wewnętrzne stany mózgu i sprzętu komputerowego różnią się w zasadniczy sposób.

W danej chwili czasu rejestry procesora komputera mają określone wartości binarne, decydujące o dalszym działaniu systemu. Do jego opisu niepotrzebne są relacje przestrzenne

czy czasowe, wystarczy abstrakcyjny opis typu maszyny Turinga. Tymczasem stan bioelektrycznej aktywności mózgu ma określony rozkład przestrzenny, rozciągłą, dynamiczną strukturę. Stanowi ona pewną całość, zawierającą zmieniający się w czasie wkład od wielu procesów peryferyjnych, skojarzenia z zapamiętanymi zdarzeniami, pobudzenia sensoryczne i wiele innych elementów. Nie można takiego dynamicznego stanu utożsamiać ze zmianą wartości binarnych rejestrów procesora tak jak nie można utożsamiać deszczu z komputerową symulacją procesów atmosferycznych. Działanie komputera nie przypomina działania mózgu i dlatego za pomocą komputerów można w dużym przybliżeniu oddać tylko niektóre aspekty działania mózgu. Metafora mózg-komputer jest dobra tylko w bardzo ograniczonym zakresie. Dotyczy to np. rozwiązywania abstrakcyjnych problemów, budowania modeli mentalnych, procesów wymagających przetwarzania informacji symbolicznej, ale nie dotyczy tak podstawowych funkcji jak czucie i świadomość.

Z drugiej strony nawet najprostsze neuronowe modele pamięci asocjacyjnej wykazują zdolność do rozpoznawania uszkodzonych wzorców, czas rozpoznania nie zależy dla nich od liczby zapamiętanych wzorców, uszkodzenie części wzajemnych połączeń sztucznych neuronów nie prowadzi do zapomnienia konkretnych wzorców, pomyłki dla podobnych wzorców są częstsze niż dla wyraźnie różnych, a przepełnienie pamięci prowadzi do chaotycznego zachowania. Są to znane z psychologii cechy charakterystyczne dla ludzkiej pamięci. W ostatnich latach bardzo intensywnie rozwija się komputerowa psychiatria, dziedzina wykorzystująca proste modele oparte na sieciach neuronowych do wyjaśnienia wielu aspektów syndromów neuropsychologicznych i chorób psychicznych, począwszy od różnego rodzaju amnezji (w tym dziecięcej), procesów degeneracji funkcji poznawczych, natura schizofrenicznych halucynacji jak i pojawiania się odczuć kończyn fantomatycznych (Duch 2000). Nie ma wątpliwości, że im dokładniejsze modele mózgu uda się stworzyć, tym więcej funkcji umysłu uda się za ich pomocą uchwycić. Czy proces ten zbiega się do „prawdziwego”, świadomego umysłu?

Spróbuję uzasadnić, że mózgowopodobna organizacja przetwarzania informacji musi prowadzić do pojawienia się przestrzeni wewnętrznej, a w niej umysłu i wirtualnej świadomości. Sztuczny umysł, w miarę doskonalenia modelu przetwarzania informacji opartego na mózgowopodobnej organizacji, stanie się coraz bardziej podobny do prawdziwego umysłu. Będzie nie tylko twierdzić, iż ma wrażenia, ale i będzie sam o tym przekonany, interpretując stan swojej pamięci roboczej.

3. Czym są wrażenia?

Umysł jest układem kontrolnym systemu (organizmu) o specyficznej architekturze, pozwalającej tworzyć wewnętrzną, subiektywny model świata (por. Piłat 1999). Najbardziej istotne części tej architektury to pamięć robocza (WM), pamięć trwała, automatyczna aktywacja skojarzeń przez stany pamięci roboczej, oraz oparta na niej zdolność do rozróżniania zmieniających się w sposób ciągły stanów pamięci roboczej. Bardziej rozwinięte umysły, takie jak ludzkie, posiadają również reprezentację swojego jestestwa w postaci zespołu struktur przechowywanych w pamięci długotrwałej, mechanizm filtracji informacji zagrażającej stabilności tego zespołu, oraz możliwość symbolicznego komentowania stanu WM. „Reprezentacja” nie oznacza tu homunkulusa, a jedynie zdolność do rozpoznania takich stanów mózgu, do których istotny wkład mają te ślady pamięci i sposoby zachowań, zapisane w nauczonych strukturach neuronalnych, które odnoszą się do „ja”.

Podstawowe operacje wykonywane przez umysły to kategoryzacja, wartościowanie i odnośnienie się do modelu wewnętrznego jestestwa. Konieczne jest szybkie porównanie np. smaku próbowanego pożywienia z smakami zapamiętanymi w pamięci długotrwałej. Szczury jak i pierwotne ssaki wykształciły tę umiejętność w wysokim stopniu, gdyż ich organizm jest bardziej delikatny niż organizm gadów, a środowisko wymusiło na nich szybkie podejmowanie decyzji. Pamięć roboczą uznać można za centralny system dystrybucji informacji do wszystkich części mózgu (por. teoria Baarsa 1988). Pamięć robocza (WM) ma niewielką pojemność, oparta jest na dynamicznych stanach mózgu, przechowuje tylko informacje konieczne do podjęcia bieżących decyzji. Pamięć długotrwała ma ogromną pojemność, pozwala odnieść się do zgromadzonej w ciągu całego życia wiedzy o świecie. Jej konstrukcja umożliwia pojawienie się szybkich skojarzeń poprzez interferencję wywołaną aktualnym stanem pamięci roboczej (jest to biologiczne rozwiązanie problemu odnalezienia istotnych fragmentów wiedzy, zwane w sztucznej inteligencji „problemem ram”). Szczur smakując nowe pożywienie wprowadza informację o wyniku przetwarzania bodźców węchowych i smakowych do swojej pamięci roboczej. Informacja ta udostępniana jest pamięci trwałej i pojawia się skojarzenie, oparte na wspomnieniu – gdyby szczur mógł to skomentować, zapewne stwierdziłby: ostatnim razem po zjedzeniu czegoś podobnego ledwo przeżyłem. Pojawia się reakcja wstrętu i związane z nią wrażenia, czyli interpretacja stanu mózgu, powstałego w wyniku skojarzenia bodźców zmysłowych z zapamiętanymi epizodami.

Adekwatna reprezentacja symboliczna stanów mózgu zmieniających się w sposób ciągły nie jest możliwa i byłaby niepotrzebna. Zwierzęta i niemowlęta nie używają języka, a jednak ich wrażenia nie różnią się jakościowo od tych, jakie doznają dorośli ludzie. Czym więc są wrażenia i dlaczego istnieją? Dyskryminacja i kategoryzacja sygnałów z różnych zmysłów i różnych sygnałów tej samej modalności ułatwia porównania. Rezultaty tej dyskryminacji pojawiają się jako stany umysłu – może pojawić się myśl (reprezentacja dźwięku, symbolu, niewypowiedzianego zdania, rezultat działania „rozsądku”, czyli *common sensorum*, zmysłu wspólnego), lub wrażenie, czyli niewerbalna kategoryzacja stanu mózgu, który może się zmieniać w sposób ciągły. Powstanie wrażenia wymaga stałej, niewerbalnej aktualizacji stanu pamięci roboczej. Wrażenia można więc traktować jako „niewerbalne kategorie” pozwalające na porównanie ciągłych reprezentacji stanów zmysłów na poziomie globalnej dynamiki mózgu. Mechanizm tej kategoryzacji oparty jest prawdopodobnie na „adaptacyjnych stanach rezonansowych”, zgodnie z modelem ART Grossberga (1995).

Wrażenia różnią się między sobą jakościowo, gdyż są to różnego rodzaju pobudzenia (globalne stany dynamiczne) mózgu, obejmują aktywację większości struktur mózgu w różnym stopniu. Do powstania „normalnych” wrażeń, czyli takich, jakie zwykle mamy, konieczne są prawidłowo przebiegające procesy w mózgu, ale wrażeniom powstającym na skutek tych funkcji w mózgu skłonny jestem przypisać status ontologiczny – są, istnieją nie mniej realnie niż jakiegokolwiek inne przedmioty fizyczne. Jednocześnie podlegają interpretacji i pełnią istotną rolę w relacyjnych strukturach charakteryzujących obiekty umysłu. Czerwony kolor wywołuje we mnie specyficzne wrażenie nie tylko z powodu moich dyspozycji mentalnych (w tym skojarzeń wpływających na stan pamięci roboczej), lecz również realnych stanów mózgu/umysłu. Nie chodzi tu o redukcję stanów świadomości do stanów mózgu – raczej o teorię relacyjną, gdyż wszelka interpretacja tych stanów odwołuje się do ich wzajemnych powiązań. Zrozumienie tych powiązań musi w szczególności uwzględniać rolę środowiska, w którym dany organizm się rozwijał, procesów rozwojowych i kultury,

która go ukształtowała. Wzajemne relacje stanów danego umysłu tworzą gęstą sieć powiązań, pozwalającą zrozumieć wewnętrzny punkt widzenia tego umysłu jak i jego sposób przeżywania świata.

Takie ujęcie natury wrażeń pozwala zrozumieć wiele aspektów dotyczących ich przeżywania. Wrażenia są ściśle związane z mechanizmami uwagi, które decydują o tym, co dla organizmu jest najważniejsze dla podjęcia decyzji kontrolujących zachowanie. Większość bodźców nie wywołuje w nas wrażeń, gdyż przetwarzane są na poziomie lokalnych struktur mózgu, a mechanizmy uwagi filtrujące bodźce w danym momencie nieważne nie pozwalają im dotrzeć do pamięci roboczej. Nowe wrażenia dość szybko mogą być uznane za nieważne i przestają mieć wpływ na pamięć roboczą dzięki mechanizmom uwagi, zwanym procesami habituacji. Czyste doświadczenie fenomenalne, czyste przeżycie, które nie zmienia w żaden sposób zachowania i jest całkowicie wewnętrzną własnością umysłu, to filozoficzna fikcja, która odpowiedzialna jest za wiele nieporozumień. Myśl różni się od głośnej wypowiedzi tym, że stan mózgu z nią związany nie pobudza dostatecznie silnie ośrodków ruchowych by doprowadzić do odpowiedniego skurczu mięśni (czytanie połączone z ruchem warg jest pośrednim stanem pomiędzy myśleniem a mówieniem). Podobnie przeżycie fenomenalne nie zawsze musi prowadzić do zewnętrznie obserwowanych zmian w zachowaniu lub możliwości wyraźnej kategoryzacji i wynikającego stąd opisu symbolicznego. Nie znaczy to jednak, że nie mamy do czynienia z różnymi stanami mózgu, które w coraz większym stopniu będziemy mogli rozróżniać za pomocą technik obrazowania mózgu, podobnie jak za pomocą analizy EEG precyzyjnie rozróżniamy kilka stanów snu, chociaż nie potrafimy tego zrobić na podstawie introspekcji.

Wrażenia spełniają rolę funkcjonalną, taki przynajmniej jest ich sens ewolucyjny. Odczuwanie jakości wrażeń wymaga interpretacji stanów WM przez mechanizmy poznawcze. Jeśli te mechanizmy nie działają odczuwanie jakości wrażeń powinno zniknąć. Obserwujemy to w stanach absorpcji kontemplacyjnej. Pytanie „czym jest świadomość” nie znika, tak jak u Dennetta (1991), który sprowadza wszystko do roli dyspozycji. Świadomość pierwotna jest zdolnością do interpretacji stanów pamięci roboczej. Świadomość refleksyjna wymaga modelu jestestwa (Dasein Heideggera). Poczucie tożsamości wiąże się z wewnętrznym modelem ciała i propriocepcją, przewidywaniem skutków swoich ruchów. Wrażenie istnienia, poczucie bytu, jest przede wszystkim związane z procesami kategoryzacji reprezentacji senso-motorycznych podstawowych procesów wynikających z posiadania ciała.

4. Rola mechanizmów poznawczych w powstawaniu wrażeń.

Wrażenia świadome nie mają specjalnego statusu wśród innych zjawisk umysłowych. Do ich wyjaśnienia nie są konieczne nowe prawa natury (Chalmers 1996), psychony (Eccles 1985), efekty kwantowe (Stapp 1993, Penrose 1994) czy protofenomeny (Rosenberg 1999). Z powyższych rozważań można wyciągnąć wiele wniosków i przewidywań dotyczących sposobu odczuwania wrażeń.

Za poczucie jakości wrażeń odpowiedzialne są skomplikowane mechanizmy poznawcze. Zawieszenie działania tych mechanizmów prowadzi do zaniku wrażeń (np. habituacja, brak uwagi). Brak interpretacji docierających do mózgu sygnałów oznacza brak wrażeń. Można „patrzeć nie widząc”, nie zauważając nawet wyraźnych zmian w środowisku, np. nowych mebli w domu, pomimo prawidłowego działania, np. omijania tych mebli (por. O'Regan,

Nöe 2001). Dopiero pojawienie się faktu w pamięci roboczej powoduje, że stajemy się go świadomi i pojawiają się związane z nim wrażenia. Silna koncentracja na wykonywanym zadaniu modyfikuje lub całkowicie blokuje wrażenia (np. bólu w czasie walki), absorpcja w stanach kontemplacji może nawet spowodować całkowity zanik wrażeń. Z drugiej strony gra aktorska, udawanie emocji przy pełnej identyfikacji z odgrywaną rolą, może doprowadzić do przeżywania prawdziwych emocji.

Wyłączenie dominującej półkuli mózgu np. za pomocą próby Wada (Churchland 1986) lub drażnienia prądem zaburza mechanizmy interpretacyjne i powoduje zanik rozpoznawalnych wrażeń. Aktywność pozostałej półkuli nie wystarcza do precyzyjnej interpretacji, stąd wrażenia stają się nieokreślone, rozmyte. Brak aktywności mózgu znosi oczywiście wszelkie wrażenia. Trenowanie mechanizmów poznawczych prowadzi do „wyostrzenia zmysłów”, nowych jakości wrażeń słuchu, smaku, wzroku. Np. już po pół godzinie pracy nad dużą układanką (puzzle) można zauważyć subtelne zmiany w postrzeganiu kształtów i kolorów. Pamięć wrażeń konieczna jest do dobrej kategoryzacji.

Nowe wrażenia dostępne są również w snach, gdyż opierają się na zmianach w pamięci długotrwałej. Zmiany te powodują powstawanie innego rodzaju skojarzeń z informacją zawartą w pamięci roboczej. Odczuwanie wrażeń zachowane jest w trakcie snów, w fazie REM, w której widać dużą aktywność kory (świadcząca o działaniu pamięci roboczej). Możliwa jest kontynuacja wrażeń sennych na jawie, np. pod wpływem *chlordiazepoxidu*.

Znikające wrażenia pociągają za sobą zredukowaną kompetencję behawioralną, co widać np. w przypadku ślepoty korowej (Dennett 1991). Funkcjonalna rola wrażeń pociąga za sobą różne własności strukturalne odczuwania wrażeń. Wrażenia wzrokowe, dotykowe, czucie temperatury i bólu mają strukturę przestrzenną, w mniejszym stopniu mają ją wrażenia słuchowe, ale całkiem pozbawione są niej wrażenia smaku, zapachu, myśli czy wyobrażenia. Modalności, za pomocą których odbieramy te wrażenia pełnią rolę drugorzędną. Aparatura dotykowa, przekształcająca sygnały z kamery na pobudzenia skóry na plecach, po okresie przyzwyczajenia się do jej używania dostarcza osobom niewidomym wrażeń podobnej natury co wzrokowe, z możliwością oceny położenia przestrzennego, wielkości, kształtu i ruchu obiektu. Nie liczy się więc modalność, konkretne zmysły wprowadzające informację, ale struktura tej informacji, pozwalająca w specyficzny sposób interpretować i wykorzystywać tę informację.

Stymulacja struktur mózgu, odpowiedzialnych za przetwarzanie danych wrażeń, powinna prowadzić do ich powstawania tylko wtedy, gdy mogą one wpłynąć na globalną dynamikę działania mózgu. Dlatego zbyt słaba stymulacja nie wywołuje wrażeń. Powstanie wrażenia wymaga dostarczenia dostatecznie dużej ilości energii na tyle silne pobudzającej neurony mózgu, by wpłynęło to na globalną aktywność mózgu, a nie tylko na lokalne obszary (cf. Libet 1993).

Poczucie wstrętu może powstać w wyniku reakcji na bodźce wzrokowe, słuchowe i smakowe. Czym uzasadnione jest używanie tego samego określenia „wstręt” na wrażenia wywołane w różny sposób? Z introspekcji wiemy, że chodzi o ten sam stan umysłu. Powinniśmy więc oczekiwać, że pewne wrażenia sensoryczne obecne w pamięci roboczej wywołują interferencję z jakimiś obszarami mózgu, które zmieniają stan tej pamięci na stan określany jako wstręt. Kategoryzacja stanów pamięci roboczej może polegać na prostym procesie interferencji z obszarem mózgu specjalizującym się w rozpoznawaniu jakiegoś stanu. Tak

się istotnie dzieje – pobudzenia różnej modalności prowadzą do tego samego stanu mózgu, pobudzenia przedniej części wyspy, rozpoznawanego jako wstręt. Być może pobudzenia tego obszaru przez nerwy czuciowe dochodzące od różnych kończyn są odpowiedzialne za dziwny syndrom neuropsychologiczny jaki jest dysmorfia ciała. Pacjenci cierpiący na ten syndrom odczuwają bardzo silny wstręt do swoich kończyn decydując się czasami na ich amputację. Błędna interpretacja braku sygnałów w mózgu może też być odpowiedzialna za zaniedbanie jednostronne, czyli brak wrażeń dotyczący połowy przestrzeni. Syndrom ten pojawia się najczęściej po wylewie w obszarze prawego płata ciemieniowego i nie pozwala pacjentom wyobrazić sobie lewej połowy przestrzeni (Walsh 1998).

Z badań nad mózgiem wiadomo, że uszkodzenia pierwotnych obszarów sensorycznych lub szlaków doprowadzających do nich sygnały powodują zanik wrażeń. Uszkodzenia wtórnych obszarów sensorycznych wywołują asymbolię (Ramachandran 1999), np. asymbolię bólu, zjawisko polegające na niezdolności do interpretacji odbieranych wrażeń jako bólu. Ramachandran opisuje fascynujące eksperymenty z osobami cierpiącymi z powodu paraliżu kończyn fantomatycznych. Osobom tym wydaje się, że ich amputowane kończyny nadal istnieją i czują w nich okropny ból, np. z powodu paznokci wrastających w nieistniejącą dłoń, której nie potrafią otworzyć. Jeśli pacjentom wydaje się, że amputowana ręka jest widziana – iluzję taką można stworzyć pokazując zdrową rękę w odpowiednio ustawionym lustrze – zaczynają czuć jej ruch i w końcu udaje im się rękę otworzyć, co prowadzi do zaniku bólu. Wrażenie bólu jest więc związane z interpretacją pojawiających się w mózgu pobudzeń.

Fenomenologia bólu wydaje się być bardzo obiecującym terenem szczegółowych badań nad naturą wrażeń. Już w okresie wojennym chirurdzy zauważyli, że bez interpretacji poznawczej ból pozostaje czystym wrażeniem (np. Beecher 1946). Piekący ból pourazowy, pojawić się może pod wpływem dowolnego bodźca. Pełna koncentracja na przeżyciu bólu może doprowadzić do zaniku cierpienia. Ból zwykle wywołuje cierpienie, ale u masochistów przyjemność. Próg wrażliwości na ból zależy od oczekiwań. Analiza zjawisk związanej z przeżyciem bólu jest dobrym testem przedstawionej tu hipotezy dotyczącej natury wrażeń i świadomości.

5. Automatyzacja wyuczonych umiejętności.

Wszystkie wypowiedzi dotyczące wrażeń wymagają działania mechanizmów poznawczych. Jestem przekonany, że mam wrażenia, które nie są tylko złudzeniem, ale realnymi stanami mojego mózgu, a tym samym umysłu. Czy podobne przekonania mieć może maszyna? Jeśli tak, należy uznać, że komputery o odpowiedniej architekturze mogą być świadome i mieć wrażenia, przeżywać ich specyficzne jakości podobnie jak my przeżywamy nasze wrażenia. Mozgopodobne przetwarzanie informacji musi prowadzić w granicy coraz bardziej złożonych modeli do powstania systemów, które komentując stany swojej pamięci roboczej będą miały przekonanie, że mają wrażenia. Czy możemy uznać, iż będzie to przekonanie fałszywe? Nie widzę dobrych argumentów przeciw takiej możliwości. Eksperymenty myślowe wymyślone po to, by zaprzeczyć możliwości powstania sztucznych umysłów nie są przekonujące. Oczywiście wrażenia identyczne z ludzkimi wymagają identycznych mózgów, ale wrażenia ludzko-podobne tylko podobnych relacji pomiędzy stanami (wzorcami pobudzeń) pamięci roboczej realizowanymi przez mózgi. Te pobudzenia tworzą tkankę wewnętrzną przestrzeni, w której mieści się nasz model świata, sposoby

reakcji, typowe zachowania, skojarzenia i łańcuchy skojarzeń prowadzące do dialogu wewnętrznego, czyli to co nazywany umysłem. Iluzja, że to właśnie „duch w maszynie” rządzi naszym zachowaniem jest bardzo silna (por. Eccles 1985). Nadal brakuje dobrej fenomenologii świadomości, która by tą iluzję podważyła, potrzebna jest neurofenomenologia (Varela 1996), głęboka introspekcja, a nie powierzchowne obserwacje.

Rozważania filozoficzne dotyczące umysłu uwikłane są w problemy lingwistyczne. Używając języka potocznego do opisu umysłu skazani jesteśmy na brak precyzji. Próba wprowadzenia pośredniego języka, opisującego zdarzenia umysłowe, języka na poziomie pośrednim pomiędzy zdarzeniami neurofizjologicznymi a psychologicznymi, przedstawiona została w (Duch 1994, 1996, 1997) oraz inne ujęcie w (Churchland 1995). W istocie nie potrafimy do końca zdefiniować żadnego obiektu czy pojęcia fizycznego, możemy natomiast mówić precyzyjnie o relacjach pomiędzy nimi. Wprowadzenie precyzyjnego języka powinno się przyczynić do rozpoznania pozornych problemów dotyczących natury umysłu, np. takich jak nielokalność czy nieprzestrzenna natura umysłu, dyskutowane przez zwolenników teorii kwantowych i innych (McGinn 1990, Stapp 1993).

Czy mamy dobry test, odróżniający systemy świadome, posiadające wrażenia, od systemów przetwarzających informację? Jeśli kiedyś dojdzie do spotkania dziwacznych form inteligentnego życia czy będziemy mogli rozróżnić pomiędzy mechaniczną inteligencją a świadomą istotą? Słynny eksperyment z Chińskim Pokojem (J. Searle, 1999; Kloch 1996) nie jest tu szczególnie przydatny. Nie jest to test, gdyż jego wynik jest zawsze negatywny, z góry przesądzony. Jeśli zagłębimy do wnętrza mózgu czy Chińskiego Pokoju nie znajdziemy tam „rozumienia”. Należy się zapytać, jakie spełnione muszą być warunki by obserwator zrozumiał, co robi czy czuje dany system?

Alison Gopnik (1998) napisała wielce interesujący artykuł pod tytułem „Zrozumienie jako orgazm”. Zbudowanie modelu, zrozumienie czegoś, wymaga z ewolucyjnego punktu widzenia u inteligentnych istot sygnału nagrody, zachęcającego do tego typu działania. „Narrator”, czyli ta część mózgu, która usiłuje tworzyć symboliczne, łatwe do zapamiętania komentarze zachowania organizmu (por. Gazzaniga 1997), daje sygnał zakończenia pracy, interpretowany jako przyjemne zdarzenie. Narrator rozwinął się z obszarów zaangażowanych w tworzenie teorii o świecie, aktywnych w dzieciństwie (i u niektórych naukowców w wieku dojrzałym). Ten sam mechanizm działa na poziomie rozumienia zdania jak i złożonej teorii.

Zrozumienie w sensie emocjonalnego poczucia „rozumiem”, oraz zrozumienie intelektualne w sensie integracji nowej wiedzy i odpowiedzi na pytania z nią związane, to dwie różne rzeczy. Tłumacze kabinowi poprawnie tłumaczący zdania nie potrafią czasami odpowiedzieć na najprostsze pytania dotyczące tłumaczonej rozmowy. Koncentracja na tłumaczeniu i poczucie „rozumiem”, nie gwarantuje zrozumienia intelektualnego. Wewnętrzne poczucie zrozumienia jest tylko sygnałem mózgu: zrozumiałem, mogę działać dalej. Dysocjacja pomiędzy poczuciem rozumienia a rozumieniem intelektualnym pojawia się po zażyciu niektórych środków halucynogennych (np. LSD). Skrajnym przypadkiem braku poczucia zrozumienia pomimo rozumienia czy rozpoznania intelektualnego jest syndrom Capgrasa (Walsh 1998).

Nie warto więc odwoływać się do tajemnych mocy biologicznych neuronów by uzasadnić, że w naszych mózgach zachodzi prawdziwe rozumienie a w systemach sztucznych nie jest

ono możliwe (Searle 1999a). Nie potrafimy tego rozpoznać inaczej, niż poprzez testy behawioralne, a więc test Turinga lub jego odmiany. Jedynie doprowadzając system sztuczny i nasze mózgi do rezonansu, mapując i przekształcając struktury pobudzeń systemu sztucznego na struktury mózgu, moglibyśmy się przekonać czy jest to rzeczywiście system rozumiejący w naszym sensie. Trudno sobie jednak wyobrazić tego typu rezonans. Wymaganie, by system działał na tyle podobnie do naszego mózgu, by dało się znaleźć takie odwzorowanie prowadzące do rezonansu, jest niepotrzebnym ograniczeniem na architekturę systemu, który mógłby mieć poczucie rozumienia. Najważniejsze są relacje pomiędzy stanami wewnętrznymi takiego systemu i jego działaniami sensomotorycznymi. System taki musiałby wykazywać jakąś formę samoświadomości, „wiedzieć, że wie”, a więc „być świadkiem” stanu swojej pamięci roboczej.

Hilary Putnam (1998) doszedł do przekonania, że za całe zamieszanie w filozofii umysłu odpowiada przekonanie, iż postrzeżeniom towarzyszą jakieś wewnętrzne stany fenomenalne, identyczne za każdym razem, gdy mamy takie same wrażenia. Chociaż argument wysuwany przez Putnama na poparcie tej tezy jest zły, to sama teza wydaje się być poprawna. Putnam rozważa eksperyment, w którym mamy 100 kart, kolejne karty pomalowane białą farbą do której za każdym razem dodajemy kroplę czerwonej farby, porównywane są parami (1,2), (2,3) aż do (99,100). Karty każdej pary wydają się mieć identyczną barwę, a więc odpowiadają identycznym stanom wewnętrznym. Ponieważ relacja „bycia w określonym stanie” jest przechodnia, pisze Putnam, więc dla karty 1 i 100 stany wewnętrzne są identyczne, ponieważ dla każdej pary były identyczne. Założenie, że oglądając dwie karty jesteśmy w jednym ze stanów wewnętrznych typu „barwa nr x ” jest oczywiście błędne, bo mamy tu proces ciągły, a nie stany dyskretne.

Mózg nie jest urządzeniem dyskretnym, nie ma w nim identycznych stanów za każdym razem, gdy rozpoznajemy ten sam obiekt, są co najwyżej stany dostatecznie podobne by je zaliczyć do tej samej kategorii i by wywołały podobne skojarzenia. Prawdopodobnie nawet to nie jest potrzebne. Istotne są relacje pomiędzy stanami, a nie same stany wewnętrzne. Walter Freeman (1995) badając zachowania królików rozpoznających ten sam zapach i reagujących na niego w taki sam sposób stwierdził, że stan kory węchowej jest za każdym razem odmienny, nie ma prostych „reprezentacji wewnętrznych”. Stan kory węchowej zależy od wielu procesów zachodzących w mózgu królika. Jedynie relacje pomiędzy stanem „rozpoznano określony zapach”, a wywołanym tym rozpoznaniem zachowaniem, są istotne.

Jedną z zagadkowych cech naszych mózgów jest zdolność do automatyzacji wyuczonych umiejętności. Nauka jazdy samochodem wymaga początkowo ciągłej, świadomej uwagi, ale po dłuższym czasie staje się umiejętnością w pełni zautomatyzowaną. Metody obrazowania mózgu pokazują przy wykonywaniu nowych czynności zaangażowanie wielu obszarów mózgu, a po nauczaniu się danego zadania jedynie zlokalizowane pobudzenia określonych struktur mózgu. Jak zachodzi ten zagadkowy proces przechodzenia procesów świadomych w nieświadome w wyniku uczenia się?

Tajemniczość tego procesu wynika z założenia, że istnieje coś takiego jak świadome wykonywanie czynności. Jak jest funkcja świadomości w tym procesie? Nie jest nią sterowanie samym wykonywaniem procesu, bo to realizowane jest przez obszary mózgu, które odpowiedzialne są za pamięć wyuczonych ruchów, pamięć umiejętności. Początkowo nie działają one zbyt precyzyjnie i muszą się dopiero nauczyć prawidłowego działania. Mecha-

nizm tego uczenia wykorzystywany jest przez mózg wielokrotnie, szczególnie przy uczeniu przez analogię i imitację. W teorii maszynowego uczenia istnieje duża grupa metod, nazywana „uczeniem się zachowań”, lub „uczeniem się z krytykiem” (reinforcement learning). Informacja zwrotna, używana do określenia pożądanego kierunku zmian, jest wynikiem krytyki, opartej na ocenie wyniku działania.

Uczenie z krytykiem jest podstawą nabywania wszelkich umiejętności. Niemowlę otrzymuje od środowiska różne sygnały i może na nie reagować w różny sposób. Niektóre z tych działań są nagradzane, inne karane. Użycie uczenia z krytykiem pozwoliło na stworzenie programu komputerowego, nazwanego „maszyną dziecięcą” (child machine), która w ocenie psychologa rozwojowego osiągnęła kompetencje językowe porównywalne z możliwościami 18-miesięcznego dziecka (Treister-Goren, Hutchens, 2000). System ten wykształcił swoje zdolności językowe tylko na podstawie krytyki jego zachowań, w efekcie „gruntuując” sens używanych przez siebie symboli (czy też ogólnie wszelkich zachowań) w odniesieniu do stanów środowiska. Ten podstawowy mechanizm pozwala zrozumieć zarówno funkcję, jaką spełnia świadomość, jak i proces automatyzacji uczenia się umiejętności.

Krytyka wymaga złożonej oceny wyników własnego działania i porównania z zapamiętanymi obserwacjami. Wynikiem tej oceny jest również zaangażowanie emocji, nagroda wynikająca z sukcesu i kara wynikająca z porażki. Układ emocjonalny reguluje poziom neurotransmiterów i neuromodulatorów w korze mózgu, wpływając na plastyczność mózgu, zwiększając szybkość uczenia się pożądanых zachowań. Elementy potrzebne do oceny i wyniki tej oceny oraz reakcje emocjonalne z niej wynikające pojawiają się w pamięci roboczej. Chociaż pamięć robocza zawiera jednocześnie kilka elementów to uświadamiane są te z nich, które trzeba zapamiętać by umożliwić proces krytyki.

Świadomość nie jest więc bezpośrednio zaangażowana w proces uczenia się umiejętności, ale jest konieczna do krytyki i oceny wyników własnego działania. Jeśli wyniki tego działania są zgodne są z zapamiętanymi oczekiwaniami i poziom kompetencji wyuczonego działania jest dostatecznie wysoki proces porównywania oczekiwań i stanu bieżącego (zachodzący prawdopodobnie w podporze, w formacji hipokampa, Gray 1995) nie wywołuje emocji i wynik nie jest zapamiętywany ani komentowany. Pamięć robocza przestaje być zaangażowana w ten proces, staje się on „podświadomy” i wykonywany jest przez zlokalizowane obszary mózgu. Jeśli jednak popełnimy pomyłkę procesy podkorowe wywołają nieprzyjemne emocje, będące sygnałem kary i pojawiające się w pamięci roboczej. Wynik staje się świadomy, gdyż powinien zostać zapamiętany, pozwalając unikać podobnych błędów w przyszłości.

6. Podsumowanie

Poszukiwanie rozumienia świadomości w teoriach fundamentalnych jest błędne. Nic nie wskazuje na konieczność odwołania się do efektów kwantowych, protofenomenów, psychonów, fenomenalnych aspektów informacji ani innych koncepcji wymagających fundamentalnej rewizji naszego obrazu świata. Świadomość i wrażenia pojawiają się jedynie w normalnie działających systemach o mózgo-podobnej organizacji.

Przedstawione tu idee po raz pierwszy pozwalają na wyciągnięcie wniosków dotyczących natury wrażeń świadomych, przełamując impas w dyskusjach nad trudnym problemem świa-

domości. Wydaje mi się, że żadna alternatywna teoria nie zaszła tak daleko i nie potrafi wyjaśnić dlaczego wspominając smak ostatnio zjedzonego lodu mam wrażenia, a wspomnieniom o sposobie wiązania krawata żadne jakości wrażeń z tym związanych nie towarzyszą. W pierwszym z tych przypadków ślady pamięci pobudzają się poprzez rezonans adaptacyjny z pamięcią roboczą. W drugim mamy do czynienia z pamięcią typu umiejętności, działaniem sensomotorycznym, które nie może modyfikować pamięci roboczej a może jedynie być wykonane. Jedynie procesy zachodzące w pamięci roboczej podlegają kategoryzacji i mogą się stać treścią świadomości. Wśród wszystkich elementów pamięci roboczej tylko te, które pełnią rolę krytyka wzmacniającego bądź osłabiającego tendencje do pewnych zachowań jest uświadamiany.

Prawidłowo działające funkcje mózgu są konieczne by być świadomym wrażeń, ale wrażenia nie sprowadzają się do funkcji – tu Chalmers (1996) ma rację. Wrażenia istnieją w postaci procesów fizycznych, którym – poprzez relacje z innymi procesami i działaniami sensomotorycznymi – nadajemy sens. Neurokognitywna teoria świadomości wymaga zdefiniowania i określenia relacji pomiędzy takimi pojęciami jak świadomość, pamięć robocza, uwaga, nieświadomość, umysł. Zadaniem umysłu jest kontrolować organizm na podstawie wewnętrznego modelu świata. Na umysł składa się zbiór relacji i stanów interpretowalnych wewnątrznie zarówno w sensie kategoryzacji werbalnej jak i niewerbalnej. Struktura i możliwe relacje między tymi stanami wynikają z przeszłych stanów mózgu, zapisanych w postaci śladów pamięci i wyuczonych struktur pobudzeń mózgu odpowiedzialnych za sposób, w jaki steruje on organizmem.

W oparciu o zarysowane tu idee powinno się dać zbudować szczegółową, neurokognitywną teorię świadomości. W następnym artykule postaram się pokazać jak takie podejście wyjaśnia paradoksy oparte na filozoficznych eksperymentach myślowych, takich jak „chiński pokój” Johna Searle. Przy odpowiedniej konstrukcji sztuczne mózgi będą twierdzić, że mają wrażenia i są tych wrażeń świadome. Jednak konstrukcja sztucznego mózgu jeszcze długo będzie się różniła od mózgow prawdziwych, a tym samym sztuczne umysły będą odmienne od naszych.

Podziękowania

Artykuł ten jest rozwinięciem pomysłów przedstawionych w referatach wygłoszonych na „Symposium kognitywnym: *Subiektywność a świadomość*”, w Obrzycku, maj 2000, oraz na Konstanz Universität. Zaproszenia te zmobilizowały mnie do uporządkowania poglądów na temat natury świadomości i dlatego jestem za nie wdzięczny prof. Andrzejowi Klawiterowi i prof. Maxowi Urchsowi.

Literatura

- B.J. Baars (1988) *A cognitive theory of consciousness*. Cambridge Univ. Press.
B.J. Baars (1997) *In the Theater of Consciousness: The Workspace of the Mind*. Oxford University Press
N. Block (1995) *On a confusion about a function of consciousness*. Behavioral and Brain Sciences 18 (2): 227-287
D.J Chalmers (1995) *Facing up to the problem of consciousness*. J. of Consciousness Studies 2: 200-219

- D.J. Chalmers (1996) *The conscious mind: In search of a fundamental theory*. New York, NY, USA: Oxford University Press
- P.S. Churchland (1986) *Neurophilosophy. Toward a unified science of the mind/brain*. MIT Press, Bradford Book
- P. Churchland (1995) *Engine of reason*. MIT Press.
- C.J.S. Clarke (1995) *The nonlocality of mind*. Journal of Consciousness Studies 2:231-40
- F. Crick, C. Koch (1995) *Why neuroscience may be able to explain consciousness*. Scientific American 273(6): 84-85
- D.C. Dennett (1991) *Consciousness explained*. Little, Brown and Co, Boston, Toronto, London
- W. Duch (2000) *Sieci neuronowe w modelowaniu zaburzeń neuropsychologicznych i chorób psychicznych*. W: Biocybernetyka, Tom 6: Sieci neuronowe (red. W. Duch, J. Korbicz, L. Rutkowski i R. Tadeusiewicz), rozdz. II.18. Akademicka Oficyna Wydawnicza-Exit, Warszawa.
- W. Duch (1994) *A solution to the fundamental problems of cognitive sciences*. International Philosophical Preprint Exchange (1994)
- W. Duch (1996) *From brain to mind to consciousness without hard problems*. Sympozjum Kognitywne '96: Świadomość a Percepcja. Instytut Filozofii UAM, Poznań, 6-7 Dec. 1996.
- W. Duch (1997) *Platonic model of mind as an approximation to neurodynamics*. W: Brain-like computing and intelligent information systems, red. S-i. Amari, N. Kasabov. Springer, Singapore, rozdz. 20, str. 491-512
- J. Eccles (1985) *The Brain and the unity of conscious experience*. Cambridge University Press.
- G. Edelman (1999) *Przenikliwe powietrze, jasny ogień. O materii umysłu*. PIW, Warszawa.
- W.J. Freeman (1995) *Societies of Brains: A study in the neuroscience of love and hate*. Lawrence Erlbaum Associates
- M.S. Gazzaniga (1997) *O tajemnicach ludzkiego umysłu. Biologiczne korzenie myślenia, emocji, seksualności, języka i inteligencji*. Książka i Wiedza, Warszawa
- D. Giulini, E. Joos, C. Kiefer, J. Kupsch, I.O. Stamatescu, H. D. Zeh (1996) *Decoherence and the appearance of a classical world in quantum theory*. Berlin: Springer.
- A. Gopnik (1998) *Explanation as orgasm*. Minds and Machines 8: 101-118
- J.A. Gray (1995) *The contents of consciousness: a neuropsychological conjecture*, Behavioral and Brain Sciences 18: 659-722
- S. Grossberg (1995) *The attentive brain*. American Scientist 83: 483-449
- J. Kloch (1996) *Świadomość komputerów?* Biblios Tarnów i OBI Kraków.
- B. Libet (1993) *Neurophysiology of Consciousness. Collected papers and new essays*. Birkhäuser, Boston, Basel, Berlin.
- C. McGinn (1990) *The Problem of Consciousness*. Oxford: Blackwells.
- J. Mehra, A.S. Wightman (1995). *The collected works of E. P. Wigner* (Vol. VI). Berlin: Springer.
- J.K. O'Regan, A. Noë, *A sensorimotor account of vision and visual consciousness*. Behavioral and Brain Sciences 24(5) (2001, w druku)
- R. Penrose (1994) *Shadows of the mind*. Oxford University Press; tłum: *Cienie umysłu*. Wyd. Zysk i Ska, Poznań 2001
- R. Piłat (1999) *Umysł jako model świata*. Wyd. IFiS PAN.
- H. Putnam (1998) dodatek do "The Royce Lectures in the Philosophy of Mind", wykładów wygłoszonych na Brown University, 1998 (dziękuję autorowi za przesłanie mi manuskryptu).
- V.S. Ramachandran (1999) *Consciousness and body image: lessons from phantom limbs, Capgras syndrome and pain asymbolia*. Phil. Trans. R. Soc. Lond. B 353: 1851-1859
- G.H. Rosenberg (1996) *Rethinking nature: A hard problem within the hard problem*. Journal of Consciousness Studies 3:76-88.

- O. Sacks (1999) *Antropolog na Marsie*. Zysk i Ska, Poznań
- W. Seager (1995) *Consciousness, information, and panpsychism*. *Journal of Consciousness Studies* 2:272-88.
- J.R. Searle (1999) *The future of philosophy*. *Philosophical Transactions of Royal Society B* 354: 2069-2080
- J.R. Searle (1999) *Umysł na nowo odkryty*. PIW, Warszawa.
- O. Sporns, G. Tononi, G.M. Edelman (2000) *Connectivity and complexity: the relationships between neuroanatomy and brain dynamics*. *Neural Networks* 13: 909-922
- H.P. Stapp (1993) *Mind, matter and quantum mechanics*. Berlin: Springer.
- M. Tegmark (2000) *The importance of quantum decoherence in brain processes*. *Physical Review E* 61, 4194-4206
- A. Treister-Goren, J.L. Hutchens (2000) *Creating AI: A unique interplay between the development of learning algorithms and their education*. Technical Report, AI Enterprises, Tel-Aviv 2000. Osiągalne pod adresem <http://www.a-i.com>
- F. Varela (1996) *Neurophenomenology: A methodological remedy for the hard problem*. *J. of Consciousness Studies* 3: 330-349
- K. Walsh (1998) *Neuropsychologia kliniczna*. Wyd. Naukowe PWN, Warszawa
- E. P. Wigner (1962). W: I. J. Good (red.), *The scientist speculates: An anthology of partly-baked ideas*. London: Heinemann.

Włodzisław Duch,
Katedra Informatyki Stosowanej, Uniwersytet Mikołaja Kopernika
87-100 Toruń, <http://www.phys.uni.torun.pl/~duch>

Neurokognitywna teoria świadomości

Abstrakt

W ostatnich latach jednym z najbardziej dyskutowanych zagadnień filozofii umysłu jest „trudny problem świadomości”, czyli próba zrozumienia, czym są wrażenia i dlaczego doznania wrażeń mają specyficzne jakości (*qualia*). Pomimo intensywnej dyskusji tego zagadnienia w ostatnich latach nie dokonano żadnego postępu pozwalającego na zrozumienie struktury i przyczyn powstawania wrażeń. W artykule uzasadniam, że wynikają one ze specyficznego sposobu przetwarzania informacji przez mózg, a w szczególności możliwości komentowania stanów pamięci roboczej. Systemy sztuczne o podobnej architekturze będą przekonane, że są świadome i mają wrażenia. Takie rozumienie przyczyny powstawania świadomości pozwala wyciągnąć szczegółowe wnioski dotyczące natury i struktury odczuwanych świadomie wrażeń. W szczególności pozwala odróżnić procesy prowadzące do działań świadomych od pozostałych, oraz zrozumieć przejście od świadomego wykonywania czynności do automatycznego, podświadomego ich wykonywania w wyniku uczenia się jakichś czynności.

Neurocognitive theory of consciousness

Abstract.

Recently the „hard problem of consciousness” has been one of the most intensely debated problems in the philosophy of mind. A discussion of the qualia problem, or the reason for existence of specific qualities of phenomenal experience, has not yet led to any reasonable solution, to understanding of the nature and structure of qualia. This article presents arguments that qualia are a result of specific, brain-like information processing. In particular any system that may comment on the state of its working memory should claim to have qualia and is conscious of its state. Such understanding of the qualia allows to draw specific inferences about the structure of qualia, explains the reasons why some processes give rise to phenomenal experience while others do not, and understand how learning of new skills evolves from initial engagement of the whole brain, requiring conscious activity, to the final automatization of the skill and subconscious information processing, restricted to specific areas of the brain.