

Świadomość i dynamiczne modele działania mózgu.

Włodzisław Duch,

Katedra Metod Komputerowych, Uniwersytet Mikołaja Kopernika,
ul. Grudziądzka 5, 87-100 Toruń, <http://www.phys.uni.torun.pl/~duch>

Streszczenie

Świadomość związana jest z uwagą, pamięcią roboczą i percepcją. Strumień świadomości zawiera tylko te elementy, które pojawiają się w pamięci roboczej. Neurofizjologiczne korelaty wrażeń świadomych mogą obejmować procesy synchronizacji oscylacji 40 Hz odpowiedzialne za scalanie wrażeń (*the binding problem*). Przedstawiono alternatywne teorie rozwiązania lub uniknięcia problemu scalania wrażeń. Szczególnie obiecujące mogą być korelaty oparte na pomiarach współczynników złożoności aktywności neuronowej zaproponowane przez Tonioni i Edelmana. Omówiono teorie „bąbli aktywności” i adaptacyjnego rezonansu, wyjaśniające wiele aspektów związanych z fenomenologią wrażeń świadomych. Przedstawiono argumenty za uznaniem świadomości za proces dyskryminacji nie-symbolicznych reprezentacji pojawiających się w pamięci roboczej o niewielkiej pojemności. Te reprezentacje tworzą stany rezonansowe z reprezentacjami przechowywanymi w pamięci długotrwałej. Jest to proces, którego nie da się symulować za pomocą cyfrowego komputera.

Abstract

Consciousness is connected with attention, working memory and perception. The stream of consciousness contains only those elements that are present in the working memory. Neurophysiological correlates of conscious states are sought in form of synchronization processes based on 40 Hz oscillations needed to solve the binding problem. Alternative theories to solve or avoid the binding problem are presented. Correlates based on neural complexity index recently introduced by Tonioni and Edelman are especially promising. Theories of “activity bubbles” and of adaptive resonance explain many aspects of the conscious’ experience phenomenology. Arguments supporting the idea that consciousness is a process of discrimination of continuous, non-symbolic representations appearing in small-capacity working memory are presented. These representations create resonant states with long-term memory traces, a process that cannot be captured by digital computer simulations.

Słowa kluczowe: świadomość, układy dynamiczne,
Keywords: consciousness, dynamical systems, binding,

1 Wprowadzenie

Badania nad świadomością objęte były przed długi czas rodzajem tabu. Problem natury świadomości dyskutowany był głównie przez filozofów spekulujących na temat umysłu w całkowitym oderwaniu od neurobiologii. Wyjątkiem jest tu stworzony przez Patrycję Churchland nurt „neurofilozofii” [3], krytyczny wobec funkcjonalizmu, a w szczególności argumentów dotyczących niezależności stanów psychologicznych od stanów mózgu. Niestety niewielu

filozofów umysłu poszło w tym kierunku i w tej dziedzinie panuje nadal duże pomieszanie. Z jednej strony niektórzy filozofowie proponują odrzucenie wszelkich pośrednich poziomów opisu, twierdząc, że istnieją tylko stany neurofizjologiczne i świadomość [15], krytykując badanie procesów przetwarzania informacji przez mózg. Jest to wyjątkowo destruktywne podejście, gdyż nie można skonstruować sensownych modeli umysłu bez wprowadzenia całego szeregu poziomów pośrednich – fakty neurobiologiczne muszą być podstawą, ale bez modeli teoretycznych i symulacji komputerowych, pozwalających na stawianie i testowanie hipotez dotyczących funkcjonowania mózgu pozostaniemy na poziomie behawioryzmu. Z drugiej strony dyskusje dotyczące umysłu spychane są na manowce przez filozofujących matematyków, argumentujących – w oderwaniu zarówno od psychologii kognitywnej jak i od neurobiologii – za istnieniem tajemnych własności umysłu, których zrozumienie wymagać będzie rewizji fundamentalnych teorii fizycznych [14].

Niestety takie bezpłodne dyskusje filozoficzne odwracają uwagę od ciągłego postępu w rozumieniu związków pomiędzy stanami mózgu a stanami umysłu, postępu który dokonuje się dzięki modelom dynamicznym typu sieci neuronowych. William James i inni twórcy psychologii w XIX wieku zajmowali się przede wszystkim świadomością. Już w 1949 roku Donald Hebb w swojej książce „*Organization of Behavior*” zaproponował wyjaśnienie wielu zjawisk psychologicznych, związanych z percepcją, uczeniem się, pamięcią i zaburzeniami afektywnymi, w oparciu o procesy zachodzące w sieciach neuronów. We wstępie do swojej książki Hebb utożsamiał te procesy ze zjawiskami psychicznymi, nie posunął się jednakże do próby zrozumienia świadomości. Psychologia kognitywna pomimo odrzucenia ograniczonej perspektywy behawioryzmu stroniła od zagadnień związanych ze świadomością, bojąc się ich pozornie „nienaukowego” charakteru. Dopiero w ostatniej dekadzie XX wieku badania nad świadomością wkroczyły do głównego nurtu nauki. Przyczyniła się do tego seria konferencji organizowanych od 1996 przez University of Arizona w Tucson, oraz powstanie szeregu specjalistycznych pism, takich jak *Consciousness and Cognition* (Academic Press) oraz *Journal of Consciousness Studies* (Imprint Academic). Ważną rolę odegrał również artykuł F. Cricka i C. Kocha [4] formułujący problem świadomości wzrokowej z punktu widzenia neurobiologii oraz popularna książka F. Cricka [5].

2 Problemy związane z definicją świadomości

Trudno jest jednoznacznie zdefiniować świadomość, słowo to używane jest bowiem w wielu różnych znaczeniach i kontekstach, jakiego więc wyjaśnienia można się spodziewać? Czy świadomość da się sprowadzić do uwagi, pamięci roboczej lub percepcji? Czucie, widzenie czy słyszenie są świadomymi wrażeniami wynikającymi ze zwrócenia uwagi na czynności dotykania, patrzenia i słuchania. W języku psychologii potocznej czynimy więc rozróżnienie pomiędzy działaniem uwagi, aktywnego procesu pozwalającego na wybór bodźców, oraz samym procesem doświadczenia tych bodźców jako wrażeń. Uwaga wybiera, które z istniejących w mózgu pobudzeń będziemy świadomie doświadczać, lub pozwala aktywnie poszukiwać nowych wrażeń, np. przez kontrolę ruchu gałek ocznych pod wpływem bodźców akustycznych. Chociaż pojęcia psychologii potocznej mogą się okazać zupełnie nieadekwatne [3] nie należy ich odrzucać pochopnie. Wyjaśnienia zgodne z potocznym sposobem używania pojęć dotyczących świadomości będą się nam wydawać bardziej naturalne niż wyjaśnienia abstrakcyjne. Niestety wyjaśnienia takie nie są uniwersalne, bo w różnych językach pojęcie świadomości i metafory dotyczące umysłu funkcjonują w odmienny sposób (por. rozróżnienie pomiędzy *consciousness* i *awareness* w języku angielskim). Wszystkie wyjaśnienia potoczne

odwołują się w jawny lub ukryty sposób do *homunculusa*, a więc ostatecznie nie dadzą się zaakceptować. Kto kieruje uwagą? Problem ten nie pojawia się jeśli działanie mózgu opiszemy za pomocą układów dynamicznych (por. poniżej).

Mniej jasne jest rozróżnienie pomiędzy pamięcią roboczą (*working memory*) a świadomością. Baddeley [1] uważa, że świadomość działa przez pamięć roboczą, stanowiąc jej aktywny element konieczny do koordynacji informacji z różnych źródeł, zarówno pamięci epizodycznej zdarzeń przeszłych, jak i bieżącej percepcji. W tym ujęciu sama pamięć robocza jest tylko zbiorem „obiektów umysłu” (reprezentacji ikonograficznych, perceptów, myśli), a świadomość jest aktywnym procesem, dokonującym skojarzeń, refleksji i wolicjonalnej oceny pojawiających się możliwości. Baars [2] rozwinął podobne idee w ramach teorii globalnej przestrzeni roboczej (*Global Workspace*), próbując w pewnym stopniu powiązać ją z neurofizjologią. W jego rozumieniu uwaga stwarza dostęp do świadomości zawierającej oprócz perceptów związanych z działaniem zmysłów również intencje, pragnienia, znaczenia abstrakcyjnych pojęć i inne elementy. Świadomość jest natomiast procesem nadrzędnym, koniecznym by powiadomić wszystkie struktury mózgu o pojawiającym się problemie i zebrać cząstkowe wyniki od wyspecjalizowanych, nieświadomych procesorów, zdolnych do rozwiązania fragmentów tego problemu. Zawartość strumienia świadomości (przeżywane wrażenia) jest tożsama z treścią pamięci roboczej. Te i podobne podejścia psychologiczne usiłujące skonkretyzować pojęcia języka potocznego wydają się atrakcyjne, pozwalając stworzyć pomost pomiędzy dynamicznymi modelami procesów zachodzących w mózgu, a treścią umysłu.

Świadomość wiąże się z subtelnymi pytaniami: dlaczego w ogóle mamy wrażenia świadome? Wrażenia mają jakości (zwane w filozofii umysłu *qualia*), które nie dają się opisać za pomocą bardziej podstawowych pojęć. Czy istnieniu tych jakości można przypisać jakąś funkcję? Czy mogą one powstać w elektronicznych symulatorach mózgu lub programach wykonywanych przez komputery? Jakie jest ewolucyjne znaczenie wrażeń? Czy świadomość refleksyjną należy uznać za inny rodzaj świadomości niż świadomość pierwotną? Powszechnie uważa się, że są to pytania, na które nie można znaleźć odpowiedzi. Można jednak znaleźć argumenty pokazujące, że system sztuczny, którego działanie wzorowane jest na modularnych sieciach neuronowych mózgu, musi twierdzić, że odczuwa wrażenia [9].

Najbardziej konkretne wydają się próby szukania neurofizjologicznych korelatów odróżniających stan świadomy od nieświadomego. Można znaleźć wiele kontrastujących ze sobą par zjawisk umysłowych, z których jedno kojarzone jest ze świadomym działaniem a drugie z nieświadomym. Dotyczy to różnych rodzajów pamięci: pamięć krótkotrwała – pamięć długoterminowa; epizodyczna – semantyczna; deklaratywna – proceduralna; świadome uczenie się i pamięć – uczenie się i pamięć utajona (*implicit learning, implicit memory*). Kontrolę i wnioskowanie świadome przeciwstawić można kontroli i automatycznemu działaniu, lub też bardziej ogólnie procesy wymagające umysłowego wysiłku procesom wykonywanym spontanicznie lub bez wysiłku. Świadoma percepcja kontrastuje z nieświadomą recepcją bodźców, a mechanizmy uwagi wyróżniają bodźce nowe, znaczące, silne lub znajdujące się w centrum uwagi, spychając do nieświadomej recepcji bodźce powtarzające się, nieistotne, słabe lub znajdujące się na peryferiach uwagi. Stan jawy i marzeń sennych przeciwstawić można stanowi głębokiego lub komie. Eksperymenty mogą pokazać różnice pomiędzy stanami mózgu w czasie świadomego działania i działania nieświadomego.

3 Neurofizjologiczne korelaty świadomości

Kontrasty pomiędzy czynnościami i stanami świadomymi a nieświadomymi można badać w testach psychologicznych. Jednakże przez długi czas trudno było znaleźć neurofizjologiczne korelaty wrażeń świadomych. Metody obrazowania mózgu pozwoliły co prawda dostrzec obszary zaangażowane w przetwarzanie informacji jednakże samo pobudzenie płatów czołowych w czasie rozwiązywania zadań dotyczących planowania nie jest precyzyjnie skorelowane z wrażeniami świadomymi. Rozdzielczość czasowa fMRI lub PET nie jest wystarczająca do zaobserwowania takich korelacji. Pamięć krótkotrwała jest bez wątpienia związana z bioelektrycznymi, szybkozmiennymi pobudzeniami przejawiającymi się w czynności EEG kory. Crick i Koch [4] zastanawiali się nad **problemem scalania** (*the binding problem*) aktywności różnych części układu wzrokowego [21] w procesie widzenia. Scalanie aktywności, konieczne do powstania wrażenia przechowywanego w pamięci roboczej, wymaga synchronizacji wyładowań neuronów. Wielu badaczy wsunęło hipotezę o istotnej roli oscylacji o częstotliwościach rzędu 40 Hz w tym procesie. Singer [16] pisze: „Możliwe, że tylko te pobudzenia przekraczają próg świadomości, które są dostatecznie zorganizowane i koherentne”. Synchronizacja powinna dotyczyć nie tylko intrakortykalnych pobudzeń, lecz również obejmować struktury limbiczne oraz twór siatkowaty.

Pomimo licznych eksperymentów z udziałem ludzi i zwierząt potwierdzających istnienie związku pomiędzy wrażeniami świadomymi i synchronizacją neuronów (por. przegląd w *Temporal Binding*, numer specjalny *Consciousness and Cognition* Vol. 8, No. 2, czerwiec 1999) hipoteza 40 Hz jest w dalszym ciągu kontrowersyjna. Przede wszystkim nie jest wcale jasne czy sam problem scalenia wrażeń wymaga rozwiązania, albowiem niektórzy filozofowie podważają przekonanie o konieczności traktowania wrażeń świadomych w całościowy sposób. Jeśli udałoby się wykazać, że jedność świadomości jest pewnego rodzaju iluzją i należy raczej mówić o strumieniu wrażeń interpretowanych jako całość problem scalenia okazałby się pozorny. Zeki i Bartels [21] przychylają się do pośredniego poglądu, w którym modułarna budowa mózgu pozwala na powstawanie synchronicznych stanów wewnątrz modułów, lecz nie w obrębie całego mózgu. Byłby to więc zbiór „mikroświadomości”, zamiast „teatru świadomości” [2]. Podobną teorię głosi A. Damasio [6]: grupy neuronów w znajdujących się w różnych obszarach mózgu – „strefach konwergencji”, czyli obszarach wielomodalnych, odbierających i wysyłających sygnały do wielu rejonów mózgu – mogą pobudzać obszary czuciowe i obszary odpowiedzialne za ekspresję emocji, rekonstruując stany mózgu podobne do tych, które wywołane zostały przez pierwotne bodźce. W tej teorii scalenie wrażeń nie wymaga synchronicznych oscylacji, chociaż powinny istnieć korelacje czasowe pomiędzy aktywnością różnych obszarów widoczne na obrazach fMRI i PET.

W niektórych badaniach nie zaobserwowano istnienia korelacji pomiędzy synchronizacją oscylacji 40 Hz a wrażeniami świadomymi, co może świadczyć o słuszności wspomnianych tu teorii alternatywnych. Sytuacja jest zapewne bardziej skomplikowana niż początkowo zakładano. Newman i Grace [11] zastanawiają się nad mechanizmem integracji i filtrowania różnych perceptów scalonych za pomocą synchronizacji rozproszonych grup neuronów w pamięci roboczej. Analizując eksperymentalne dane dochodzą oni do wniosku, że za synchronizację procesów zachodzących na najwyższych piętach przetwarzania, w obszarze dolno-skroniowym (IT) i przedczołowym, tworzących strumień świadomości, odpowiedzialne są procesy selekcji (*gating*) dokonywane przez ośrodki układu limbicznego, brzuszłą część prążkowiec (a zwłaszcza jądro półleżące) i jądro siatkowate wzgórza (*nucleus reticularis thalami*, NRT). Jest to rozwinięcie zaproponowanego wcześniej „rozszerzonego siatkowo-

wzgórzowego schematu aktywacji” (Extended Reticulo-Thalamic Activation System, ERTAS), w obrębie którego powstawać miały pobudzenia odpowiedzialne za świadomość (szczegóły neuroanatomiczne opisane są w pracach [11],[17],[19],[20]).

William James sądził, że świadomość jest wynikiem działania całego mózgu, gdyż niewiele uszkodzeń mózgu prowadzi do jej całkowitej utraty. W badaniach wykorzystujących rywalizację obuoczną (*binocular rivalry*) u małp wykazano brak korelacji pomiędzy wrażeniami świadomymi a aktywnością większości neuronów niższych pięter kory wzrokowej (V1, V4, MT) w czasie gdy aktywność 95% neuronów w wyższych piętrach (IT, STS) jest z nimi skorelowana. Podobne wyniki otrzymano za pomocą metod magnetoencefalograficznych (MEG) u ludzi. Wydaje się rzeczą oczywistą, że aktywność tylko tych obszarów mózgu, w których istnieją stabilne, niezmiennicze reprezentacje perceptów, może wykazywać korelacje z świadomymi odczuciami. Oko przemieszczając się ruchami sakkadycznymi wysyła z siatkówki do niższych pięter układu wzrokowego sygnały zbyt chaotyczne, by mogły one bezpośrednio być uświadamiane. Obszary kory związane ze szlakiem grzbietowym układu wzrokowego działają w znacznej mierze niezależnie od świadomości, pozwalając na przykład na szybką, automatyczną reakcję na wykrywanie ruchu. Pobudzenia szlaku brzuszego, prowadzącego do płatów skroniowych, prowadzą bezpośrednio do wrażeń wzrokowych.

Wiele innych procesów automatycznych pobudza części układu wzrokowego i słuchowego wpływając jedynie w pośredni sposób na treść wrażeń świadomych. Oczywiście wszystkie obszary mózgu są ze sobą sprzężone i w dłuższym okresie czasu mogą na siebie oddziaływać, ale nie wszystkie mogą to zrobić w krótkim okienku czasowym potrzebnym na powstanie świadomego wrażenia (rzędu ułamków sekund). Tylko najważniejsze informacje dają bezpośredni wkład do pamięci roboczej, scalającej stan tych obszarów mózgu, które w istotny sposób wpływają na kontrolę zachowania. Habitacja powoduje zanikanie skorelowanej z danym bodźcem aktywności poza obszarami sensorycznymi. Długotrwała pamięć epizodyczna zapisuje przede wszystkim stan pamięci roboczej, pozwalając go odtworzyć w późniejszym okresie i umożliwiając użyteczne skojarzenia. Jedną z funkcji świadomości jest więc rolę filtra, pozwalającego zapamiętać z morza kombinacji docierających do mózgu sygnałów tylko najważniejsze.

Chociaż nie ma sensu mówić o „neuronach świadomości” nie wszystkie obszary mózgu w jednakowym stopniu mają bezpośredni wpływ na treść pamięci roboczej, a więc na treść świadomości. Tajemnica świadomości z pewnością wydawała by się mniejsza gdyby udało się znaleźć takie cechy w możliwych do zaobserwowania sygnałach aktywności mózgu, które jednoznacznie korelują się z subiektywnymi wrażeniami. Znaczny postęp w tym kierunku stanowi praca Tononi i Edelmana [19], podkreślająca dwie sprzeczne ze sobą cechy świadomej aktywności: scalenia różnych aspektów doświadczanych wrażeń i jednoczesnego dużego zróżnicowania, pozwalającego na szybkie zmiany przeżywanych, złożonych epizodów. Scalenie aktywności grup neuronów można zmierzyć wprowadzając miarę funkcjonalnego skupienia aktywności F_A , opartą na informacji wzajemnej. W określonym okienku czasowym grupa funkcjonalna powinna charakteryzować się większym indeksem F_A dla elementów należących do tej grupy niż dla pozostałych elementów.

Niech X oznacza zbiór elementów neuronowych (np. minikolumn); wybierzmy z niego j -ty podzbiór X_j^k o k elementach (z k elementów utworzyć można wiele kombinacji) i zbadajmy na ile aktywność tego podzbioru zależy od pozostałych neuronów $X - X_j^k$. Wprowadzając miarę ilości informacji (entropii) zbioru $H(X)$ można określić informację wzajemną $H_I(X_j^k; X - X_j^k) =$

$H(X_j^k) + H(X - X_j^k) - H(X)$, która równa jest sumie ilości informacji zawartej w podzbiorze X_j^k i jego dopełnieniu $X - X_j^k$ minus ilość informacji w całym zbiorze X . Za miarę informacji wynikającej z integracji podzbioru X_j^k możemy przyjąć różnicę między sumą informacji wszystkich elementów podzbioru traktowanych jako niezależne, czyli $I(X_j^k) = \sum_i H(x_i \in X_j^k) - H(X_j^k)$. Iloraz $F_A(X_j^k) = I(X_j^k) / H_I(X_j^k; X - X_j^k)$ określa miarę skupienia aktywności a jego maksymalizacja pozwala tak dobrać podzbiór elementów by informacja wewnętrzna była jak największa przy najmniejszej informacji wzajemnej. Kolejny krok to wprowadzenie miary złożoności określającej zdolność systemu do przyjmowania stanów, które (z jego własnego punktu widzenia) są istotnie różne. Jeśli uznać, że system może obserwować sam siebie to dzieląc go na X_j^k i $X - X_j^k$ możemy zmierzyć informację wzajemną $H_I(X_j^k; X - X_j^k)$, która jest duża jeśli obydwie części mają dużą pojemność informacyjną w stosunku do całego zbioru. Tonioni i Edelman [19] proponują przyjąć jako miarę złożoności neuronalnej zbioru $C_N(X)$ sumę po wszystkich k -elementowych podzbiórach uśrednioną po możliwych podzbiórach, czyli:

$$C_N(X) = \sum_k \langle H_I(X_j^k; X - X_j^k) \rangle_j$$

Duża złożoność oznacza współlistnienie specjalizacji funkcjonalnej elementów należących do aktywnych skupień, przyjmujących liczne, wzajemnie od siebie zależne stany. Zastosowanie tej miary do symulowanych sieci złożonych z tysięcy neuronów daje najbardziej interesujące zachowania sieci i duże współczynniki $C_N(X)$ dla „naturalnej” topografii połączeń, a niewielkie dla innych topografii. Gęste przypadkowe połączenia dają hipersynchroniczne pobudzenia o niewielkim zróżnicowaniu a rzadkie połączenia dają chaotyczne, zróżnicowane, ale niezależne od siebie pobudzenia, a więc i złożoność neuronalna jest niewielka. Tonioni i Edelman [19] mają nadzieję, że zastosowanie takich miar informacyjnych pomoże w identyfikacji struktur mózgu biorących w istotny sposób udział w generacji stanów świadomych.

4 Teoretyczne modele świadomości

John Taylor postawił sobie bardzo ambitne zadanie wyjaśnienia, w oparciu o modele neuronowe, własności charakteryzujących przeżycia świadome [17],[18]. *Qualia*, monadyczne własności wrażeń, nieredukowalne do własności prostszych, znane tylko z introspekcji. Problemem jest tu oczywiście ustalenie fenomenologii wrażeń, gdyż naiwne twierdzenie, że przeżycia znane są bezpośrednio a nie interpretowane w świetle własnej teorii umysłu, nie jest prawdziwe. Patrzymy na świat przez swoje wrażenia, a nie „doświadczamy wrażeń”. Subiektywna treść doświadczenia wewnętrznego jawi się nam w intymny, bezpośredni sposób, nie możemy się oderwać, spojrzeć na treść umysłu z zewnątrz tak, by zobaczyć jakie mechanizmy stoją za sceną „teatru świadomości”. Monadyczność wrażeń, niemożliwość ich dekompozycji, powoduje, że trudno jest je odnieść do stanu neuronów, gdyż nie wiadomo, jakie efekty mają powodować te stany. Na poziomie neuronowym możemy mówić o funkcjach odpowiadających przetwarzaniu informacji, ale tu nie chodzi o przetwarzanie informacji. Wrażenia mają swój wymiar czasowy, są obecne, trwają, mają okres latencji (powstanie wrażenia wymaga kilkuset milisekund) i niespostrzeżenie przechodzą w inne. Trwanie może obejmować kilka sekund, w których następuje integracja wrażeń, np. potrzebna do zrozumienia sensu zdania. Większość stanów psychicznych jest intencjonalna, wskazuje na coś w świecie. Inne aspekty to tendencja do postrzegania całości (gestalt), rozróżnienie wrażeń centralnych i peryferyjnych. Doświadczenia świadome przypisywane są i odnoszone do centrum, „ja”, ale jest to własność świadomości refleksyjnej, której mogą nie posiadać zwierzęta.

Własności przeżyć świadomych mogą ulec radykalnej zmianie w przypadku uszkodzeń lub chorób mózgu. Analizując te własności Taylor [17] wyprowadza z nich dość oczywiste kryte-

ria, jakie powinny spełniać modele neuronowe. Np. „patrzenie przez wrażenia” związane jest z efektywnością działania systemu łączącego moduły dostarczające dane związane z wrażeniami do modułów na wyższych piętrach, w tym modułów odpowiedzialnych za reprezentację perspektywy egocentrycznej. Następnie rozważa przestrzenne, czasowe i emergentne własności konieczne do powstania przeżyć wewnętrznych, podając rodzaje sieci neuronowych, jakie mogą realizować tego typu funkcje. Własności przestrzenne, lub ogólniej topograficzne, zrealizować jest najłatwiej za pomocą samoorganizujących się map neuronów. Własności temporalne związane z trwaniem, latencją i przejściami pomiędzy wrażeniami zrealizować można za pomocą lokalnych rekurencyjnych sieci konkurujących ze sobą neuronów z odpowiednią dynamiką, w których przejścia między stanami atraktorowymi są dostatecznie szybkie w porównaniu z czasem przebywania w poszczególnych stanach. Najciekawsze są własności emergentne – tu autor proponuje „bąble aktywności”, odpowiadające wrażeniom powstającym na najwyższym poziomie hierarchii modułów systemu. Takie bąble tworzą się w modelach traktujących tkankę neuronową jako ciągłą dwuwymiarową warstwę. Są to zlokalizowane pobudzenia, trwające przez jakiś (skończony) czas nawet po zniknięciu sygnałów, które je wywołały.

W sieciach atraktorowych pobudzenia neuronów szybko prowadzą do metastabilnych stanów, które zacierają ślady swojego pochodzenia, gdyż są takie same, niezależnie od początkowego punktu wyjścia w obrębie basenu atrakcji. Ciąg świadomych wrażeń pojawia się kolejno, bez możliwości prześledzenia, skąd się one biorą. Taylor sądzi [18], że kluczową rolę w kontrowaniu konkurencji pomiędzy różnymi fragmentami reprezentacji, składającymi się na całościowe wrażenia, pełni jądro siatkowate (NRT), trójwarstwowa sieć hamujących neuronów znajdujących się w części brzusznej wzgórza. NRT ma obustronne połączenia ze wszystkimi obszarami kory oraz pozostałymi częściami wzgórza. Symulacje komputerowe pokazały, że istotnie przejścia pomiędzy stanami atraktorowymi w układzie NRT-kora-wzgórza mają odpowiednie własności. Jeśli dodać do tego model kory w postaci ciągłego pola neuronalnego (*Continuum Neural Field Theory*, CNFT), w którym powstają bąble aktywności, łącząc lokalne moduły CNFT w warstwy można zaobserwować szereg własności, które zgodne są z kryteriami dla systemu realizującego świadome przeżycia. Bąble w kolejnych warstwach mogą się stawać coraz dłuższe, zacierając szczegóły struktury na niższych poziomach. Pobudzenie krótkim impulsem wzrokowym może wywołać aktywność w obszarze V1 o charakterystycznym czasie życia rzędu 200 ms, wydłużającą się w wyższych piętrach układu wzrokowego wielokrotnie. Może to być źródłem złudzeń wzrokowych, np. ruchu pozornego. Bąble aktywności mogą się przesuwac pod wpływem nowych bodźców a ruch tych bąbli może się przez jakiś czas utrzymywać już po ustaniu bodźców. Jedynie ta aktywność, która trwa dostatecznie długo może stać się źródłem świadomych wrażeń.

Takie teorie zbliżają nas do szczegółowych modeli dynamiki mózgu o własnościach, które można ściśle powiązać z własnościami przeżywanymi stanów psychicznych. Pozostaje jednak wiele pytań. Aktywność których obszarów mózgu odpowiada bezpośrednio za treść świadomości? Taylor proponuje tu (informacja prywatna) głównie boczne części płatów ciemieniowych. Propozycja ta wymaga eksperymentalnego zbadania, np. przez zrobienie mapy czasów życia bąbli aktywności w różnych częściach kory za pomocą MEG lub EEG. Program badania czasów aktywności można realizować zarówno dla bodźców podprogowych jak i uświadomianych. Specyficzna treść wrażeń, takich jak doświadczenie czerwieni zachodzącego słońca, lub smaku czekolady, jest wynikiem wzajemnych relacji powstających stanów i stanów zapamiętanych. Dzięki tym relacjom powstają skojarzenia i interpretacje. Taylor nie twierdzi, że jego teoria bąbli przekracza przepaść pomiędzy psychiką a fizjologią, nie odpo-

wiada bowiem na fundamentalne pytanie „dlaczego wrażenia są w ogóle odczuwane”, uważa jednak, że jest to pierwszy rzut oka na drugą stronę tej przepaści (*explanatory gap*).

Dla Stephena Grossberga problem świadomości jest częścią ogólniejszej teorii związanej z uczeniem się i uwagą [11]. Podstawowa hipoteza brzmi: wszystkie stany świadome są stanami rezonansowymi. Z tego powodu wiele procesów w mózgu, wykorzystujących uczenie konkurencyjne lub porównywanie z prototypami, to procesy nieświadome. Rozwijana przez Grossberga przez ostatnie 30 lat teoria adaptacyjnego rezonansu (*Adaptive Resonance Theory*, ART) znajduje licznie zastosowania wyjaśniając np. iluzje wzrokowe, akustyczne, tworzenie kategorii czy integrację wrażeń w okienkach czasowych. Np. słysząc zdanie: „It was found that the ...eel is on the ...” (stwierdzono, że ...eel jest na ...), gdzie ... zastąpione jest szumem, świadome wrażenie związane z niepełnym słowem ...eel zmienia się pod wpływem nadchodzącego później słowa końcowego! Słyszymy zdania: „peel is on the orange” (skórka jest na pomarańczy), „wheel is on the wagon” (koło jest na wozie), „heel is on the shoe” (obcas jest na bucie) lub „meal is on the table” (posiłek jest na stole). Wrażenia nie powstają więc bezpośrednio na podstawie sygnałów zmysłowych lecz są wynikiem odwoływania się do już istniejącej wiedzy, przechowywanej w pamięci długotrwałej. W ciągu sekundy pobudzone są reprezentacje tych pojęć, które na podstawie kontekstu mogą nadać pełniejszą interpretację reprezentacjom w pamięci roboczej, po czym zawartość pamięci roboczej ulega zmianie.

Można na ten proces spojrzeć jak na dopasowanie dochodzącej informacji do z góry ustalonych oczekiwań (dzięki wiedzy zdobytej na podstawie wcześniejszych doświadczeń). Dzięki oczekiwaniom z szumu wybierane są odpowiednie formanty (jest to konkluzja eksperymentów psychofizycznych). Ten wzajemny wpływ oczekiwań i istniejącej w danym momencie w pamięci roboczej interpretacji wytwarza stan rezonansowy, przedłużający i nasilający pobudzenia, łączący w tym przypadku fonetykę z semantyką zdania.

ART stworzono po to, by rozwiązać problem stabilności-plastyczności w nabywaniu nowej wiedzy. Stabilność jest konieczna by nie zapomnieć użytecznych informacji sprzecznych z nowo napływającymi; plastyczność konieczna jest by nauczyć się nowych informacji. Modele ART starają się dobudować nową wiedzę w kontekście już znanej, tworząc stany rezonansowe. Stabilność uczenia implikuje jego intencjonalność, albowiem jest ono oparte na oczekiwaniach wobec świata dostarczającego systemowi danych. Oczekiwania są również podstawą mechanizmu uwagi, skupiania się systemu na „interesujących” go danych. Pobudzenie rezonansowych stanów, nawet jeśli nie będzie na tyle silne by pojawić się w pamięci roboczej, może utorować drogę do późniejszego wyboru oczekiwanych odpowiedzi (*priming*). Procesy rezonansowe zachodzą w układzie wzrokowym na różnych poziomach, łącznie z układem orientacji gałek ocznych, w układzie słuchowym, prawdopodobnie są również podstawą procesów czuciowych, procesów warunkowania klasycznego i interakcji kognitywno-afektywnych. Można je nazwać „dopasowaniem pobudzającym”, podczas gdy procesy motoryczne jak i procesy związane z ustalaniem położenia w przestrzeni (wzrokowy szlak grzbietowy) są komplementarnymi procesami dopasowania hamującego. Kompetencje motoryczne i mapy przestrzenne mogą być doskonalone w ciągły sposób, nie ma tu niebezpieczeństwa zniszczenia stabilności. Ponieważ nie ma tu stanów rezonansowych pamięć proceduralna nie prowadzi do wrażeń świadomych. W przypadku amnezji następczej pamięć proceduralna nie jest zaburzona, nabywane są nowe umiejętności ale nie można sobie przypomnieć związanych z tym wrażeń. Zaburzenia rozpoznawania obiektów i niezdolność do nauczenia się rozpoznawania nowych obiektów nie oznacza braku zdolności do uczenia się nowych sposobów manipulacji tymi obiektami, ale te umiejętności nie są związane z świadomymi wrażeniami.

5 Modele dynamiczne i neurofizjologia

Dlaczego stany mózgu miałyby prowadzić do powstania świadomości? Dotychczas rozwinięte teorie, chociaż potrafią częściowo odpowiedzieć na pytania „które stany mózgu mogą być świadome”, nie próbują odpowiedzieć bezpośrednio na to pytanie. Filozofowie zajmujący się umysłem wyobrażają sobie, że mogą istnieć zombi, istoty zachowujące się tak jak ludzie ale niczego świadomie nie odczuwające [2]. Przytacza się tu przykład ślepoty korowej (*blind-sight*), w której szczątkowe widzenie pozostaje pomimo braku wrażeń wzrokowych. Dlaczego to całe przetwarzanie informacji przez mózg nie zachodzi w ciemności, bez wrażeń?

Rozważmy typowe zadanie, przed jakim stoi ssak, np. szczur, odżywiający się różnorodnym pokarmem. Smak pożywienia nie jest wielkością dyskretną, nie można go opisać symbolicznie, a zbiór pobudzeń czopków smakowych nie jest inwariantną wielkością, która może stać się bezpośrednio świadomym wrażeniem. Szczur musi porównać nowy smak 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. Musi więc być utrzymana w pamięci roboczej dostatecznie długo, by pobudzić wiele obszarów mózgu i by nastąpiły procesy skojarzeniowe (interferencja reprezentacji). Mamy więc do czynienia z ciągłą, niewerbalną aktualizacją reprezentacji wrażeń smakowych w pamięci roboczej, powstawaniem stanów rezonansowych. Ciągła aktualizacja konieczna jest ze względu na szybki zanik informacji w pamięci roboczej i jej niewielką pojemność. Jeśli nie pojawią się żadne negatywne skojarzenia szczur zacznie jeść, w przeciwnym przypadku kolejnym stanem jego mózgu będzie pobudzenie ośrodków strachu lub wstępu i porzucenie jedzenia. Najważniejszym procesem jest więc dyskryminacja reprezentacji wrażeń smakowych, od której zależy zachowanie. Jak to wygląda z punktu widzenia mózgu szczura?

Szczur wie jak postąpić, a jego wiedza ma charakter wrażeń wynikających z działania mechanizmów poznawczych. Gdyby jego mózg posiadał zdolność werbalnego komentowania stanów pamięci roboczej, tak jak posiada ją mózg ludzki, z pewnością stwierdziłby, że odczuwa smak, widzi przedmioty i ma z tym skojarzone wrażenia, gdyż reprezentacje wrażeń są fizycznymi stanami jego pamięci roboczej. Każdy model mózgu działający na podobnej zasadzie, jeśli będzie dostatecznie złożony, by móc komentować pojawienie się takich reprezentacji wrażeń, musi twierdzić, że jest świadomy swoich wrażeń. Własności tych wrażeń będą miały cechy podobne do opisanych przez Taylora [18]. Modele mózgu autonomicznego zwierzęcia muszą zawierać jakąś reprezentację siebie, odgraniczającą „ja” od „nie-ja”, pozwalającą odróżnić „siebie” od reszty świata. Nasze poczucie tożsamości związane jest silnie z propriocepcją, a rozwój mózgu związany był przede wszystkim z koniecznością powstania wewnętrznego modelu ciała w celu przewidywania wyników ruchu kończyn. W systemie pojawi się więc perspektywa „moje wrażenia”, wynikająca z utrzymywania się stanów rezonansowych i możliwości komentowania stanu pamięci roboczej.

Świadomość w tym ujęciu nie jest więc niczym innym jak procesem dyskryminacji ciągłych reprezentacji stanów pamięci roboczej, stanów tworzących „przestrzeń wewnętrzną” dla subiektywności. Wszelkie zaburzenia świadomości muszą być związane z upośledzeniami funkcji kognitywnych i odwrotnie, zaburzenia tworzenia naturalnych reprezentacji muszą pogarszać odpowiednie kompetencje poznawcze. Ślepota korowa (wynikająca z uszkodzenia kory wzrokowej) prowadzi do upośledzenia funkcji wzrokowych, pomimo dostępności in-

formacji z nerwu wzrokowego w wielu strukturach mózgu. Informacja ta ma jednak odmienny wkład do dynamiki mózgu odpowiedzialnej za pamięć roboczą niż zwykle wrażenia wzrokowe. Dyskryminacja nowych 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 nigdy nie będzie zbyt wysoki, gdyż do precyzyjnej analizy sygnałów z nerwu wzrokowego brakuje wyspecjalizowanych struktur. Takie rozumienie przyczyn pojawiania się wrażeń świadomych prowadzi do wielu wniosków i przewidywań eksperymentalnych dotyczących fenomenologii wrażeń (Duch, w przygotowaniu).

W jaki sposób można modelować tak skomplikowane zagadnienia? Jeden z pomysłów [8] polega na próbie zdefiniowania serii przybliżeń, od poziomu molekularnego, przez pojedyncze neurony, grup neuronów, sieci neuronowych z rekurencją, aż do skrajnego uproszczenia opisu w postaci probabilistycznych lub deterministycznych reguł zachowania. Na każdym poziomie możliwe jest badanie innych form zachowania i innych zjawisk. Przejścia do opisu działania umysłu na poziomie psychologicznym można dokonać wprowadzając model pamięci roboczej oparty o popularne w psychologii przestrzenie psychologiczne. Strumień świadomości można wówczas modelować jako zdarzenia w przestrzeniach psychologicznych – chociaż są one cieniem neurodynamiki to traktować je można jako niezależny, emergentny poziom stanowiący pomost pomiędzy neurofizjologią a psychologią.

6 Dyskusja

Przedstawiony tu krótki przegląd problemów i teorii związanych ze świadomością pokazuje, że jesteśmy na dobrej drodze i potrafimy zadawać coraz bardziej precyzyjne pytania, na które odpowiedzi dają eksperymenty i symulacje komputerowe. Pomimo ogromnej złożoności mózgu symulacje komputerowe mogą uchwycić niektóre mechanizmy przyczynowe, pozwalając nie tylko na odtworzenie funkcji struktur mózgu, lecz również zrozumieć związki pomiędzy specyficznymi uszkodzeniami neuroanatomicznymi, zmianami neuropatologicznymi i wynikającymi stąd symptomami. Szczegółowe modele oparte na neurofizjologicznie poprawnych, impulsujących neuronach, wykazują jakościowe zachowania, które można odtworzyć w modelach uproszczonych [10],[13]. Modele komputerowe mogą w istotny sposób przyczynić się do zrozumienia zaburzeń pamięci, sposobów używania i rozumienia mowy, zaburzeń neurologicznych i chorób psychicznych. Często jest to rozumienie jakościowe, metaforyczne, posługujące się luźnymi analogiami ale i to stanowi spory postęp.

Czym różni się model komputerowy działania mózgu od swojego biologicznego pierwowzoru? Wrażenia, myśli, percepty i inne obiekty umysłu są fizycznymi stanami bioelektrycznych pobudzeń istniejących w określonym momencie czasu. Pamięć robocza zawiera w sobie wkład od procesów peryferyjnych, interferencję (rezonans) z podobnymi stanami przeżywanymi w przeszłości. Postrzeganie lub przypominanie jest aktualizacją potencjalnie dostępnych stanów mózgu w ulotnej formie bioelektrycznych pobudzeń. W komputerze reprezentacje symboliczne są aktualizowane przez przesyłanie ich do rejestrów procesora, jednakże są to reprezentacje martwe, gdyż binarne stany rejestrów nie zawierają w sobie żadnego bogactwa asocjacji. Dynamicznych procesów tworzących „ubrane” reprezentacje powstające w pamięci roboczej nie da się symulować za pomocą cyfrowego komputera. Wrażenia świadome są rzeczywiste, gdyż są realnymi stanami fizycznymi, a nie tylko stanami przetwarzania informacji.

Symulacje umysłu metodami sztucznej inteligencji nie doprowadzą więc do świadomych komputerów, ale nie oznacza to, że istnieją jakieś ograniczenia związane ze sprawnością automatyzacji procesów myślenia. Wbrew powszechnemu mniemaniu nie ma dowodów na to, że świadomość jest aktywnie zaangażowana w rozwiązywanie problemów [21]. Przegrana Kasparowa z systemem komputerowym *Deep Blue* w 1997 roku pokazuje, że maszyna, której szybkość przetwarzania informacji jak i pojemność pamięci nie przekracza promila możliwości ludzkiego mózgu może, w dobrze określonej dziedzinie wymagającej sprawności intelektualnej, osiągnąć lepsze rezultaty od człowieka. Do roku 2005 IBM zamierza zbudować komputer *Blue Gene* o możliwościach porównywalnych z ludzkim mózgiem. Kompetencje systemów ekspertowych realizowanych na takich komputerach mogą znacznie przekroczyć ludzkie możliwości. Z technicznego punktu widzenia nie ma więc potrzeby budowy świadomych maszyn. Konstrukcja takich maszyn musiałaby zawierać jako kluczowy element analogową pamięć roboczą, zdolną do przechowywania reprezentacji interferujących z reprezentacjami przechowywanymi w pamięci długotrwałej.

Literatura

- [1] Baddeley A.D, *Consciousness and working memory*. *Consciousness and Cognition* 1 (1992) 3-6
- [2] Baars B.J (1996) *In the theater of consciousness: the workspace of the mind*. (Oxford University Press, London)
- [3] Churchland P.S, *Neurophilosophy. Toward a unified science of the mind/brain*. MIT Press, Bradford Book, 1986 (7-me wydanie 1993)
- [4] Crick F, Koch C, *Towards a neurobiological theory of consciousness*. *Seminars in the Neurosciences* 2 (1990) 263-275
- [5] Crick F, *The Astonishing hypothesis. The scientific search for the soul*. Charles Scribner's sons, New York, 1994 (tłumaczenie: *Zdumiewająca hipoteza*, Prószyński i S-ka, Warszawa 1997)
- [6] Damasio A. R. (1990). *Synchronous activation in multiple cortical regions: A mechanism for recall*. *Seminars in Neurosciences* 2, 287–296.
- [7] Dennett, D.C. (1991) *Consciousness explained*. Boston: Little, Brown.
- [8] Duch W. (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
- [9] Duch W, *Conscious mind as a limit of brain-like computing*. *Minds and Machines* (wyślane, 1999)
- [10] Duch W (2000) *Sieci neuronowe w modelowaniu zaburzeń neuropsychologicznych i chorób psychicznych*. W: *Biocybernetyka 2000, Tom 6: Sieci neuronowe* (red. W. Duch, J. Korbicz, L. Rutkowski i R. Tadeusiewicz), rozdz. II.18
- [11] Grossberg S. (1999) *The Link between Brain Learning, Attention, and Consciousness*. *Consciousness and Cognition* 8, 1–44
- [12] Newman J, Grace A.A. (1999) *Binding across Time: The Selective Gating of Frontal and Hippocampal Systems Modulating Working Memory and Attentional States*. *Consciousness and Cognition* 8, 196-212

- [13] Parks R.W, Levine D.S, Long D, red. (1998): *Fundamentals of Neural Network Modeling*. (MIT Press, Cambridge, MA)
- [14] Penrose R. *Nowy umysł cesarza*. (Oxford University Press 1994)
- [15] Searle J.R., *Umysł na nowo odkryty*. PIW, Warszawa 1999.
- [16] Singer, W. (1994) *The organization of sensory motor representations in the neocortex: A hypothesis based on temporal coding*. W: Umiltà C, Moscovitch M. (eds.), *Attention and performance XV: Conscious and nonconscious information processing*, str. 77–107. Cambridge, MA: MIT Press.
- [17] Taylor J.G. (1998) *Cortical Activity and the Explanatory Gap*. *Consciousness and Cognition* 7, 109-148
- [18] Taylor J.G. (1999) *The race for consciousness*. MIT Press, Cambridge, MA
- [19] Tononi G, Edelman G. (1998) *Consciousness and complexity*. *Science* 282, 1846-1851
- [20] Treisman, A. (1996) *The binding problem*. *Current Opinion in Neurobiology* 6, 171–178
- [21] Wróbel A. (1997) *W poszukiwaniu integracyjnych mechanizmów działania mózgu*. W: *Mózgi i zachowanie*, PWN Warszawa (red. T. Górska, A. Grabowska, J. Zagordzka)
- [22] Velmans M. (1991) *Is human information processing conscious?*, *Brain and Behavioral Science* 14, 651-726.
- [23] Zeki S., Bartels A. (1999) *Toward a Theory of Visual Consciousness*. *Consciousness and Cognition* 8, 225-259