
Geometryczny model umysłu.

Włodzisław Duch,

Katedra Informatyki Stosowanej, Uniwersytet Mikołaja Kopernika

Wielkim wyzwaniem dla nauki jest stworzenie takiego modelu świata, który da się zrozumieć. Istotną częścią tego świata są nasze umysły. Po trwającej tysiące lat dyskusji filozoficznej natura umysłów staje się coraz bardziej rozumiana dzięki postępom neuronauk (*neurosciences*), czyli różnych gałęzi nauki badających budowę i aktywność neuronów mózgu. Mózg ludzki jest najbardziej skomplikowanym obiektem w znanym wszechświecie a umysł najbardziej tajemniczym. Przez „umysł” rozumiemy kompleks zdolności umożliwiających powstanie subiektywnego obrazu świata dzięki procesom poznawczym i aktywnej eksploracji. Subiektywny obraz świata obejmuje świat fizyczny, relacje społeczne oraz wyobrażenia dotyczące samego siebie. Poznawanie obejmuje postrzeganie, pamiętanie, myślenie i ocenianie. Aktywna eksploracja, działanie, związane jest z planowaniem, aktami podejmowanie decyzji, kontrolą zachowania i uwagą. Specyficzną własnością umysłu jest świadomość, zdolność do „zdawania sobie sprawy” z własnych procesów poznawczych.

Czy fizyka wystarczy by zrozumieć działanie mózgu i umysłu? Nie jest to przekonanie wśród fizyków powszechne [1]. Podobne wątpliwości na początku XX wieku dotyczyły natury życia [2]. Introspekcja przekonuje nas, że zjawiska mentalne i procesy zachodzące w mózgu mają całkiem odmienną naturę. Jak należy rozumieć relacje pomiędzy mózgiem a umysłem i jak przejść od badania zachodzących w mózgu procesów do opisu umysłu? Badanie mózgu na poziomie molekularnym (neurochemii) lub badanie własności pojedynczych neuronów niewiele nam tu pomogą, podobnie jak pełna znajomość oddziaływań na poziomie cząstek elementarnych nie wystarczy do zrozumienia aktywności biologicznej białek.

Kilka lat temu zaproponowałem [3] by traktować procesy mentalne jako rezultat neurodynamiki na poziomie globalnej aktywności bioelektrycznej mózgu. Otwiera to drogę do tworzenia modeli na poziomie pośrednim pomiędzy działaniem mózgu i umysłu, modeli operujących językiem redukowalnym do zjawisk neurofizjologicznych z jednej strony i rozszerzalnym do pojęć psychologicznych z drugiej. Postaram się poniżej uzasadnić swój optymizm i przedstawić podstawowe idee tego modelu. Zanim jednak do tego dojdę konieczne będzie krótkie wprowadzenie wyjaśniające relacje pomiędzy mózgiem i umysłem.

1. Relacja umysł - mózg

Mózg jest substratem, w którym zachodzą procesy konieczne do powstawania procesów mentalnych. Neuropsychologia [4] bada związki pomiędzy zaburzeniami funkcji psychicznych a uszkodzeniami różnych obszarów mózgu i połączeń pomiędzy tymi obszarami. Związki te można często zrozumieć jako specyficzne zaburzenia przetwarzania informacji, prowadzące do zaskakujących zmian w sposobie działania umysłu (np. dysocjacji pomiędzy rozpoznawaniem afektywnym i kognitywnym). Modele neuronowe chorób psychicznych i syndromów neuropsychologicznych pozwalają zrozumieć przynajmniej w jakościowy sposób przyczyny patologii i normalnego funkcjonowania mózgu [5]. Mamy obecnie bardzo bogaty materiał doświadczalny wymagający teoretycznego uporządkowania.

Struktura mózgu jest wynikiem milionów lat rozwoju, adaptacji ewolucyjnych umożliwiających sprawne działanie pozwalające na przetrwanie gatunku w zmiennych, niekorzystnych warunkach. Procesy zachodzące w mózgu są warunkiem koniecznym istnienia umysłu, ale

samo badanie tych procesów nie pozwoli w pełni zrozumieć indywidualnego umysłu. Potrzeby organizmu i jego możliwości poznawcze, znajdujące odbicie w strukturze mózgu, stwarzają ramy dla powstania subiektywnego obrazu świata. Struktura umysłu jest rezultatem niepowtarzalnej historii jednostki. Procesy neurofizjologiczne odpowiedzialne za zdarzenia mentalne mają rację bytu tylko ze względu na istnienie umysłu. Pomiędzy procesami mentalnymi i stanami mózgu istnieje ścisła odpowiedniość. Np. rozpoznanie jakiejś melodii wywołuje szereg skojarzeń, stanów mentalnych realizowanych przez różne pobudzenia mózgu. Zależą one od kultury, w której wychowała się dana osoba, od jej przeszłych przeżyć. Relacje pomiędzy następującymi po sobie stanami mentalnymi mają sens tylko na poziomie umysłu. Zachodzące w mózgu procesy neurofizjologiczne umożliwiają powstanie stanów mentalnych, ale ich nie wyjaśniają. Dlatego nie można w pełni wyjaśnić stanów umysłu sprowadzając je do stanów mózgu. Modele umysłu wymagają innego poziomu opisu niż modele mózgu.

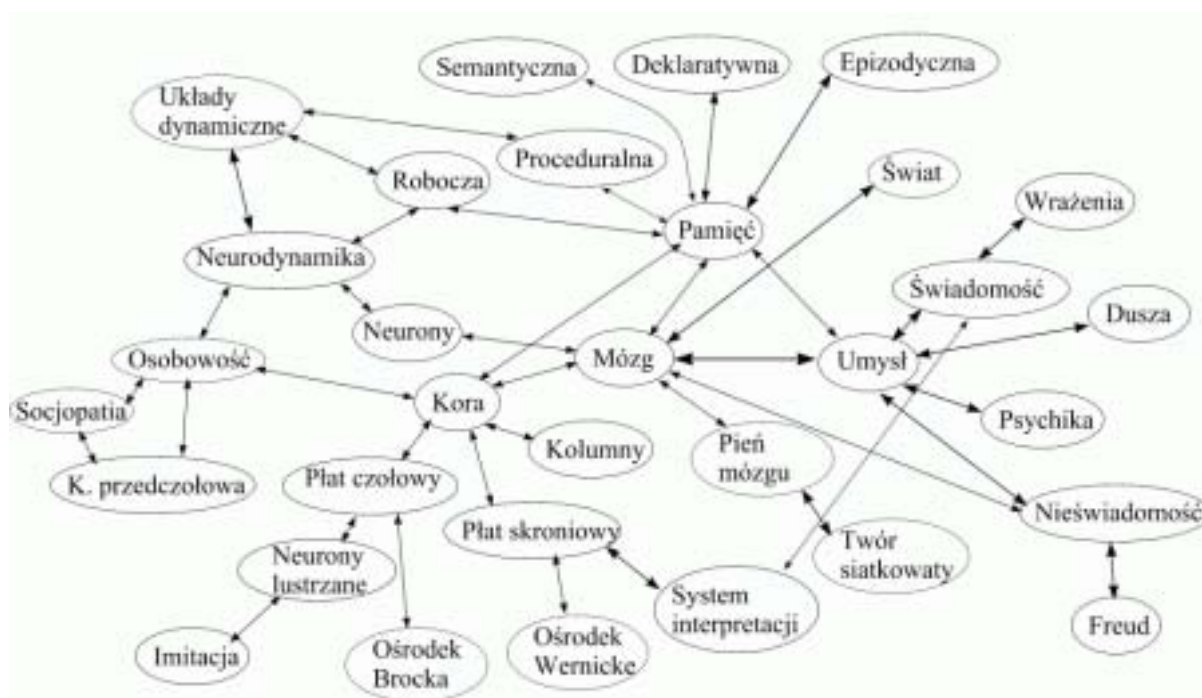
Procesy neurofizjologiczne odpowiadające zdarzeniom mentalnym muszą charakteryzować się odmiennymi prawami i własnościami niż te, które dotyczą zachowania się pojedynczych neuronów. Niezwykle złożone oddziaływania ogromnej liczby neuronów tworzą nowy, emergentny poziom organizacji. Wzajemne relacje stanów neurofizjologicznych mózgu i wzajemne relacje stanów mentalnych mają analogiczną strukturę, nie ma natomiast mowy o bezpośrednim podobieństwie. Wewnętrzną dynamikę stanów mózgu w przypadku swobodnym (bez bodźców zewnętrznych) scharakteryzować można za pomocą prawdopodobieństw przejść pomiędzy stanami atraktorowymi neurodynamiki [5]. Tylko takie stany istnieją dostatecznie długo by mogły być związane z treścią zdarzeń mentalnych. Odpowiadające im stany umysłu składają się z wspomnień, epizodów, myśli, wyobrażeń. Elementy te będę nazywał „obiettami umysłu”. Prawdopodobieństwa przejść pomiędzy potencjalnie możliwymi stanami umysłu powinny być takie same jak pomiędzy odpowiadającymi im stanami mózgu.

Umysł rozumiany jako zbiór stanów, relacji pomiędzy tymi stanami i funkcji poznawczych, pozwalających na tworzeniu nowych stanów, jest czymś abstrakcyjnym i niematerialnym. Mózg jest substratem pozwalającym na jego istnienie. Budowa mózgu ogranicza potencjalnie możliwe abstrakcyjne stany umysłu do takich, które można fizycznie zrealizować. Najprostszy statyczny model umysłu może mieć postać grafu, którego węzły reprezentować będą potencjalnie dostępne stany, a łuki prawdopodobieństwa przejść między nimi (Rys. 1). Takie modele znane są jako sieci powiązań przyczynowych lub sieci Bayesowskie [6].

Chociaż teoretycznie umysły mogą być realizowane w oparciu o inny substrat, to szczegółowe odtworzenie relacji zachodzących pomiędzy stanami ludzkiego umysłu wymaga ludzkiego mózgu, ciała i historii rozwoju. Możemy jedynie tworzyć prymitywne modele tłumaczące niektóre aspekty działania ludzkiego umysłu w dobrze określonych warunkach eksperymentalnych. Dotyczy to również najbardziej kontrowersyjnych zagadnień związanych z świadomością. Na obecnym etapie badania tego zjawiska warto skupić się na analizie powiązanych ze sobą zjawisk umysłowych próbując zrozumieć, dlaczego jedne z nich są związane z **świadomością** a inne nie [7].

Pamięć długotrwała ma charakter epizodów, zapamiętanych przeżyć, które możemy sobie uświadomić jako zespół wrażeń. Świadomości wymaga też pamięć deklaratywna, czyli pamięć rzeczy dających się poznać dzięki werbalnym opisom. Innym rodzajem jest pamięć semantyczna, związana z hierarchizacją pojęć i relacjami gramatycznymi, których używamy, chociaż zwykle nie jesteśmy ich świadomi. Pamięć długotrwała zawiera elementy, których nie jesteśmy świadomi do momentu ich aktywacji i przeniesienia do pamięci roboczej [8]. Nie jesteśmy za to nigdy świadomi elementów pamięci proceduralnej, pamięci takich umiejętności jak pływanie czy jazda na rowerze. Nie mamy np. świadomych wrażeń związanych z umiejętnością wiązania krawata czy sznurowania butów (wrażenia związane są z epizodem, ale nie z umiejętnością) i nie da się takich umiejętności nauczyć czytając werbalne opisy. Mo-

żemy badać, dlaczego stan jawy związany jest ze świadomością a koma lub stan snu głębokiego nie. Z tego rodzaju badań wyłania się obraz procesu zwanego świadomością, pozwalający na coraz bardziej precyzyjne zdefiniowanie samego zjawiska.



Rys. 1. Sieć powiązań pomiędzy różnymi pojęciami dotyczącymi umysłu. Istotne są relacje, a nie bezpośrednia reprezentacja pojęć.

Świadome działania i procesy poznawcze zawdzięczają swoją sprawność istnieniu nieświadomych procesów mózgowych. Umysł jest związany z tą częścią zachodzących w mózgu procesów, które mogą stać się, przynajmniej potencjalnie, świadome. Wiele czynności po ich doskonałym opanowaniu przestaje wymagać świadomych decyzji – dotyczy to zarówno czynności manualnych jak i percepcji. Również pamięć jest często niezależna od naszej woli, gdyż pomimo wysiłków nie możemy sobie czegoś przypomnieć, a poszukiwana informacja pojawia się później spontanicznie. Nie potrafimy świadomie zaplanować i kontrolować szczegółów koordynacji setek mięśni potrzebnych do wykonania ruchu lub choćby wypowiedzenia jednego słowa. Działania nieświadome złożone są z automatyzmów, leżących u podstaw zachowania. Nieświadome działania mózgu aktywnie wpływają na postać wrażeń świadomych, percepcji, pamięci i zachowania.

Takie rozumienie nieświadomości, powstałe już w połowie XIX wieku, jest bardzo podobne do rozumienia współczesnego. Tymczasem pojęcie nieświadomości przedstawione w psychoanalizie Freuda było całkiem odmienne. Jego idee oparte były na błędnych analogiach hydraulicznych powstawania „ciśnienia psychicznego”, które musiało znaleźć odpowiednie ujście. Freud zauważył, że nieświadome elementy psychiki mogą mieć wpływ na zachowanie, a uświadomienie sobie tego faktu może mieć terapeutyczny wpływ na pacjentów. Rolą nieświadomości było dla niego filtrowanie informacji dostępnej świadomości, tłumienie i cenzurowanie myśli, które mogłyby prowadzić do konfliktów. Według Freuda „Ja”, świadoma część naszej psychiki, „nie jest nawet panem we własnym domu, lecz poprzestać musi na skąpych wieściach o tym, co odbywa się nieświadomie w jego życiu duchowym”.

2. Relacyjna teoria umysłu

Wielu procesów regulujących stan organizmu nie jesteśmy w stanie postrzegać, czyli sobie ich uświadomić. Szczegóły wielu procesów biochemicznych zachodzących w organizmie nie są ważne z punktu widzenia decyzji, jakie trzeba podjąć. Kiedy poziom glukozy w krwi spada zwierzę musi wiedzieć, że czas szukać jedzenia. Stan krwi monitorowany jest przed podwzgórze, składające się z wielu wyspecjalizowanych, drobnych skupisk neuronów. Jądro boczne podwzgórza reaguje na spadek poziomu glukozy a jądro brzuszno-przyśrodkowe na jej narastanie. Sygnały wysyłane do kory przez jądro boczne interpretowane są jako uczucie głodu, a sygnały z jądra brzuszno-przyśrodkowego jako uczucie nasycenia. Interpretacją zajmuje się kora mózgu, w tym przypadku kora zakrętu obręczy, leżąca na powierzchni boczno-przyśrodkowej mózgu [9]. „Interpretacja” oznacza tu uruchomienie programów działania i pojawienie się wewnętrznych dyspozycji i związanych z nimi stanów mózgu.

Układy sensoryczne przesyłają sygnały z receptorów zmysłowych (oka, ucha, języka) przez wzgórze, pełniące rolę centralnej stacji przekaźnikowej, do pierwotnej kory zmysłowej. Wzgórze przesyła sygnały do kory nowej i do ośrodków podkorowych, specjalizujących się w wykrywaniu zagrożeń i umożliwiających szybkie działanie w oparciu o mało precyzyjną analizę sygnału. Dzięki temu „instynktowne” działanie może poprzedzać świadome rozpoznanie. Wzgórze i inne ośrodki podkorowe regulują dopływ sygnałów do kory biorąc udział w procesach uwagi. Pierwszorzędowa kora zmysłowa szczegółowo analizuje różne cechy sygnału. Kora czuciowa reaguje na miejsce i siłę dotyku, ból, swędzenie, temperaturę i wibracje a kora wzrokowa na kształty (nachylenie krawędzi), kolory, ruch, lub sygnały z lewego albo prawego oka.

Rozpoznawanie obiektów na podstawie informacji dochodzącej ze zmysłów jest bardzo trudne. Mózg przygotowany został przez ewolucję do rozpoznawania często spotykanych obiektów. Zadanie to jest wykonalne dzięki właściwym oczekiwaniom, hipotezom generowanym przez obszary kory skojarzeniowej mieszczącej się w płatach czołowych i skroniowych. O dużej specjalizacji kory świadczą liczne agnozje, czyli niezdolności do rozpoznawania obiektów na podstawie informacji zmysłowej. Są one wynikiem specyficznych uszkodzeń wtórnej i trzeciorzędowej kory sensorycznej (Rys. 2). Np. uszkodzenie kory obręczy spowodować może asymbolię bólu, czyli niezdolność do interpretacji sygnałów bólu jako nieprzyjemnych, przy w pełni zachowanej zdolności do czucia i umiejscowienia tych sygnałów. Uszkodzenie dolnej części zakrętu skroniowego może spowodować prosopagnozję, czyli niezdolność do rozpoznawania twarzy. Uszkodzenia obszaru ciemieniowego w części przyśrodkowej prowadzi do astereognozji, czyli niezdolności do rozpoznawania przedmiotów za pomocą dotyku. Jest też wiele rodzajów agnozji wzrokowych [4].

Wrażenia zmysłowe, uczucie pragnienia czy głodu, wyobrażenia, wszystko, co stanowi treść umysłu, jest rezultatem interpretacji dokonywanej przez korę mózgu. Pojawienie się nowej informacji zmysłowej wymaga interpretacji, a więc odwołania się do już istniejącej wiedzy. Jest to możliwe dzięki zdolności kory mózgu do odtworzenia stanu, w którym mózg się poprzednio już znajdował, czyli dzięki pamięci. Wyjaśnienia wymaga struktura i dynamika stanów mózgu, od reprezentacji wewnętrznej należy jedynie wymagać, by zawierała dostatecznie dużo informacji by w oparciu o nią podjąć odpowiednie działania. Moja interpretacja wrażeń zmysłowych wynika z pojawiających się w mózgu skojarzeń, i zdolności rozróżniania takich stanów jak „przyjemna czerwień zachodzącego słońca” czy „czerwień policyjnego lizaka”. Stan mózgu przy jednej i drugiej obserwacji jest zupełnie odmienny, wynika z wewnętrznych pobudzeń, stanu ciała i receptorów zmysłowych. Wrażenia są rezultatem procesu, w którym mózg interpretuje („komentuje”) swoje własne stany, zmieniając je w wyniku pojawiających się nowych skojarzeń. W odróżnieniu od stanu procesora komputera stan dyna-

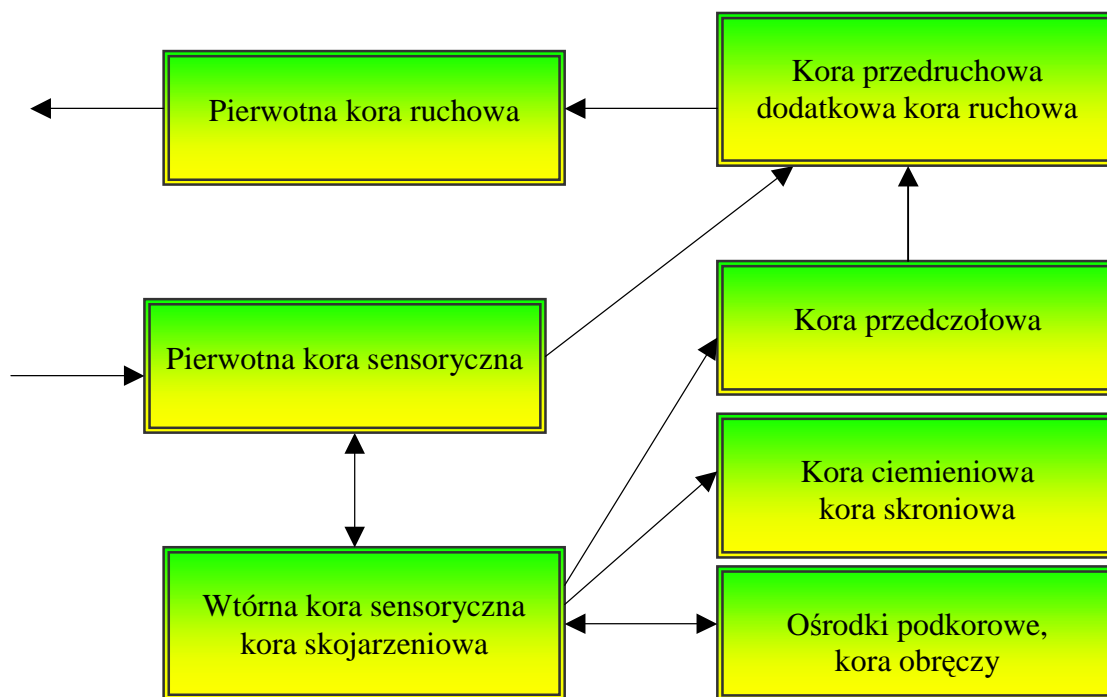
3. Jak działa mózg?

Obecna wiedza o działaniu mózgu, chociaż niepełna i kontrowersyjna, pozwala na sformułowanie prawdopodobnych odpowiedzi na pytania dotyczące natury umysłu i świadomości. Zwierzęta, w szczególności małpy naczelne, mają mózgi podobne do ludzkich. Na poziomie kory mózgu różnice są jednak znaczne, w szczególności w obszarach kory skroniowej i przedczołowej. Jedną z podstawowych cech decydujących o inteligencji jest zdolność do przechowywania informacji o bieżącej sytuacji w pamięci roboczej, utrzymywania ich „w umyśle”. Bez pamięci roboczej [8] niemożliwa byłaby wyobraźnia, wykroczenie poza bieżącą chwilę („teraz i tutaj”), a więc również i świadomość. Wszystko, czego jesteśmy świadomi, musi znajdować się w pamięci roboczej, chociaż nie jesteśmy świadomi wszystkiego, co w niej się znajduje. Świadomość można więc uznać za proces wyróżniający pewne elementy pamięci roboczej, pozwalający na wykorzystanie znajdującej się w niej informacji do działania. To działanie niekoniecznie musi wyrażać się skurczami mięśni, może to być rozpoczęcie planowania działania, umożliwienie zapamiętania lub przywołanie jakiegoś skojarzenia z pamięci. W ten sposób powstaje „strumień świadomości”, ciąg niewypowiedzianych zdań (myśli), obrazów, pobudzeń emocjonalnych.

Pojemność pamięci roboczej jest niewielka i wynosi u ludzi około 7 ± 2 porcje informacji (np. grup cyfr czy przypadkowych słów) lub 4 ± 2 bardziej złożone obiekty [10]. W czasie wykonywania czynności wymagających wykorzystania zapamiętanej informacji kilkanaście do kilkudziesięciu sekund po jej przedstawieniu główną rolę gra kora przedczołowa. Szczególnie ważne jest pamiętanie relacji przestrzennych, pozwalające na prawidłowe działanie nawet jeśli wykryte obiekty (wzrokiem, słuchem lub dotykiem) przestają chwilowo być postrzegane. W te czynności zaangażowane są różne obszary płatów przedczołowych.

Świadomość z pewnością wydawałaby się procesem bardziej zrozumiałym gdyby udało się znaleźć takie cechy w możliwych do zaobserwowania sygnałach aktywności mózgu, które jednoznacznie korelują się z subiektywnymi wrażeniami. W eksperymentach z rywalizacją obuoczną [10] małpa widzi każdym okiem inny obraz, ale jest świadoma przez kilka sekund tylko jednego z nich, co sygnalizuje odpowiednim przyciskiem. Nie stwierdzono korelacji pomiędzy świadomym rozpoznanem danego obrazu a aktywnością większości neuronów niższych piętter kory wzrokowej (obszary V1, V4, MT). Jest to zrozumiałe, bo obszary te dokonują ciągłej analizy elementarnych cech sygnału wzrokowego. Aktywność 90% neuronów w obszarze dolnego i górnego zakrętu skroniowego (IT, STS) jest za to skorelowana z rozpoznawaniem konkretnego obiektu. Podobne wyniki otrzymano za pomocą metod magnetoencefalograficznych (MEG) u ludzi [12].

Nie oznacza to jednak, że znalezienie obszarów kory mózgu aktywnych w czasie rozpoznawania konkretnego obiektu wyjaśnia, czym jest widzenie. Aktywność neuronów nie produkuje wrażeń wzrokowych, ale jest konieczna do eksploracji wzrokowej świata, podjęcia odpowiednich działań, interpretacji stanów układu wzrokowego umożliwiającej takie działania [12]. Strumień świadomości jest ciągiem interpretacji stanów części pamięci roboczej, dokonywanych przez **system interpretacji** związany z działaniem lewej półkuli i płatów czołowych [14]. Pamięć robocza pozwala na rozpowszechnienie informacji o bieżącej sytuacji organizmu do wszystkich części mózgu, dzięki czemu informacja ta może zostać uzupełniona lub zmieniona w procesach kojarzeniowych. System interpretujący ma dostęp do pamięci roboczej, do wewnętrznego modelu stanu organizmu, ma wpływ na planowanie działań motorycznych oraz generowanie mowy, pozwala też na zapamiętanie wyników swojego działania w pamięci trwałej.



Rys. 3. Od postrzeżenia do działania – obszary kory mózgu.

Za częściową synchronizację procesów zachodzących na najwyższych piętrach przetwarzania, w obszarze dolnoskroniowym (IT) i przedczołowym, tworzących strumień świadomości, odpowiedzialne są prawdopodobnie procesy selekcji informacji dokonywane przez ośrodki układu limbicznego, brzuszną część prądkowia (a zwłaszcza jądro półleżące) i jądro siatkowate wzgórza, oraz tworzącego siatkowatego pnia mózgu. Tylko najważniejsze informacje dają bezpośredni wkład do pamięci roboczej, scalającej stan tych obszarów mózgu, które w istotny sposób wpływają na kontrolę zachowania. Długotrwała pamięć epizodyczna zapisuje przede wszystkim stan pamięci roboczej, pozwalając go odtworzyć w późniejszym okresie, powrócić do minionej chwili. Jedną z funkcji świadomości jest filtrowanie informacji, pozwalające zapamiętać z morza kombinacji docierających do mózgu sygnałów tylko te najważniejsze.

Istotną rolę w powstaniu systemu interpretacji grają neurony brzusznej części kory przedruchowej, aktywne w czasie wykonywania specyficznych ruchów, np. łapania, podnoszenia lub popychania ręką. Te same neurony u małpy uaktywniają się również w czasie obserwacji innej małpy lub eksperymentatora wykonującego te same czynności [15]. Podstawą uczenia się przez imitację, zdolności do empatii i rozumienia innych istot jest możliwość powstawania podobnych procesów zachodzących w obu mózgach.

Wrażenia nie powstają bezpośrednio z sygnałów zmysłowych, lecz są wynikiem odwoływania się do już istniejącej wiedzy o świecie, 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. Wrażenia zmysłowe są więc wynikiem niewerbalnej interpretacji stanu pierwotnej i wtórnej kory zmysłowej. Można na ten proces spojrzeć jak na dopasowanie dochodzącej informacji do z góry ustalonych oczekiwań dzięki wiedzy zdobytej na podstawie wcześniejszych doświadczeń. Kiedy np. rozmawiamy w hałaśliwym miejscu z szumu wybierane są odpowiednie formanty dzięki oczekiwaniom i istniejącej w danym momencie w pamięci roboczej interpretacji. W mózgu powstają wówczas

mięci roboczej interpretacji. W mózgu powstają wówczas stany rezonansowe, przedłużające i nasilające pobudzenia, łącząc w tym przypadku fonetykę z semantyką zdania.

Emocje są wynikiem interpretacji stanu ośrodków podkorowych, takich jak ciało migdałowe w przypadku strachu, oraz wpływu tych ośrodków na stany pamięci roboczej. Wrażenia i emocje mają jakościowo różny charakter, gdyż odpowiadają im odmienne stany mózgu o różnych własnościach strukturalnych. Emocji nie można oddzielić od procesów poznawczych – interpretacja jest procesem poznawczym a pobudzenia emocjonalne ukierunkowują jego działanie pozwalając na dokonanie wyboru pomiędzy różnymi działaniami.

Psychologia naukowa odrzuciła pojęcie nieświadomości sformułowane przez Freuda jako nie podlegające eksperymentalnej weryfikacji. W podejściu psychoanalitycznym, chociaż błędnym w szczegółach, zawarte jest jednak ziarno prawdy. Z psychologii rozwojowej wiemy, że na poziomie świadomym mamy jedynie dostęp do teorii na temat siebie, teorii budowanej we wczesnym dzieciństwie, opartej na wiedzy o świecie i sobie jako istoty w określonym środowisku. System interpretujący ma teorię „Ja”, pozwalająca mu odróżniać procesy odnoszące się do własnego organizmu oraz do własnej teorii wiedzy o świecie [14].

Działanie, jak też i wyobrażenie działania lub obserwacja wykonywania jakiejś czynności, wymaga aktywności kory przedczołowej. Podjęcie działania interpretowane jest jako wolny wybór dokonany przez „ja”, **proces wolicjonalny**. Sam proces interpretacji trwa około 0.5 sekundy – pobudzenie kory motorycznej widoczne jest o tyle wcześniej w stosunku do momentu uświadomienia sobie, że zamierzamy podjąć jakieś działanie, np. nacisnąć przycisk. Potknięcie na schodach prowadzi zwykle do działania celem utrzymania równowagi zanim uświadomimy sobie co się dzieje i zanim poczujemy strach. Świadomość nie jest czynnikiem sprawczym, podejmującym decyzję, lecz jedynie wynikiem interpretacji stanu mózgu. Wiele wyspecjalizowanych procesów w mózgu zachodzi bez konieczności komentowania skutków ich działania przez system interpretacji, dlatego nie są uświadamiane.

4. Od mózgu do umysłu.

Modele pamięci i funkcji poznawczych dotyczą działania mózgu. Zrozumienie umysłu wymaga „wejścia do środka” sieci relacji pomiędzy stanami mózgu. Czy można tak opisać neurodynamikę, by dostrzec w niej zjawiska mentalne? Umysł jest częścią tego, co robi mózg lub, wyrażając to metaforycznie, umysł jest cieniem neurodynamiki. Procesy, mające istotne znaczenie dla kontroli organizmu na najwyższym poziomie, mają swój wkład do pamięci roboczej. Uświadomienie sobie tych elementów polega na interpretacji stanów pamięci roboczej, „zdawaniu sobie sprawy” przez jedne obszary mózgu ze stanu innych. Neurodynamika jest znacznie bogatsza niż zdarzenia mentalne. Nawiązując do słynnej alegorii Platona możemy powiedzieć, że widzimy cienie prawdziwej rzeczywistości na ścianie jaskini. Prawdziwa rzeczywistość odbija się w „aktywnym lustrze” naszych mózgów, a to, co dostrzegamy w naszym umyśle, jest cieniem tego odbicia.

Próby zrozumienia świata umysłu na podobieństwo świata fizycznego czynione były w psychologii od dawna. Kurt Levin już w 1938 roku [16] proponował rozpatrywanie zdarzeń mentalnych w odpowiedniej przestrzeni psychologicznej, w której agent (autonomiczne działający system poznawczy) poddawany jest działaniu „sił kognitywnych”. George Kelly zaproponował analizę zjawisk psychologicznych bardziej w oparciu o geometrię niż logikę [17] w ramach swojej „psychologii osobistych konstruktów”. Kelly chciał doprowadzić do powstania oprogramowania dla symulacji osobowości i wizualizacji wielu procesów psychologicznych. Proponowano, by przyjąć jego idee za centralny model nauk kognitywnych [18], nie są one jednak powszechnie znane.

Roger Shepard, bardzo wpływowy psycholog ze Stanfordu, usiłował znaleźć w świecie umysłu uniwersalne prawa, na wzór praw fizycznych [19]. Według niego „nie potrzeba nam

więcej danych lub bardziej szczegółowych danych, lecz innego podejścia do problemu”. W fizyce opis ruchu jest prosty dzięki symetriom, pozwalającym zdefiniować niezmienniki ruchu w odpowiednich przestrzeniach. W przestrzeniach Euklidesowych można zdefiniować transformację Galileusza, w pseudo-Euklidesowych (3+1) transformację Lorentza a w przestrzeniach Riemanna transformacje w układzie przyspieszającym. Prawa psychologiczne należy również formułować w odpowiednich przestrzeniach.

Jako przykład Shepard [20] pokazuje dane psychofizyczne dla zwierząt i ludzi dotyczące uogólniania bodźców wizualnych, słuchowych i smakowych. Dla N par bodźców bada się generalizację G , czyli prawdopodobieństwo tego, że zachowanie wyuczone dla jednego z nich pojawi się dla drugiego. Procedura skalowania wielowymiarowego, stosowana często w psychometrii, przekształca te dane porządkując je monotonicznie tak, by wzajemne relacje podobieństwa przedstawić w postaci odległości D w jednym wymiarze. Jeśli wykreślić $G(D)$ dla dowolnego eksperymentu tego typu otrzymamy zależność eksponencjalną $G(D)=\exp(-\alpha D)$.

Inne przykłady uniwersalnych zależności opierają się na symetriach. 24-godzinny zegar biologiczny jest bezpośrednią konsekwencją prawa zachowania moment pędu. Skąpe informacje docierające z siatkówki wystarczają do odtworzenia wrażenia ciągłego obrazu poruszającego się przedmiotu dzięki internalizacji własności geometrycznych Euklidesowej przestrzeni. Obiekty nie znikają w przypadkowy sposób, więc mózg wykorzystuje „zasadę zachowania obiektów” tworząc wrażenie pozornego ruchu, np. wtedy, gdy dwa znajdujące się obok siebie podobne obiekty naprzemiennie znikają i pojawiają się. Możemy postrzegać trójwymiarowe obroty na komputerowo generowanych obrazach, ale nie jesteśmy zdolni do wyobrażenia sobie obrotów w czterech wymiarach. Jeśli obiekty używane do wywołania wrażenia ruchu pozornego mają różną orientację to postrzegany ruch zachodzi zwykle wokół osi symetrii danego obiektu. Podobnie dzieje się przy eksperymentach z mentalną rotacją, w których należy porównać dwie asymetryczne figury geometryczne.

Intuicyjna fizyka ruchu wyobrażanych obiektów opiera się bardziej na geometrii kinematycznej (jest to gałąź matematyki zajmująca się charakteryzacją ruchu obiektów geometrycznych) niż na dynamice Newtona [20]. Trajektorie postrzeganego ruchu pozornego to linie geodezyjne obrotu i przesunięcia, a więc jest to ruch helikalny. Położenie sztywnego, asymetrycznego obiektu opisać można za pomocą położenia wybranego punktu i kątów określających orientację. Przestrzeń rozróżnialnych położeń jest 6-wymiarową rozmaitością a 3 jej wymiary (rotacje) są zwinięte kołowo. Rozmaitość ta ma symetrię grupy będącej iloczynem półprostym dwóch grup $E^+=\mathbb{R}^3\otimes\text{SO}(3)$. Linie geodezyjne w tej przestrzeni, tworzące rodzinę jednoparametrowych podgrup, odpowiadają helikalnym trajektoriom geometrii kinematycznej. Dla obiektów posiadających częściową symetrię struktura przestrzeni ulega uproszczeniu. Można ją bezpośrednio powiązać z rezultatami eksperymentów psychofizycznych. Interpretacja kształtów obiektów jest bardziej skomplikowana i wymaga większej liczby wymiarów. I w tym przypadku można sobie wyobrazić proces rozpoznawania jako ruch po geodezyjnej trajektorii w stronę prototypu kształtu danego obiektu [20],[21].

Charakteryzacja różnych aspektów przestrzeni, w których należy analizować wrażenia związane z percepcją jest pierwszym krokiem w kierunku fizyki umysłu. Postrzeganie koloru – tematem tym zajmował się Newton, Helmholtz, Maxwell i Schrödinger – jest tu również dobrym przykładem. Pomimo zmiany rozkładu spektralnego światła odbitego od powierzchni postrzegany kolor jest stabilny. Przestrzeń kolorów przedstawić można w postaci sfery, podając intensywność, nasycenie i barwę, która opisana jest przez zmienną kątową. Już Newton zauważył, że skrajne kolory widma, ciemna czerwień i fiolet (po zmieszaniu dające wrażenie purpury), są bardziej podobne do siebie niż do kolorów leżących w środku widma. Jest to oczywiście cecha naszych umysłów a nie własność fizyczna światła. Niezmienniczość postrzeganego koloru $P(x,\lambda) = I(\lambda)S(x,\lambda)$ w punkcie x odbijającej powierzchni S przy różnym

oświetleniu $I(\lambda)$ wymaga odróżnienia wkładu $S(x, \lambda)$ dzięki informacji zaledwie z trzech receptorów koloru. Rozważania teoretyczne pokazują [22], że do wykonania tego zadania potrzeba przynajmniej 6 typów receptorów, nie widzimy więc w pełni wszystkich cech powierzchni odbijającej. Jednakże zmienność naturalnych warunków oświetlenia można w przybliżeniu opisać za pomocą 3 parametrów (jest to jeden z alternatywnych sposobów opisu przestrzeni kolorów [23]): natężenia światła, przesunięcia maksimum widma światła od czerwieni do zieleni, zależnego od absorpcji atmosfery, oraz przesunięcia tego maksimum pomiędzy kolorem niebieskim i żółtym, zależnie od rozpraszania światła. Układ wzrokowy dokonuje kompensacji w tych trzech wymiarach zachowując stałość postrzeganego koloru [23].

Makroskopowe własności obiektów i ich ruchu postrzegane na poziomie umysłu są wynikiem sposobu przetwarzania informacji na poziomie mikroskopowym, przez sieci neuronowe mózgu. Struktura tych sieci wykształciła się w wyniku ewolucji tak, by z informacji dostarczanej przez zmysły wydobywać najbardziej przydatne i wiarygodne cechy. Świat umysł wykorzystuje tylko te cechy, które są przydatne z punktu widzenia sprawnego działania. Dotyczy to również rozpoznawania kategorii obiektów. Porównywanie obiektów pełni podstawową rolę i dokonywane jest za pomocą najszybszych transformacji przy zachowaniu niezmienniczych własności obiektów. Ważną rolę pełni tu pamięć.

5. Modele pamięci i sieci atraktorowe.

Na ile prawdziwa jest metafora mózg-komputer? W jakim sensie mózg prowadzi obliczenia: czy tak, jak maszyna Turinga realizowana przez cyfrowe komputery, czy też w sposób analogowy, tak jak zwijające się białko minimalizuje swoją energię? Zjawiska fizyczne powinny być modelowane za pomocą ciągłych modeli i symulowane za pomocą ciągłych systemów obliczeniowych [24]. Procesy zachodzące w tkance nerwowej również do takich należą. Modele sieci neuronowych oparte na zliczaniu impulsów czy uśrednionej aktywacji neuronów są jedynie przybliżeniem do opisu ewolucji takich układów. Kolejne stany tej ewolucji interpretujemy jako wyniki procesów obliczeniowych. Lepszym modelem stanu tkanki neuronowej byłby model fizyczny, np. analogowy układ elektryczny złożony z bardzo wielu lokalnych oscylatorów.

Kora mózgu ma budowę modułową, składa się z obserwowalnych pod mikroskopem kolumn o średnicy ułamka milimetra, zawierających około 100.000 neuronów. Typowa kolumna jest obiektem bardzo złożonym, zawiera tysiące mikroobwodów w niewielkim stopniu sprzężonych ze światem zewnętrznym. Neurony kilkunastu rodzajów pobudzają się wzajemnie przy pomocy różnorodnych synaps reagujących na dziesiątki neurotransmiterów i neuromodulatorów. Struktura tak złożonego obiektu musi być w znacznej mierze przypadkowa, gdyż informacja przekazywana w genach nie wystarcza by ją w pełni określić. W jaki sposób reprezentowana jest informacja w takich strukturach?

Podstawowym zadaniem kolumny jest rozróżnianie dochodzących do niej sygnałów, tworzenie unikalnej sygnatury, którą inne kolumny będą mogły odczytać. Układ tysięcy przypadkowych mikrooscylatorów zaburzony przez zewnętrzne sygnały będzie przez krótki czas pobudzony w charakterystyczny sposób [25]. Odzworowuje to w nieliniowy sposób dochodzące sygnały w przestrzeń o dużej liczbie wymiarów. Inne kolumny kory otrzymują sygnały o jej aktywności i potrafią łatwo rozróżnić, w jakim jest ona stanie, za pomocą zwykłego mechanizmu uczenia się w oparciu o wzmacnianie siły połączeń synaptycznych jednocześnie aktywnych neuronów (nazywa się to uczeniem Hebbowskim, na cześć Donalda Hebba). Większość połączeń wewnątrz kolumn może być przypadkowa i nie musi podlegać uczeniu.

Maass i współpracownicy nazwali ten model „płynem neuronowym” [25] i pokazali, że układ takich kolumn ma moc równą maszynie Turinga dla zadań obliczeniowych w czasie

rzeczywistym. Taki układ działa stabilnie nie wymagając powtarzalnych stanów wewnętrznych. Zastosowano go między innymi do identyfikacji słów reprezentowanych przez zsumowane ciągi impulsów. Reprezentacją rozpoznawanego obiektu jest chwilowy rozkład częstości oscylacji mikroobwodów. Można go przedstawić jako rozkład gęstości prawdopodobieństwa (PDF) pojawienia się sygnału w jakiejś części przestrzeni parametrów wewnętrznych kolumny.

Taki mechanizm może prawdopodobnie działać przez ułamki sekund, co wystarcza do dyskryminacji różnych bodźców, rozpoznawania słów czy wzrokowego rozpoznawania obiektów lub ich fragmentów. Pierwszorzędowa kora zmysłowa zawiera wyspecjalizowane obwody wykrywające różne aspekty sygnału, np. kolor, nachylenie krawędzi, kontury, ruch. Kolumny mogą specjalizować się w wykrywaniu różnych cech sygnałów, gdyż otrzymują je z różnych obszarów kory zmysłowej. Drugorzędowa kora zmysłowa, odpowiedzialna za nadawanie sensu odbieranym bodźcom (uszkodzenia tej kory powodują *asymbolię*, czyli niezdolność do rozumienia znaczenia bodźców przy nienaruszonej zdolności odczuwania) może działać w ten sposób. Jednakże pojawienie się jakiegoś obiektu w pamięci roboczej, a następnie jego uświadomienie, wymaga procesów zachodzących w dłuższej skali czasu. Mamy tu do czynienia ze sprzężeniem zwrotnym pomiędzy kolumnami [26].

Nawet najprostsze modele neuronowe pamięci mają wiele własności typowych dla pamięci biologicznych. Pamięć zapisana jest w sposób rozproszony w wielu połączeniach synaptycznych, więc zniszczenie części sieci prowadzi do łagodnych zmian (*graceful degradation*) a nie zapomnienia poszczególnych wzorców. Dzięki temu możliwe jest również rozpoznawanie uszkodzonych wzorców i skojarzenia nowych wzorców z najbardziej podobnymi wśród zapamiętanych. Adresowalność kontekstowa oznacza, że pobudzenie sieci fragmentem zapamiętanego sygnału prowadzi do przypomnienia całego wzorca. W takich sytuacjach mogą zdarzać się pomyłki i skojarzenia fonologiczno-semantyczne. W szczególności uszkodzone modele neuronowe zamiast przypominać sobie sekwencję zapamiętanych wzorców mogą produkować halucynacje, fałszywe stany atraktorowe, poskładane z fragmentów zapamiętanych wzorców. Przepełnienie pamięci prowadzi do gwałtownego wzrostu liczby pomyłek i powstania fałszywych wzorców pamięci. Czas przypomnienia nie zależy od liczby zapamiętanych wzorców, w odróżnieniu od baz danych, gdy trzeba kolejno porównać zapamiętane wzorce.

Realistyczne modele pamięci wymagają uwzględnienia faktu, że oprócz pamięci długoterminowej, zapisanej w połączeniach neuronów kory mózgu, mamy również pamięć średnio-terminową, w starej korze hipokampa. Zaletą dwóch systemów pamięci jest możliwość budowania stabilnego modelu świata, przenoszenia do pamięci trwałej tylko ważnych faktów, pasujących do wcześniej zapamiętanych, powolnej modyfikacji lub dodawania nowych faktów bez katastroficznej interferencji ze starymi. Wyjaśnienie własności różnych rodzajów amnezji: wstecznej, następczej, całkowitej przejściowej, trudności z uczeniem się, wymaga dodatkowo uwzględnianie wpływu neuromodulacji na plastyczność hipokampa i kory [5]. Mechanizm neuroplastyczności jest dość skomplikowany, związany zarówno z mechanizmami uwagi jak i emocji.

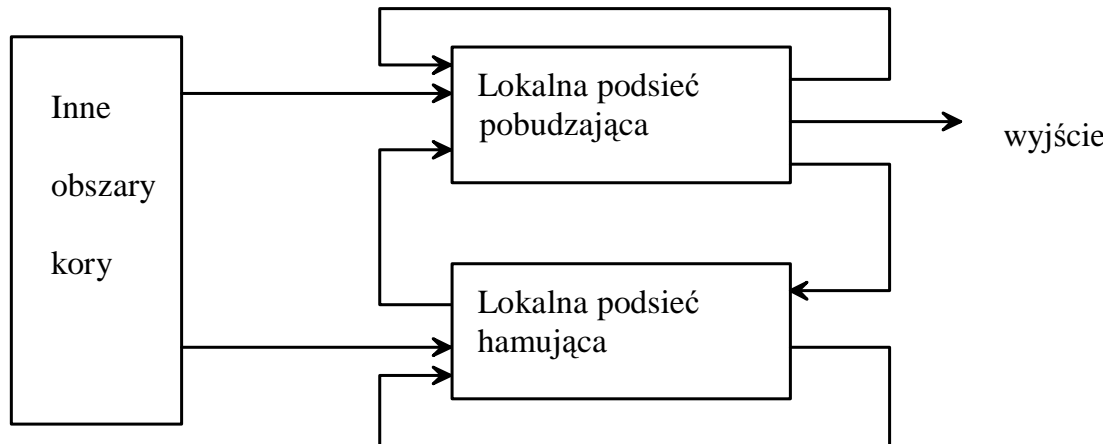
Sieci neuronowe stosowane dotychczas w modelowaniu funkcji mózgu są skrajnie uproszczonymi modelami. Amit i jego współpracownicy zrobili ważny krok na drodze ku stworzeniu prostych modeli sieci neuronowych poprawnych z neurobiologicznego punktu widzenia [26], [27]. W jaki sposób powstają w sieciach neuronowych i jak zachowują się stany odpowiedzialne za pojawienie się treści umysłu (elementy pamięci roboczej)? Jak można je pogodzić z istnieniem aktywności spoczynkowej mózgu? Badanie dynamiki sieci przydaje się do interpretacji i daje się bezpośrednio porównać z wynikami badań doświadczalnych na zwierzętach. Zastosowanie takich sieci w modelowaniu opisanych powyżej zjawisk pozostaje

nadal wyzwaniem dla badaczy. Przy konstrukcji modelu zgodnego z danymi neurobiologicznymi warto zrobić następujące założenia:

1. Aktywność tła ma charakter stochastyczny – jest to założenie dobrze spełnione w tak dużym układzie jak mózg.
2. Jednorodność: neurony w czasie aktywności spoczynkowej są, z dokładnością do fluktuacji statystycznych, w identycznym środowisku. Aktywność spoczynkowa nie wiąże się ze specyficznym działaniem kory (w szczególności z procesami pamięci) i może być obserwowana tylko w stanach snu głębokiego, pozbawionego marzeń sennych. Fluktuacje konieczne są do powstania aktywności spoczynkowej.
3. Impulsy wysyłane przez różne neurony w stanie spoczynku nie są skorelowane. Wynika to z różnicy czasów pomiędzy długością trwania impulsu (milisekunda) i procesami integracji a czasami pomiędzy impulsami, które dla aktywności spoczynkowej wynoszą 200-1000 milisekund.
4. Aktywacja neuronu jest sumą wkładów synaptycznych. To założenie jest dobrze spełnione, jeśli mamy wiele nieskorelowanych kanałów wejściowych dochodzących w ciągu czasu integracji impulsów.
5. Dynamika aktywności spoczynkowej i selektywnej daje się opisać przez Gaussowski rozkład wkładów synaptycznych.
6. Do opisu wystarczy aktywność neuronu mierzona liczbą impulsów na sekundę - oznacza to, że zaniedbujemy procesy związane z synchronizacją aktywności neuronów.
7. Kora ma budowę modułową, każdy neuron otrzymuje około 80% impulsów z lokalnych obwodów pobudzających i 20% z hamujących. Tylko jednostki pobudzające mają modyfikowalne synapsy.

Procesy uczenia w takich modelach, związane z długotrwałym wzmocnieniem synaptycznym (LTP), przebiegają w dwóch etapach: w początkowej fazie uczenia moduł biorący udział w rozpoznawaniu zwiększa w nieselektywny sposób częstość impulsacji dla wszystkich dochodzących do niego sygnałów. Powyżej pewnej krytycznej wartości wzmocnienia LTP na tle globalnej aktywności pojawiają się lokalne atraktory. Są one odbiciem struktury pobudzeń wywołanej przez sygnały otrzymywane w procesie uczenia się. Pojawienie się rozpoznawanego sygnału powoduje podniesienie aktywności od spoczynkowej (kilka Hz) do około 20 Hz. Aktywność ta pozostaje przez pewien czas nawet po zniknięciu bodźca – odpowiada to utrzymywaniu aktywnej reprezentacji bodźca w pamięci. Nie wszystkie neurony lokalnego modułu biorą udział w rozpoznawaniu czy też kodowaniu wewnętrznej reprezentacji, stąd tylko niektóre będą w stanie podwyższonego pobudzenia.

Przewidywania modelu dobrze zgadzają się z obserwacjami neurofizjologicznymi dotyczącymi pomiarów aktywności neuronów małych w czasie wykonywania zadań wymagających aktywnej pamięci pokazywanego przez krótki czas bodźca (*delayed match-to-sample*). Podwyższoną aktywność spontaniczną obserwuje się w trakcie uczenia rozpoznawania, dopiero po nauczeniu widać aktywność związaną z lokalnymi atraktorami.



Rys. 4. Schemat modułu kory mózgu.

Makrokolumna, zajmująca około jednego mm^2 kory mózgu, ma w przybliżeniu 10^5 neuronów oddziaływujących z wielką liczbą neuronów znajdujących się w pozostałych częściach kory. Neurony mają trzy typy połączeń: pobudzające i hamujące wewnątrz modułu, oraz pobudzające dochodzące z zewnątrz (dochodzące od długich aksonów komórek piramidowych przekazujących impulsy od innych modułów). Pobudzenia wewnętrzne są silniejsze niż pobudzenia dochodzące z zewnątrz, wpływając przede wszystkim na utrzymanie spontanicznej aktywności. Można je modelować za pomocą rozkładu Poissona. Każdy neuron lokalnej sieci reaguje na depolaryzację $V(t)$ swojej membrany zgodnie z równaniem:

$$\tau \dot{V}(t) = -V(t) + I(t)$$

gdzie prądy synaptyczne $I(t)$ wyrażone są w jednostkach napięcia (pomnożone przez stałą oporności membrany). Stała czasowa $\tau \sim 10$ ms jest szybkością całkowania depolaryzacji w pobliżu wierzchołka aksonowego. Kiedy depolaryzacja osiągnie próg θ neuron wysyła impuls i po czasie refrakcji rzędu 2 ms wraca do potencjału spoczynkowego H . Zakładamy, że neuron ma około $C \approx 20.000$ synaps, przy czym sprawność synaptyczna (siła połączeń lub waga) J_i ($i=1..C$) rozkłada się w przypadkowy sposób, zgodnie z rozkładem Poissona $P(J)$ wokół średniej J z wariancją $J\Delta$. Prądy synaptyczne opisywane są przez równanie:

$$\tau_1 \dot{I}(t) = -I(t) + \sum_{i=1}^C J_i \tau \sum_{\alpha} \delta(t_i^{\alpha} - t)$$

gdzie sumowanie przebiega po wszystkich synapsach oraz po impulsach α dochodzących w czasie t_i^{α} do danej synapsy. Stała czasowa τ_1 określa tempo zmian przewodności synaps i jest znacznie krótsza (rzędu ułamków milisekund) od stałej czasowej τ . W dłuższych odcinkach czasu prądy zdominowane są przez sumę po prawej stronie powyższego równania a więc:

$$\tau \dot{V}(t) = -V(t) + \sum_{i=1}^C J_i \tau \sum_{\alpha} \delta(t_i^{\alpha} - t)$$

Dla pojedynczego impulsu rozwiązaniem tego równania jest funkcja eksponencjalna opisująca potencjał postsynaptyczny EPSP, $V(t) = J e^{-t/\tau}$. Oceny J wahają się od 0.05 do 0.5 mV. Jeśli przyjmą $J=0.2\text{mV}$ i próg rzędu 20mV to potrzeba około 100 impulsów dochodzących w

czasie τ by doprowadzić do wzbudzenia powyżej progu. Jeśli do każdej synapsy dochodzi średnio ν niezależnych impulsów to z powodu dużej liczby kanałów wejściowych rozkład całkowitej liczby impulsów jest gaussowski o średniej $\mu(\nu) = JC\nu\tau$ i wariancji $\sigma(\nu)$. Prawdopodobieństwo wysłania impulsu w przedziale czasowym o długości τ przez tak pobudzany neuron wynosi:

$$P(\nu) = \int_{\theta}^{\infty} \exp\left\{-\frac{(I - \mu(\nu))^2}{2\sigma^2(\nu)}\right\} \frac{dI}{\sqrt{2\pi\sigma^2(\nu)}}$$

Wprowadzając zrenormalizowaną „odległość” średniego pobudzenia od progu:

$$a(\nu) = \frac{\theta - \mu(\nu)}{\sigma(\nu)}$$

wzór na prawdopodobieństwo na jednostkę czasu mierzoną w stałych czasowych τ , określający średnią liczbę impulsów w czasie τ , przyjmuje prostą postać:

$$P(\nu) = \int_{a(\nu)}^{\infty} \exp\left(-\frac{z^2}{2}\right) \frac{dz}{\sqrt{2\pi}}$$

Dla typowych neuronów $\tau = 10\text{ms}$, $C = 20.000$, $\nu = 0.02$ (2 impulsy/sek) mamy $\mu = 400J$ oraz $\sigma = 20J$. Dokładniejszy model pokazuje, że jest to dobre przybliżenie dla aktywności nie przekraczających 100 impulsów na sekundę.

W sieci złożonej z takich neuronów można rozwiązać problem stabilności sieci i poszukiwać spontanicznych częstości impulsacji. Wiadomo, że neurony w mózgu wykazują ciągłą aktywność – przejście w stan spoczynku oznaczałoby śmierć. W dynamice mózgu musi więc istnieć bardzo stabilny, rozległy atraktor, gwarantujący stabilność działania sieci neuronów. Z problemem tym nie potrafią sobie poradzić prostsze sieci (np. sieci Hopfielda), oparte na atraktorach punktowych. W przypadku braku hamowania rozwiązania są niestabilne względem fluktuacji częstości. Dodając około 10% synaps pochodzących od hamujących interneuronów można otrzymać stabilne rozwiązania. W modelu mamy następujące parametry:

- liczbę synaps pobudzających $C_E = 20.000$;
- liczbę synaps hamujących $C_I = 2.000$;
- ułamek impulsów pobudzających od innych neuronów wewnątrz modułu w stosunku do neuronów zewnętrznych,
- średnie wartości sprawności synaptycznych $J_{EE}, J_{EI}, J_{IE}, J_{II}$,
- czasy integracji neuronów hamujących τ_I oraz pobudzających τ_E .

Należy obliczyć średnie częstości ν_I oraz ν_E dla samoreprodukującej się aktywności spontanicznej. Globalna stabilność w największym obszarze zmienności tych parametrów osiągnięta jest dla stałych czasowych synaps hamujących wyraźnie krótszych od pobudzających. Jednakże przy czasach τ_I bardzo krótkich (ok. 2 ms) sieć staje się wrażliwa na opóźnienia czasowe związane z przesyłaniem impulsów między neuronami i przestaje być stabilna, gdyż opóźnienia te wzmacniają lokalne fluktuacje i rośnie prawdopodobieństwo przejścia do stanu absolutnego spoczynku.

Tak skonstruowana sieć działająca spontanicznie może uczyć się reprezentacji sygnałów dochodzących z zewnątrz korzystając z lokalnej reguły plastyczności Hebb'a. Załóżmy, że moduł ma się nauczyć p reprezentacji. Niech każda z nich zwiększa aktywność f procent neuronów pobudzających, przy czym każda reprezentacja kodowana będzie przez niezależne neurony modułu, $f p < 1$. Jest to wygodne przybliżenie działające przy niewielkiej liczbie reprezentacji i stosunkowo dużych sieciach. W sytuacjach doświadczalnych (uczenie małp rozpoznawania obrazów) $p=50$ a $f=0.01$. Mamy więc podsieć pfN neuronów reagujących na jeden z bodźców p i pozostałych neuronów działających w sposób spontaniczny. Rozkład prawdopodobieństwa $P_+(J)$ sprawności synaptycznych podsieci reagującej na pobudzenie jednym z p wzorców ma średnią $J_{+} > J_{EE}$ oraz odpowiednią wariancję. Jednocześnie część $(1-f)C_{EE}$ synaps neuronów pobudzających nie reaguje na wybrany wzorzec i opisać je można rozkładem $P_-(J)$ o średniej $J_- < J$. Średnio rzecz biorąc wzmocnienie synaptyczne jednych synaps równoważone jest osłabieniem innych. Możliwe są dwa scenariusze uczenia się: stochastyczna rekrutacja coraz większej liczby synaps znajdujących się w dyskretnych stanach lub ciągłe, stopniowe zwiększanie wzmocnień synaptycznych. W sieciach tego typu, przy dużej liczbie neuronów, przybliżenie dyskretne wydaje się funkcjonować dobrze.

Wyniki symulacji programem SpikeNet dla 2000-400.000 neuronów (jeden moduł złożony jest z 2000 neuronów) pokazują, że spontaniczna aktywność jest stabilna w czasie lokalnego uczenia się, chociaż w module uczącym się pojawia się podwyższona częstość impulsacji wśród neuronów biorących udział w kodowaniu któregoś wzorca i nieco obniżona wśród pozostałych neuronów (jest to zapewne wynikiem przyjętego założenia o równowadze pomiędzy wzmocnieniem a osłabieniem synaptycznym). To podwyższenie średniej częstości impulsacji przy prezentacji wzorców zapowiada pojawienie się nowego atraktora. Przy wzrastającym stosunku J_+/J w pewnym momencie pojawia się bifurkacja i dwa rozwiązania stabilne, odpowiadające spontanicznej aktywności i lokalnemu atraktorowi.

W eksperymentach opóźnionego wyboru (*Delayed Match to Sample*, DMS) makak uczy się zapamiętać sekwencję wielu przypadkowych kształtów (by uniknąć przypadkowych korelacji stosuje się obrazy fraktalne); w fazie testu po krótkiej prezentacji kilku obrazów i przerwie rzędu 30 s małpa ma pokazać właściwą sekwencję obrazów. W tym celu musi je utrzymać w pamięci roboczej, pobudzając odpowiednie kolumny kory zaangażowane w tworzenie śladów pamięci. Kolumny te znajdują się w obszarze zakrętu dolnego (IT) kory skroniowej oraz w korze przedczołowej (PF). Wysoka aktywność (20 Hz) neuronów w tych obszarach utrzymuje się do 30 s po prezentacji i widoczna jest w pomiarach potencjałów z wielu elektrod [28].

Korelacje czasowe wynikające z zapamiętania kolejności prezentacji obrazów po wielu powtórzeniach przechodzą w korelacje aktywności neuronów biorących udział w kodowaniu śladów pamięci. W trakcie uczenia się moduł biorący udział w rozpoznawaniu początkowo zwiększa w nieselektywny sposób częstość impulsacji dla wszystkich sygnałów. Powyżej krytycznej wartości wzmocnienia LTP pojawiają się lokalne atraktory na tle globalnej aktywności - struktura sygnału uczącego. W pierwszym etapie neurony mikrokolumn reagują na nauczone bodźce tylko gdy są pobudzane. W drugim etapie aktywność spoczynkowa rośnie do około 20 Hz i utrzymuje się po zniknięciu bodźca - jest to sygnał aktywnej reprezentacji bodźca w pamięci. Pobudzenia wewnętrzne, silniejsze niż zewnętrzne, utrzymują spontaniczną aktywność. W trzecim etapie powstają korelacje aktywności pomiędzy kolumnami, co pozwala na odtworzenie kolejności pojawiania się obrazów [28].

6. Przestrzenie umysłu.

Neurodynamika decyduje o zdarzeniach na poziomie umysłu: postrzeganych cechach, skojarzeniach, rozróżnianych obiektach i decyzjach podejmowanych na podstawie tych informacji. Opis zachodzących w umyśle zjawisk na poziomie neurodynamiki jest bardzo trudny. Opis na poziomie symbolicznym, za pomocą konstrukcji logicznych, jest z kolei zbyt uproszczony i często nienaturalny. K. Lewin próbował zdefiniować przestrzenie psychologiczne jako obszary, w których można umieścić elementy naszego doświadczenia, zdarzenia mentalne [16]. Jakich przestrzeni należy użyć by dostrzec ogólne prawa działania umysłu? Jeśli takie prawa istnieją, powinny być odbiciem ewolucyjnie wykształconych własności mózgu, internalizacją istotnej dla przetrwania naszego gatunku wiedzy o świecie.

Nie należy się spodziewać, by jeden model wystarczył do opisu wszystkich zjawisk – różne działy fizyki posługują się różnymi modelami. Należy dążyć do minimalizacji liczby wymiarów używanych przestrzeni i jak najprostszej reprezentacji stanu umysłu. Przestrzeń i czas są od czasów Newtona areną zdarzeń fizycznych, pozwalając na ilustrację procesów mechanicznych. Chociaż początkowo myślano, że są to pojęcie dane *a priori* Einstein wprowadził abstrakcyjną czasoprzestrzeń, w której czas i położenie związane są z rezultatami pomiarów. Przestrzenie fazowe i przestrzenie Hilberta są konieczne by uogólnić mechanikę Newtonowską. Przestrzenie psychologiczne rozpięte na osiach związanych z rezultatami pomiarów psychofizycznych, traktowane jako arena zdarzeń umysłowych, są niezbędne by stworzyć język pozwalający na precyzyjny opis zdarzeń mentalnych.

Neurodynamika posługuje się ogromną liczbą parametrów opisujących aktywność poszczególnych neuronów i ich synaps. Należy szukać takich przybliżeń do neurodynamiki, które będą prowadziły do najprostszych modeli zgodnych z psychofizyką. Na proces przetwarzania informacji w mózgu można patrzeć z punktu widzenia redukcji ilości informacji na rzecz jej jakości. Ilość informacji dostępna na poziomie świadomym, potrzebna do podejmowania decyzji, jest w porównaniu z ilością informacji otrzymywanej przez receptory mózgu bardzo niewielka. Stany umysłu są w porównaniu ze stanami mózgu bardzo uproszczone, postrzegamy jedynie cień neurodynamiki.

Poznanie jednego konkretnego przykładu obiektu nowego rodzaju pozwala na generalizację, stworzenie nowej kategorii. Jest to zrozumiałe, gdyż w sieci neuronowej każdy zapamiętany wzorzec ma swój basen atrakcji. W przestrzeni zdefiniowanej przez postrzegane cechy przypisać można zapamiętanym obiektom rozkłady gęstości prawdopodobieństwa (PDF) przypisania danego obiektu do jakiejś kategorii przy coraz większych deformacjach różnych cech zapamiętanego obiektu. Takie rozkłady wynikają z basenów atrakcji na poziomie neurodynamiki. Odległości pomiędzy różnymi obiektami O_i , O_j będą maleć wraz z wzrastającym podobieństwem obiektów. Sensowne odległości D_{ij} można ustalić za pomocą procedury skalowania wielowymiarowego na podstawie postrzeganego podobieństwa. Jak pokazał eksperymentalnie i teoretycznie Shepard [19] prawdopodobieństwo reakcji ludzi i zwierząt na wyuczony bodziec O_i przy pojawieniu się bodźca O_j wynosi $P(O_i, O_j) = \exp(-\alpha D_{ij})$ dla szerokiej gamy bodźców. Rozkłady PDF reprezentujące obiekty w przestrzeniach psychologicznych powinny więc zanikać eksponencjalnie. Zamiast proponowanych wcześniej [2] rozkładów Gaussowskich lepiej jest użyć naturalnych z punktu widzenia neurobiologii [29] funkcji sigmoidalnych, np. funkcji logistycznych $\sigma(x) = 1/(1 + \exp(-x))$. Różnica $\sigma(x+a) - \sigma(x+b)$ zanika w sposób eksponencjalny, mogąc jednocześnie utrzymać stałą wartość w większym obszarze, w którym deformacje cechy x nie mają znaczenia dla przypisania obiektów do danej kategorii.

Modele zdarzeń mentalnych wykorzystujące przestrzenie psychologiczne mogą pozwolić na integrację informacji behawioralnej otrzymanej metodami psychologii eksperymentalnej, oraz informacji dotyczących neurodynamiki otrzymanej metodami neurofizjologii. Od pomia-

rów aktywności neuronów możemy przejść do oceny siły bodźców, które należy następnie związać ze skalami dla poszczególnych wymiarów przestrzeni psychologicznej. Za pomocą wielu elektrod mierzymy aktywność neuronów r_i przy stosowaniu bodźca s otrzymując prawdopodobieństwa $P(r_i|s)$. Korzystając z twierdzenia Bayesa można teraz obliczyć prawdopodobieństwo posteriorne $P(s|r) = P(\text{stymulacja} | \text{reakcja})$:

$$P(s|r) = P(s|r_1, r_2, \dots, r_N) = P(s) \prod_{i=1}^N P(r_i|s) / \sum_{s'} P(s') \prod_{i=1}^N P(r_i|s')$$

Stwarza to pomost pomiędzy psychologią i neurofizjologią, prostszy niż próba użycia sieci neuronowych do modelowania w psychologii, zachowując zalety modeli sub-symbolicznych, posługujących się opisem ciągłych procesów, a więc bardziej szczegółowym niż opis za pomocą logiki. Upraszczając neurodynamikę model powinien reprezentować jej atraktory za pomocą rozkładów gęstości prawdopodobieństwa w niskowymiarowych przestrzeniach psychologicznych obrazując powstawanie kategorii i relacje między nimi. Takie upraszczanie może postępować wielostopniowo: należy określić aktywność poszczególnych kolumn kory (odpowiadającą ustaleniu wartości poszczególnych cech obiektu) a następnie całej populacji.

W wersji statycznej model ten ma za zadanie opisywać behawioralne (sensomotoryczne) lub kognitywne (oparte na pamięci) reakcje mózgu zachodzące w czasie nie dłuższym niż około sekundy. Pojawienie się bodźca wywołuje pewien stan umysłu i związane z nim skojarzenia modelowane w oparciu o strukturę przestrzeni psychologicznej odpowiedniej dla danego eksperymentu. Reprezentacja relacji, podobieństw pomiędzy obiektami, wystarcza do zrozumienia kategoryzacji i rozpoznawania obiektów [30], a więc realizacji podstawowych procesów poznawczych. Jak duża powinna być przestrzeń by można w niej było rozróżnić jednoznacznie wszystkie pojęcia? Analiza wielkiego korpusu tekstów zawierającego 60.000 słów doprowadziła Landauera i Dumais do wniosku [31], że wystarczy około 300 wymiarów. Nauczenie każdego nowego słowa ustawia go niejako w relacji ze wszystkimi już poznanymi słowami. Dzięki temu dowiadując się niewiele, dzieci szybko uczą się prawidłowo wielu relacji.

Teoria statyczna zastosowana została do problemów uczenia się kategorii [32]. W eksperymentach z kategoryzacją bada się szybkość uczenia i prawdopodobieństwa błędów w różnych warunkach. Pokazywane obiekty (np. karty lub figury geometryczne o kilku różnych kształtach, kolorach i wielkościach) dzieli się na kilka kategorii zgodnie z jakąś ukrytą regułą. Jeśli własności są binarne i tylko jedna z nich określa kategorię (np. kolor jest czarny lub biały) to łatwo jest zgadnąć odpowiednią regułę. Jeśli istotne są dwie własności możemy mieć np. regułę typu XOR: kategoria A jeśli Obiekt jest czarny i duży, lub Obiekt jest biały i mały, a kategoria B dla kombinacji (czarny, mały) lub (biały, duży). Uczenie się właściwej kategoryzacji zajmuje wówczas dłużej. Teoria powinna przewidywać tempo uczenia się i prawdopodobieństwa błędów.

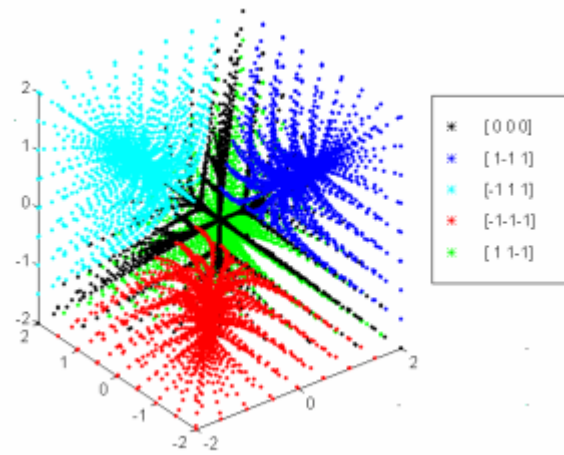
Co dzieje się w mózgu w czasie uczenia się definicji kategorii na podstawie przykładów? Złożone procesy neurodynamiczne realizują w końcu stosunkowo prostą dynamikę. Dla wszystkich reguł logicznych użytych w eksperymentach można napisać odpowiednia równania zawierające istotne cechy i kategorie. Np. dla problemów typu $z = x.xor.y$, gdzie $z = \pm 1$ jest kategorią, zaś $x, y = \pm 1$ to są cechy obiektów, od których ona zależy, najprostsza dynamika ma postać:

$$V(x, y, z) = 3xyz + \frac{1}{4}(x^2 + y^2 + z^2)^2$$

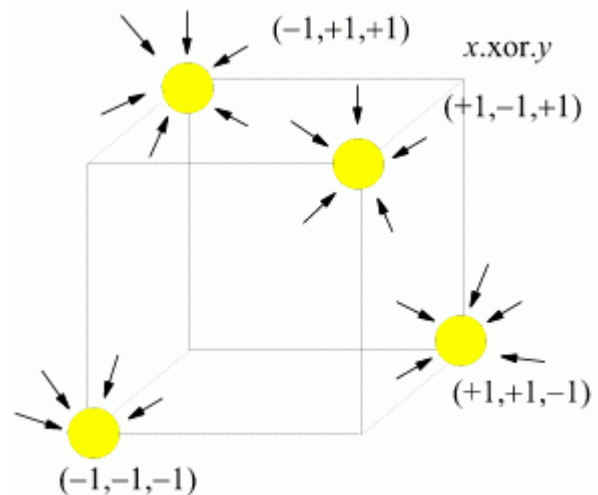
$$\dot{x} = -\frac{\partial V}{\partial x} = -3yz - (x^2 + y^2 + z^2)x$$

$$\dot{y} = -\frac{\partial V}{\partial y} = -3xz - (x^2 + y^2 + z^2)y$$

$$\dot{z} = -\frac{\partial V}{\partial z} = -3xy - (x^2 + y^2 + z^2)z$$



Na rysunku przedstawiono trajektorie obrazujące baseny czterech atraktorów punktowych dla tych równań. Choć u różnych osób zachodzące w mózgu procesy uczenia będą się od siebie znacznie różnić to efekt końcowy da się opisać za pomocą układu dynamicznego z 4 atraktorami (w dokładniejszym modelu nie będą one oczywiście punktowe). Jeszcze prościej można to opisać w przestrzeni psychologicznej rozpiętej na 3 użytych tu wymiarach. Strzałki pokazują tu kierunek ruchu w stronę stanu końcowego (x,y,z) przy startowaniu z różnych stanów początkowych.



Ekspertyzy z kategoryzacją dają czasami zaskakujące wyniki [33]. Mamy dwie choroby, często występującą C i rzadką R. Choroba C ma dwa symptomy, SC i SI, a choroba R też dwa symptomy, SR i SI. Po przestudiowaniu listy przypadków uczestnicy eksperymentu odpowiadają prawidłowo na takie pytania jak: jeśli zaobserwowano symptomy (PC, SI) to mamy prawdopodobnie chorobę ... O jakiej chorobie świadczy sam symptom SR? Oczywiście o R. A o jakiej sam symptom SI? Większość odpowiada rozsądnie „PC”, bo ta choroba występowała częściej. Podobnie dla kombinacji symptomów (SC, SI, SR) około 60% osób odpowiada „PC”. Jednak dla kombinacji symptomów SC i SR taki sam procent osób odpowiada „PR”. Dlaczego w tym przypadku ludzie nie stawiają na najczęściej występującą chorobę?

Sytuacja jest tu bardziej skomplikowana niż poprzednio, gdyż przestrzeń psychologiczna ma 5 wymiarów, $\{C, R, SI, SC, SR\}$. Baseny atraktorów neurodynamiki można ocenić na podstawie odpowiedzi. Interpretacja psychologiczna jest tu mało przekonująca: symptom SR ma większe znaczenie gdy pojawia się z SC ponieważ jest to symptom wyróżniający, chociaż występuje rzadziej. Dlaczego dodanie symptomu SI zmienia w takim razie odpowiedzi?

Rozwiązanie wydaje się leżeć w sposobie formowania się basenów atrakcji w czasie uczenia się. Kombinacja $\{SC, SI, C\}$, powtarzająca się często, tworzy szybciej duży basen atrakcji, zaś basen dla $\{SR, SI, R\}$ jest mniejszy. Aktywacja $\{SR, SC\}$ przy nieokreślonym SI częściej prowadzi do odpowiedzi R ponieważ gradient w kierunku R jest większy. Zwykle udaje się podać racjonalne przyczyny decyzji, ale w niektórych sytuacjach eksperymentalnych może to być trudne. Nie mamy wpływu na percepcję, np. ulegamy iluzjom wzrokowym, co ułatwia badanie mechanizmów działania układu wzrokowego. Podejmowanie decyzji przypomina percepcję, gdyż postrzegamy w swoim umyśle wynik przebiegających w mózgu procesów, a procesy interpretacji dostarczają nam racjonalizacji odnosząc decyzję do modelu „ja” i swoich

potrzeb. Być może w eksperymentach z kategoryzacją mamy okazję dostrzec załamanie się tych mechanizmów.

7. Model dynamiczny.

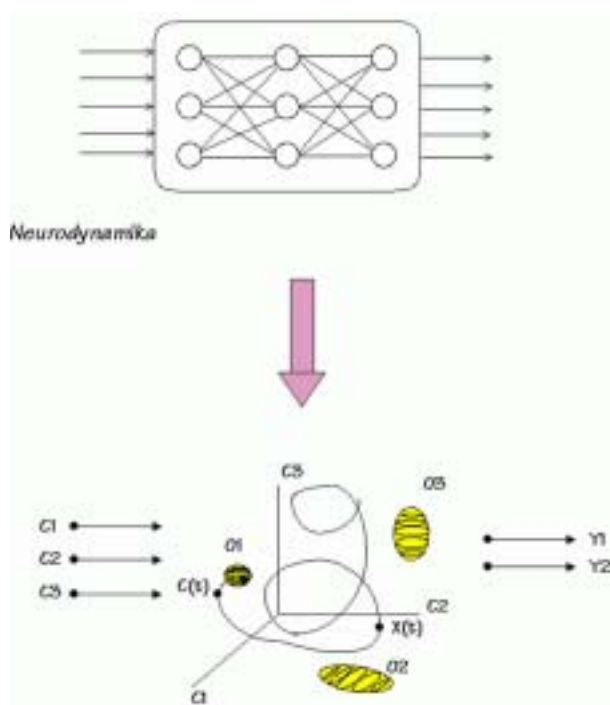
Opisany powyżej model statyczny przydatny jest do interpretacji szybkich reakcji. W przestrzeniach psychologicznych określone są rozkłady prawdopodobieństw rozpoznania obiektów przy określonej kombinacji cech. Odpowiada to elementom trwałej pamięci związanym z atraktorami neurodynamiki. Przestrzeń psychologiczna zawiera wymiary odpowiadające cechom wewnętrznym, które traktować możemy jako pomiary wykonane przez obszary mózgu zajmujące się przetwarzaniem bodźców zmysłowych. Umysł działa tu jako układ kontrolny, podejmując decyzje na podstawie wyników takich pomiarów. Przestrzeń umysłu powinna uwzględniać zarówno cechy wewnętrzne jak i podejmowane działania.

W doświadczeniach ze zwierzętami mamy do czynienia z sensomotorycznymi obiektami umysłu: po rozpoznaniu jakiegoś bodźca zwierzę wykonuje jedną z kilku wyuczonych czynności. Np. w eksperymentach z układem węchowym królików rozpoznawanych jest kilka rodzajów zapachów, na które królik reaguje w specyficzny sposób [34]. Zagadnienie można więc rozpatrywać w przestrzeni o liczbie wymiarów nie większej niż 10, podczas gdy procesy neurodynamiczne zachodzą w przestrzeniach o wymiarach setek milionów. Atraktry opisujące zachowanie trudno jest w takich warunkach dostrzec wśród ogólnych zmian stanu kory mózgu.

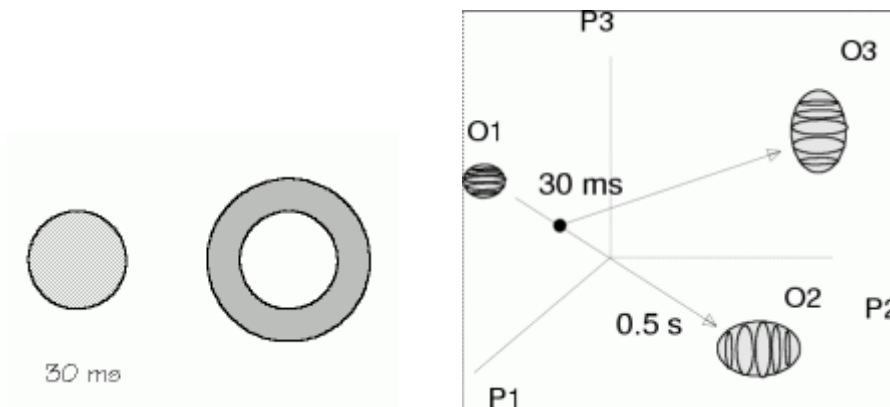
W normalnych warunkach doświadczamy zmiennych stanów umysłu, określanych mianem „strumienia świadomości”. Stany mózgu przechodzą w kolejne możliwe stany z pewnym prawdopodobieństwem. Chwilowy stan umysłu możemy powiązać z określonym punktem lub zlokalizowanym obszarem przestrzeni psychologicznej. Taki obszar odpowiada części aktywnej pamięci długotrwałej, „kopiowanej” do pamięci roboczej. Dynamika aktywacji kolejnych obiektów umysłu przedstawianych w przestrzeniach cech powinna odtwarzać dynamikę stanów mózgu (neurodynamikę).

Jeśli w chwili t_0 obiektem aktywnym jest O_1 to stan umysłu reprezentowany jako punkt w przestrzeni psychologicznej znajduje się w obszarze dużej wartości PDF dla tego obiektu. Pod wpływem bodźców zewnętrznych jak i w wyniku wewnętrznej dynamiki stan ten będzie się zmieniał podążając po „najmniejszej linii oporu” od obiektu do obiektu. W najprostszym przypadku ruch ten obrazuje proces aktywacji kolejnych elementów pamięci, np. serii wyuczonych zachowań.

Jedną z możliwości opisu takiej dynamiki wzorowana jest na mechanice klasycznej. Stan umysłu ma pewną bezwładność (związaną ze stabilnością atraktora reprezentującego obiekt podlegający aktywacji). Można mu przypisać też pęd utrudniający zmianę kierunku trajektorii („kierunek myślenia”). Przejście od aktywacji jednego obiektu do drugiego wymaga energii. Mózg jest układem otwartym, zużywającym aż 20% całej energii organizmu. Trajektorie obrazujące stan umysłu będą szybko przeskakiwać pomiędzy obiektami i pozostawać w ich pobliżu.



Obiektom można przypisać pewne potencjały, określone przez funkcję zadaną w przestrzeni psychologicznej. Funkcja ta określa strukturę potencjalnie dostępnych stanów umysłu. Dynamikę w takim układzie można symulować podobnie jak dynamikę ładunku próbnego w układach elektrostatycznych, z uwzględnieniem sił stochastycznych i zewnętrznych sił wymuszających. Zadaniem takich modeli jest odtworzenie prawdopodobieństw przejść między atraktorami w mikroskopowym opisie neurodynamicznym, a więc sieci relacji pomiędzy stanami umysłu i zachowaniami organizmu. W przestrzeni psychologicznej miarą odległości aproksymującą te prawdopodobieństwa będzie najmniejsze działanie Hamiltona, wyznaczające trajektorie geodezyjne. Taka miara odległości używana jest w przestrzeniach Finslera [35]. Nie jest ona symetryczna, albowiem nie mamy tu procesów odwracalnych.

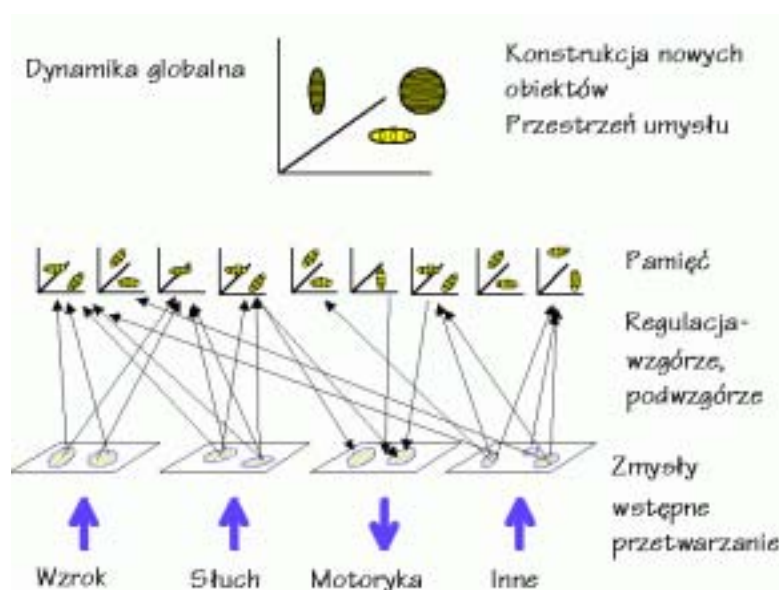


Przydatność takich pojęć jak pęd czy bezwładność stanu umysłu można dostrzec przy analizie eksperymentów psychofizycznych dotyczących maskowania [36]. Jeśli po ekspozycji pierwszego obrazu, np. wypełnionego kółka, pokazany jest szybko drugi obraz, np. pierścienia, obserwator widzi tylko drugi obraz. Brak dobrego języka by opisywać takie zagadnienia przyczynia się do kontrowersji typu [36]: czy drugi obraz wymazuje wrażenia powstałe w wyniku pojawienia się pierwszego (zmienia przeszłość), czy też nie dopuszcza do jego powstania? Proces ten można opisać w sposób przypominającym zderzenie lub rozproszenie. Początkowo stan umysłu znajduje się w pobliżu O1 (patrzmy na pusty ekran), potem pojawia się bodziec O2, ale zanim stan umysłu znajdzie się w jego obszarze (rozpoznane zostanie kółko) bodziec maskujący O3 zmienia trajektorię stanu umysłu i nie dochodzi do aktywacji O2. Wrażenia związane z pierwszym bodźcem nie powstaną, jeśli obiekt O3 będzie podobny do O2 oraz znajdzie się w pobliżu. Obiekty całkiem odmienne, np. o innej modalności, nie mają wpływu na aktywizację O2. Nawet jeśli dojdzie do zmiany trajektorii – można sobie wyobrazić kaskadę takich rozprożeń blokującą powstawanie wrażeń – część energii dociera do O2 i pomimo tego, że nie wystarcza to do aktywacji obiektu (przeniesienia go do pamięci roboczej) to może mieć wpływ na takie decyzje, jak zgadywanie czy były dwa czy jeden bodziec [36]. Podobnym językiem można się posłużyć przy analizie wielu eksperymentów psychofizycznych lub badaniach pamięci związanych z poprzedzaniem (ang. *priming*, tłumaczone czasem jako „prymowanie”).

Z powyższych rozważań widać, że przestrzenie psychologiczne mogą być przydatne w dobrze określonych sytuacjach eksperymentalnych. W geometrycznym modelu działania umysłu przestrzenie psychologiczne wykorzystywane są na różnych etapach [2]:

- wykrywania istotnych cech sygnałów zmysłowych – odpowiada za to kora sensoryczna, a głównym mechanizmem analizy są mapy topograficzne;
- rozpoznawania podstawowych obiektów – kombinacje cech określają prawdopodobieństwo pobudzenia się reprezentacji obiektów w pamięci długotrwałej;

- pamięci roboczej – najbardziej aktywne elementy kopiowane są z pamięci długotrwałej do pamięci roboczej, umożliwiając podejmowanie decyzji i kontrolę zachowania.



8. Dyskusja

Fizyka odwróciła się w znacznej mierze od zagadnień psychofizycznych dopiero na początku XX wieku. Działanie mózgu było wówczas całkowitą tajemnicą a psychologia eksperymentalna dopiero zaczynała się rozwijać. Obecnie sytuacja uległa radykalnej zmianie. Coraz lepiej rozumiemy działanie mózgu na poziomie mikroskopowym jak i na poziomie systemowym. Neurofizjologia bada szczegóły procesów zachodzących w mózgu, ale modeli działania umysłu dostarczyć może tylko fizyka. W tym celu trzeba poszukiwać nie tylko dobrych modeli aktywności mózgu na poziomie neuronowym, ale i cech pozwalających na niskowymiarowe reprezentacje zdarzeń mentalnych i sposobów upraszczania neurodynamiki.

Naszkieowany tu język opisu zdarzeń w przestrzeniach psychologicznych pozwala na modelowanie niektórych zdarzeń mentalnych, a jednocześnie jest redukowalny do zdarzeń neurofizjologicznych. Pierwotne obiekty umysłu skonstruowane są z danych sensorycznych i reakcji ruchowych, czyli działań sensomotorycznych, nadających im sens związany z bezpośrednimi przeżyciami „bycia w świecie”. Takie pojęcia jak kamień, zwierzę, siła czy pęd odwołują się do pierwotnych obiektów umysłu, dzięki czemu są intuicyjnie dobrze zrozumiałe. Rozwój mowy spowodował powstanie coraz bardziej abstrakcyjnych pojęć, których znaczenie można określić przez ich relacje z pierwotnymi obiektami. Związek takich abstrakcyjnych pojęć z rzeczywistością staje się jednak intuicyjnie trudny do uchwycenia – tak jest np. z pojęciem funkcji falowej.

Próba opisu zjawisk mentalnych jest oczywiście znacznie trudniejsza niż typowych zjawisk fizycznych. Trudności techniczne wynikają zarówno z niepełnej wiedzy dotyczącej działania mózgu, sposobu analizy sygnałów zmysłowych mogących stanowić podstawę dla opisu obiektów umysłu, czy też samej złożoności modeli neurodynamicznych i sposobów ich upraszczania. Na najwyższym poziomie niskowymiarowe przestrzenie psychologiczne opisujące pamięć roboczą można zdefiniować tylko zakładając szybką zmienność liczby i rodzaju

cech rozpinających przestrzeń. Wybór obiektów pojawiających się w tej „przestrzeni umysłu” związany jest z działaniem mechanizmów uwagi, które dopiero uczymy się modelować.

Geometryczne unaocznienie nawet prostych eksperymentów wymaga wielowymiarowych przestrzeni. Przydatność takich modeli jak i ich ograniczenia nie są jeszcze jasne. Celem jest nie tyle wyjaśnienie wszystkich zagadek umysłu za pomocą jednego prostego modelu, co próba zmniejszenia przepaści pomiędzy badaniami nad mózgiem a psychologią przez dostarczenie precyzyjnego języka pozwalającego na analizę rezultatów konkretnych eksperymentów w wzorowany na fizyce sposób.

Podziękowania: jestem wdzięczny prof. Piotrowi Jaśkowskiemu z Akademii Bydgoskiej im. Karola Wielkiego za wnikliwe przeczytanie pracy i szereg krytycznych uwag.

Literatura

- [1] R. Penrose, *Nowy umysł cesarza* (Oxford Univ. Press, 1994)
- [2] E. Schrödinger, *What is life* (Cambridge Univ. Press, 1944)
- [3] W. Duch, „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 1997), rozdz. 20, str. 491-512
- [4] K. Walsh, *Neuropsychologia kliniczna* (Wyd. Naukowe PWN, Warszawa 1998)
- [5] W. Duch, „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. Akademicka Oficyna Wydawnicza-Exit, Warszawa.
- [6] P. Cichosz, *Systemy uczące się* (WNT, Warszawa 2000)
- [7] W. Duch, „Świadomość i dynamiczne modele działania mózgu”. *Neurologia i Neurochirurgia Polska* **34**(50), Supl. 2, pp. 69-84 (2000)
- [8] A.D. Baddeley, „Consciousness and working memory”. *Consciousness and Cognition* **1**, 3-6 (1992)
- [9] E. Kandel, J. Schwartz, T. Jessell, *Principles of neural science* (3rd ed. Prentice Hall Int. 1991)
- [10] N. Cowan, „The magical number 4 in short-term memory: A reconsideration of mental storage capacity”. *Behavioral and brain sciences* **24**(1) (2001) 87-114
- [11] D.A. Leopold, N.K. Logothetis, „Multistable phenomena: changing views in perception”. *Trends in Cognitive Studies* **3**(7), 254-265 (1999)
- [12] S. Zeki, „Localization and globalization in conscious vision”. *Annual Review of Neuroscience* **24**:57-86 (2001)
- [13] J.K. O'Regan, A. Noë, „A sensorimotor account of vision and visual consciousness”. *Behavioral and Brain Sciences* **24**(5) (2002, w druku)
- [14] M.S. Gazzaniga, red. *The Cognitive Neurosciences* (A Bradford Book 1995); por. też M.S. Gazzaniga, *O tajemnicach ludzkiego umysłu. Biologiczne korzenie myślenia, emocji, seksualności, języka i inteligencji* (Książka i Wiedza, Warszawa 1997); M.S. Gazzaniga, *The mind's past*. Univ. of California Press 1998.
- [15] G. Rizzolatti, L. Fadiga, V. Gallese, L. Fogassi, „Premotor cortex and the recognition of motor actions”. *Cognitive Brain Research* **3**, 131-141 (1996)
- [16] K. Lewin, *The conceptual representation and the measurement of psychological forces* (Durham, N.C, Duke University Press 1938)
- [17] G.A. Kelly, *The Psychology of Personal Constructs* (Norton, New York 1955)

- [18] M.L.G. Shaw, B.R. Gaines, “Kelly's ‘Geometry of Psychological Space’ and its Significance for Cognitive Modeling”. *The New Psychologist* **10**, 23-31 (1992)
- [19] R.N. Shepard, “Toward a universal law of generalization for psychological science”. *Science* **237**, 1317-1323 (1987)
- [20] R.N. Shepard, “Perceptual-Cognitive Universals as Reflections of the World”. *Psychonomic Bulletin & Review* **1**, 2-28 (1994)
- [21] M. Leyton, *Symmetry, causality, mind* (MIT Press 1992)
- [22] L.T. Maloney, “Evaluation of linear models of surface spectral reflectance with small numbers of parameters”. *J. Optical Soc. of America A* **3**, 1673-1683 (1986)
- [23] R.N. Shepard, “On the physical basis, linguistic representation, and conscious experience of colors”, w: G. Harman (red.), *Conceptions of the mind*, str. 217-245 (Lawrence Erlbaum Associates, Hillsdale, NJ 1993).
- [24] J. F. Traub, “A Continuous Model of Computation”, *Physics Today* **5**, 39-43 (1999)
- [25] W. Maass, T. Natschläger, H. Markram, “Real-time computing without stable states: A new framework for neural computation based on perturbations” (preprint, Technische Universität Gratz, 2001)
- [26] D.J. Amit, “The Hebbian paradigm reintegrated: Local reverberations as internal representations”. *Behavioral and brain sciences* **18**(4): 617-26 (1994)
- [27] D.J. Amit, N. Brunel, “Model of global spontaneous activity and local structured (learned) delay activity in cortex”. *Cerebral Cortex* **7**, 237-252 (1997)
- [28] V. Yakovlev, S. Fusi, E. Berman, E. Zohary, “Inter-trial neuronal activity in inferotemporal cortex: a putative vehicle to generate long term associations”. *Nature Neuroscience* **1**(4), 310-31 (1998)
- [29] J.A. Anderson, *An Introduction to Neural Networks* (MIT Press/Bradford Book 1995)
- [30] S. Edelman, *Representation and recognition in vision* (MIT Press 1999)
- [31] T.K. Landauer, S.T. Dumais, “A solution to Plato's Problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge”. *Psychological Review* **104**, 211-240 (1997)
- [32] W. Duch, “Categorization, Prototype Theory and Neural Dynamics”. Proc. of the 4th International Conference on Soft Computing, Iizuka, Japonia 1996, red. T. Yamakawa, G. Matsumoto, pp. 482-485
- [33] D.L. Medin, S.M. Edelson, “Problem structure and the use of base-rate information from experience”. *Journal of Experimental Psychology: General* **117**, 68-85 (1988)
- [34] W.J. Freeman, *Societies of Brains: A study in the neuroscience of love and hate* (Lawrence Erlbaum Associates 1995)
- [35] P.L. Antonelli, R.S. Ingarden, M. Matsumoto, *The Theory of Sprays and Finsler Spaces with Applications in Physics and Biology* (Kluwer Academic, Dodrecht 1993)
- [36] B.G. Breitmeyer, *Visual Masking* (Oxford: Clarendon Press 1984)
- [37] D.C. Dennett, M. Kinsbourne, “Time and the observer: The where and when of consciousness in the brain”. *Behavioral and Brain Sciences* **15**(2), 183-247 (1995)